Power analysis of projection-pursuit independence tests

Kai Xu and Liping Zhu

Anhui Normal University

Renmin University of China and Zhejiang Gongshang University

This supplement contains proofs of Theorems 1–4, Proposition 1 and (1.2).

S1. Proof of Theorem 1

For any $\mathbf{t}_1 \in \mathbb{R}^p$ and $\mathbf{t}_2 \in \mathbb{R}^q$, given independent and identically distributed data $\{(\mathbf{x}_i, \mathbf{y}_i), i = 1, \ldots, n\}$, the characteristic function of $(\mathbf{x}_{i_k}, \mathbf{y}_{j_l})$ is equal to

$$n^{-2} \sum_{i,j=1}^{n} \exp(i\mathbf{t}_1^{\mathrm{T}}\mathbf{x}_i + i\mathbf{t}_2^{\mathrm{T}}\mathbf{y}_j),$$

which, by the law of large numbers, converges to $E\{\exp(i\mathbf{t}_1^{\mathrm{T}}\mathbf{x}_1 + i\mathbf{t}_2^{\mathrm{T}}\mathbf{y}_2)\}$. By Lévy's theorem, $(\mathbf{x}_{i_k}, \mathbf{y}_{j_l})$ converges in distribution to $(\mathbf{x}_1, \mathbf{y}_2)$. By continuous mapping theorem, it follows that $a(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \mathbf{x}_{i_3}, \mathbf{x}_{i_4}, \mathbf{x}_{i_5})a(\mathbf{y}_{j_1}, \mathbf{y}_{j_2}, \mathbf{y}_{j_3}, \mathbf{y}_{j_4}, \mathbf{y}_{j_5})$ and $a(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \mathbf{x}_{i_3}, \mathbf{x}_{i_4}, \mathbf{x}_{i_5})a(\mathbf{y}_{j_1}, \mathbf{y}_{j_2}, \mathbf{y}_{j_3}, \mathbf{y}_{j_4}, \mathbf{y}_{j_6})$ converge in distribution to $a(\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2, \widetilde{\mathbf{x}}_3, \widetilde{\mathbf{x}}_4, \widetilde{\mathbf{x}}_5)a(\widetilde{\mathbf{y}}_1, \widetilde{\mathbf{y}}_2, \widetilde{\mathbf{y}}_3, \widetilde{\mathbf{y}}_4, \widetilde{\mathbf{y}}_5)$ and $a(\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2, \widetilde{\mathbf{x}}_3, \widetilde{\mathbf{x}}_4, \widetilde{\mathbf{x}}_5)a(\widetilde{\mathbf{y}}_1, \widetilde{\mathbf{y}}_2, \widetilde{\mathbf{y}}_3, \widetilde{\mathbf{y}}_4, \widetilde{\mathbf{y}}_6)$ respectively, where $\{\widetilde{\mathbf{x}}_i, \widetilde{\mathbf{y}}_i\}_{i=1}^n$ denotes an independent and identically distributed sample from $(\mathbf{x}_1, \mathbf{y}_2)$. Because of $| a(\cdot) | \le 4\pi$, $a(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \mathbf{x}_{i_3}, \mathbf{x}_{i_4}, \mathbf{x}_{i_5}) a(\mathbf{y}_{j_1}, \mathbf{y}_{j_2}, \mathbf{y}_{j_3}, \mathbf{y}_{j_4}, \mathbf{y}_{j_5})$ is uniformly integrable. Thereby, the above convergence also holds in L_2. This, together with Theorem 2.1 of Leucht and Neumann (2009), completes the proof for the first claim.

Following similar arguments, we show that $b(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \mathbf{x}_{i_3}, \mathbf{x}_{i_4}) b(\mathbf{y}_{j_1}, \mathbf{y}_{j_2}, \mathbf{y}_{j_3}, \mathbf{y}_{j_4})$ converges in distribution to $b(\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2, \widetilde{\mathbf{x}}_3, \widetilde{\mathbf{x}}_4) b(\widetilde{\mathbf{y}}_1, \widetilde{\mathbf{y}}_2, \widetilde{\mathbf{y}}_3, \widetilde{\mathbf{y}}_4)$. For convenience, write $b^*_{\mathbf{x}} \stackrel{\text{def}}{=} b(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \mathbf{x}_{i_3}, \mathbf{x}_{i_4})$ and $b^*_{\mathbf{y}} \stackrel{\text{def}}{=} b(\mathbf{y}_{j_1}, \mathbf{y}_{j_2}, \mathbf{y}_{j_3}, \mathbf{y}_{j_4})$. For any $\epsilon > 0$, $E^*\{| b^*_{\mathbf{x}} b^*_{\mathbf{y}} | I(| b^*_{\mathbf{x}} b^*_{\mathbf{y}} | \ge \epsilon)\}$ is less than or equal to $E^*\{| b^*_{\mathbf{x}} b^*_{\mathbf{y}} | I(| b^*_{\mathbf{x}} | \ge \epsilon^{1/2})\} + E^*\{| b^*_{\mathbf{x}} b^*_{\mathbf{y}} | I(| b^*_{\mathbf{y}} | \ge \epsilon^{1/2})\}$. By the definition of permutation as well as the law of large numbers, we can show $E^*\{| b^*_{\mathbf{x}} b^*_{\mathbf{y}} | I(| b^*_{\mathbf{x}} | \ge \epsilon^{1/2})\} = E\{| b(\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2, \widetilde{\mathbf{x}}_3, \widetilde{\mathbf{x}}_4) | I(| b(\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2, \widetilde{\mathbf{x}}_3, \widetilde{\mathbf{x}}_4) | \ge \epsilon^{1/2})\} E\{| b(\widetilde{\mathbf{y}}_1, \widetilde{\mathbf{y}}_2, \widetilde{\mathbf{y}}_3, \widetilde{\mathbf{y}}_4) |\} + o(1)$. By Markov's inequality, $E\{| b(\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2, \widetilde{\mathbf{x}}_3, \widetilde{\mathbf{x}}_4) | I(| b(\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2, \widetilde{\mathbf{x}}_3, \widetilde{\mathbf{x}}_4) | \ge \epsilon^{1/2})\} \le \epsilon^{-1/2} E\{b^2(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)\}$. By triangle inequality, we have $b^2(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) \le C(\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2 + \|\mathbf{x}_3\|^2 + \|\mathbf{x}_4\|^2)$ and $| b(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) | \le C(\|\mathbf{y}_1\| + \|\mathbf{y}_2\| + \|\mathbf{y}_3\| + \|\mathbf{y}_4\|)$ for some constant $C > 0$. Together with $E\{\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2\} < \infty$, we have that for any ϵ large enough, $E^*\{| b^*_{\mathbf{x}} b^*_{\mathbf{y}} | I(| b^*_{\mathbf{x}} | \ge \epsilon^{1/2})\} \to 0$ in probability. Similarly, we can show that for any ϵ large enough, $E^*\{| b^*_{\mathbf{x}} b^*_{\mathbf{y}} | I(| b^*_{\mathbf{y}} | \ge \epsilon^{1/2})\} \to 0$ in probability. Consequently, $b^*_{\mathbf{x}} b^*_{\mathbf{y}}$ is uniformly integrable. The rest proof can be done through using similar arguments for proving the first statement. $\square$

S2. Proof of Proposition 1

We only prove the first claim. The proofs for the other two assertions are very similar and hence omitted. Let $\{K_i\}_{i=1}^B$ and $\{\widetilde{K}_i\}_{i=1}^B$ be two independent sequences of uniformly distributed random variables on $\{1, \ldots, (n)_n\}$. Then introduce for all $k \in \{1, \ldots, (n)_n\}$ and for all $i \in \{1, \ldots, B\}$ the random variables

$$Y_{ik} = I(K_i = k) \quad \text{and} \quad \widetilde{Y}_{ik} = I(\widetilde{K}_i = k).$$

By definition, both Y_{ik} and $\widetilde{Y}_{ik}$ have a Bernoulli distribution with parameter $\{(n)_n\}^{-1}$. Furthermore, we can write

$$\begin{aligned} B^{-1} \sum_{b=1}^{B} I\left\{n\ \widehat{\mathrm{PC}}(\mathbf{x}^b, \mathbf{y}^b) \le t\right\} &= \sum_{k=1}^{(n)_n} \sum_{\widetilde{k}=1}^{(n)_n} B^{-1} \sum_{i=1}^{B} Y_{ik} \widetilde{Y}_{i\widetilde{k}} I\left\{n\ \widehat{\mathrm{PC}}(\mathbf{x}^k, \mathbf{y}^{\widetilde{k}}) \le t\right\} \\ &= \sum_{k=1}^{(n)_n} \sum_{\widetilde{k}=1}^{(n)_n} I\left\{n\ \widehat{\mathrm{PC}}(\mathbf{x}^k, \mathbf{y}^{\widetilde{k}}) \le t\right\} B^{-1} \sum_{i=1}^{B} Y_{ik} \widetilde{Y}_{i\widetilde{k}}. \end{aligned}$$

By the strong law of large numbers, it follows that given the data,

$$\begin{aligned} \lim_{B \to \infty} B^{-1} \sum_{b=1}^{B} I\left\{n\ \widehat{\mathrm{PC}}(\mathbf{x}^b, \mathbf{y}^b) \le t\right\} &= \{(n)_n\}^{-2} \sum_{k=1}^{(n)_n} \sum_{\widetilde{k}=1}^{(n)_n} I\left\{n\ \widehat{\mathrm{PC}}(\mathbf{x}^k, \mathbf{y}^{\widetilde{k}}) \le t\right\} \\ &= \mathrm{pr}\{n\ \widehat{\mathrm{PC}}(\mathbf{x}^b, \mathbf{y}^b) \le t \mid \mathcal{D}_n\}, \end{aligned}$$

almost surely. This, together with Lemma 2.11 of van der Vaart and Wellner (1996), completes the proof of the first statement. □

S3. Proof of Theorem 2

It is apparent that

$$\lim_{n\to\infty} \inf_{H^{(n)}_{\mathbf{x},\mathbf{y}}} \mathrm{pr}(\Phi^{PC}_{\alpha} = 1 \mid H_1) \quad = \quad 1 - \lim_{n\to\infty} \sup_{H^{(n)}_{\mathbf{x},\mathbf{y}}} \mathrm{pr}\{n\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y}) < q^{PC}_{\alpha,n} \mid H_1\}.$$

To establish the first statement, it suffices to show $\lim_{n\to\infty} \sup_{H^{(n)}_{\mathbf{x},\mathbf{y}}} \mathrm{pr}\{n\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y}) \leq q^{PC}_{\alpha,n} \mid H_1\} = 0$. Invoking the fact $\mid a(\cdot) \mid \leq 4\pi$ and Lemma 5.2.1A of Serfling (1980), we have $\mathrm{var}\{\widehat{\mathrm{PC}}(\mathbf{x}^b, \mathbf{y}^b) \mid \mathbf{x}_1, \cdots, \mathbf{x}_n\} \leq O(n^{-1})$. This implies $q^{PC}_{\alpha,n} = O_p(n^{1/2}) = o_p(n)$. This, together with $\mathrm{PC}(\mathbf{x},\mathbf{y}) \geq \varpi$, indicates that there exists n_0 large enough such that for all $n \geq n_0$, $\varpi/2 \geq n^{-1} q^{PC}_{\alpha,n}$. Using Chebyshev's inequality, we have that $\lim_{n\to\infty} \sup_{H^{(n)}_{\mathbf{x},\mathbf{y}}} \mathrm{pr}\{n\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y}) < q^{PC}_{\alpha,n} \mid H_1\}$ is equal to

$$\begin{aligned}
& \lim_{n\to\infty} \sup_{H^{(n)}_{\mathbf{x},\mathbf{y}}} \mathrm{pr}\left[\frac{\mathrm{PC}(\mathbf{x},\mathbf{y}) - \widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y})}{\mathrm{var}^{1/2}\{\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y})\}} > \frac{\mathrm{PC}(\mathbf{x},\mathbf{y}) - n^{-1} q^{PC}_{\alpha,n}}{\mathrm{var}^{1/2}\{\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y})\}} \mid H_1\right] \\
\leq \; & \lim_{n\to\infty} \sup_{H^{(n)}_{\mathbf{x},\mathbf{y}}} \mathrm{pr}\left[\frac{\mathrm{PC}(\mathbf{x},\mathbf{y}) - \widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y})}{\mathrm{var}^{1/2}\{\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y})\}} \geq \frac{\mathrm{PC}(\mathbf{x},\mathbf{y})}{2\mathrm{var}^{1/2}\{\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y})\}} \mid H_1\right] \\
\leq \; & \lim_{n\to\infty} \sup_{H^{(n)}_{\mathbf{x},\mathbf{y}}} \frac{4\mathrm{var}\{\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y})\}}{\mathrm{PC}^2(\mathbf{x},\mathbf{y})} \leq \lim_{n\to\infty} \frac{4C}{n\varpi^2} = 0
\end{aligned}$$

where the positive constant C is generic and its exact value may vary at each appearance. Hence, $\lim_{n\to\infty} \inf_{H^{(n)}_{\mathbf{x},\mathbf{y}}} \mathrm{pr}(\Phi^{PC}_{\alpha} = 1 \mid H_1) = 1$. As the kernel associated with $\widehat{\mathrm{mBKR}}(\mathbf{x},\mathbf{y})$ is bounded, apply arguments similar to those for dealing with $\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y})$ to obtain $\lim_{n\to\infty} \inf_{H^{(n)}_{\mathbf{x},\mathbf{y}}} \mathrm{pr}(\Phi^{mBKR}_{\alpha} = 1 \mid H_1) = 1$.

For the distance correlation test, we choose $H_{12}^{(n)}$ properly to show that this statistic is dominated by the samples drawn from the contaminations $H_{\mathbf{x},\mathbf{y}}^{(n)}$. Let $(\mathbf{0}, \boldsymbol{\Sigma})$ denote any random vector with mean zero and covariance $\boldsymbol{\Sigma}$. Specifically, we pick $H_{\mathbf{x},\mathbf{y}}^{(n)} = H_{\mathbf{x}}^{(n)} H_{\mathbf{y}}^{(n)}$, where $H_{\mathbf{x}}^{(n)}$ and $H_{\mathbf{y}}^{(n)}$ have the same distributions as $(\mathbf{0}_p, \sigma_{1,n}^2 \boldsymbol{\Sigma}_1)$ and $(\mathbf{0}_q, \sigma_{2,n}^2 \boldsymbol{\Sigma}_2)$, respectively. Here $\sigma_{1,n}^2$ and $\sigma_{2,n}^2$ will be specified later, $\mathrm{tr}(\boldsymbol{\Sigma}_1) + \mathrm{tr}(\boldsymbol{\Sigma}_2) < \infty$, and $(\mathbf{0}_p, \boldsymbol{\Sigma}_1)$ is independent of $(\mathbf{0}_q, \boldsymbol{\Sigma}_2)$. Under this setting, we define the truncated random vectors $\mathbf{x}^{(T)}$ and $\mathbf{y}^{(T)}$ given by

$$\mathbf{x}^{(T)} = \begin{cases} \mathbf{0}_p & \text{if } \mathbf{x} \sim F_{\mathbf{x}}, \\ \mathbf{x}/\sigma_{1,n} & \text{if } \mathbf{x} \sim H_{\mathbf{x}}^{(n)}, \end{cases}, \text{ and } \mathbf{y}^{(T)} = \begin{cases} \mathbf{0}_q & \text{if } \mathbf{y} \sim F_{\mathbf{y}}, \\ \mathbf{y}/\sigma_{2,n} & \text{if } \mathbf{y} \sim H_{\mathbf{y}}^{(n)}, \end{cases}$$

where $F_{\mathbf{x}}$ and $F_{\mathbf{y}}$ are the marginal distributions of $\mathbf{x}$ and $\mathbf{y}$, respectively. By the construction, $\mathbf{x}^{(T)}$ and $\mathbf{y}^{(T)}$ have the respective mixture distributions as

$$\mathbf{x}^{(T)} \sim (1-\epsilon)\mathrm{pr}(\boldsymbol{\omega}_p = \mathbf{0}_p) + \epsilon H_{\mathbf{x}}^{(1)}, \text{ and } \mathbf{y}^{(T)} \sim (1-\epsilon)\mathrm{pr}(\boldsymbol{\omega}_q = \mathbf{0}_q) + \epsilon H_{\mathbf{y}}^{(2)},$$

where $\boldsymbol{\omega}_d$ has the degenerating distribution at $\mathbf{0}_d$, i.e., $\mathrm{pr}(\boldsymbol{\omega}_d = \mathbf{0}_d) = 1$, and $H_{\mathbf{x}}^{(1)}$ and $H_{\mathbf{y}}^{(2)}$ are the marginal distributions of $(\mathbf{0}_p, \boldsymbol{\Sigma}_1)$ and $(\mathbf{0}_q, \boldsymbol{\Sigma}_2)$, respectively.

If the distance correlation test using the original samples is asymptotically equivalent to that using the truncated ones drawn independently from the mixture distributions, then we can illustrate that the distance correlation test is asymptotically powerless. Suppose $\{\mathbf{x}_1^{(T)}, \mathbf{y}_1^{(T)}\}, \cdots, \{\mathbf{x}_n^{(T)}, \mathbf{y}_n^{(T)}\}$ are generated from $\{\mathbf{x}^{(T)}, \mathbf{y}^{(T)}\}$. The sample version of $\mathrm{DC}\{\mathbf{x}^{(T)}, \mathbf{y}^{(T)}\}$ is $\widehat{\mathrm{DC}}\{\mathbf{x}^{(T)}, \mathbf{y}^{(T)}\} = \widetilde{S}_1 + \widetilde{S}_2 - 2\widetilde{S}_3$, where

$$\begin{aligned}
\widetilde{S}_1 &= (n)_2^{-1} \sum_{i \neq j}^{n} \|\mathbf{x}_i^{(T)} - \mathbf{x}_j^{(T)}\| \|\mathbf{y}_i^{(T)} - \mathbf{y}_j^{(T)}\|, \\
\widetilde{S}_3 &= (n)_3^{-1} \sum_{i \neq j, j \neq k, k \neq i}^{n} \|\mathbf{x}_i^{(T)} - \mathbf{x}_j^{(T)}\| \|\mathbf{y}_i^{(T)} - \mathbf{y}_k^{(T)}\|, \\
\widetilde{S}_2 &= (n)_4^{-1} \sum_{(i,j,k,l)}^{n} \|\mathbf{x}_i^{(T)} - \mathbf{x}_j^{(T)}\| \|\mathbf{y}_k^{(T)} - \mathbf{y}_l^{(T)}\|.
\end{aligned}$$

By definition, $\mid (\sigma_{1,n}\sigma_{2,n})^{-1}\widehat{\mathrm{DC}}(\mathbf{x}, \mathbf{y}) - \widehat{\mathrm{DC}}\{\mathbf{x}^{(T)}, \mathbf{y}^{(T)}\} \mid$ is less than or equal to

$$\begin{aligned}
& Cn^{-2} \sum_{i \neq j}^{n} \mid \Delta_n(i,j,i,j) \mid + Cn^{-3} \sum_{i \neq j, j \neq k, k \neq i}^{n} \mid \Delta_n(i,j,i,k) \mid \\
& + Cn^{-4} \sum_{(i,j,k,l)}^{n} \mid \Delta_n(i,j,k,l) \mid \stackrel{\text{def}}{=} I_1 + I_2 + I_3,
\end{aligned}$$

where $\Delta_n(i,j,k,l) = \|\mathbf{x}_i/\sigma_{1,n} - \mathbf{x}_j/\sigma_{1,n}\| \|\mathbf{y}_k/\sigma_{2,n} - \mathbf{y}_l/\sigma_{2,n}\| - \|\mathbf{x}_i^{(T)} - \mathbf{x}_j^{(T)}\| \|\mathbf{y}_k^{(T)} - \mathbf{y}_l^{(T)}\|$. It is noted that there are four possible cases of the difference $\Delta_n(i,j,i,j)$. That is, $\Delta_n(i,j,i,j)$ equals

$$\begin{cases}
\frac{1}{\sigma_{1,n}\sigma_{2,n}} \|\mathbf{x}_i - \mathbf{x}_j\| \|\mathbf{y}_i - \mathbf{y}_j\|, & \text{if } \mathbf{x}_i, \mathbf{x}_j \sim F_{\mathbf{x}}, \mathbf{y}_i, \mathbf{y}_j \sim F_{\mathbf{y}}, \\
0, & \text{if } \mathbf{x}_i, \mathbf{x}_j \sim H_{\mathbf{x}}^{(n)}, \mathbf{y}_i, \mathbf{y}_j \sim H_{\mathbf{y}}^{(n)}, \\
\frac{1}{\sigma_{1,n}\sigma_{2,n}} (\|\mathbf{x}_i - \mathbf{x}_j\| \|\mathbf{y}_i - \mathbf{y}_j\| - \|\mathbf{x}_j\| \|\mathbf{y}_j\|), & \text{if } \mathbf{x}_i \sim F_{\mathbf{x}}, \mathbf{x}_j \sim H_{\mathbf{x}}^{(n)}, \mathbf{y}_i \sim F_{\mathbf{y}}, \mathbf{y}_j \sim H_{\mathbf{y}}^{(n)}, \\
\frac{1}{\sigma_{1,n}\sigma_{2,n}} (\|\mathbf{x}_i - \mathbf{x}_j\| \|\mathbf{y}_i - \mathbf{y}_j\| - \|\mathbf{x}_i\| \|\mathbf{y}_i\|), & \text{if } \mathbf{x}_j \sim F_{\mathbf{x}}, \mathbf{x}_i \sim H_{\mathbf{x}}^{(n)}, \mathbf{y}_j \sim F_{\mathbf{y}}, \mathbf{y}_i \sim H_{\mathbf{y}}^{(n)}.
\end{cases}$$

If $\mathbf{x} \sim F_{\mathbf{x}}$ and $\mathbf{y} \sim F_{\mathbf{y}}$, then $\mathbf{x}/\sigma_{1,n} = O_p(1)$ and $\mathbf{y}/\sigma_{2,n} = O_p(1)$ because $E(\|\mathbf{x}\| + \|\mathbf{y}\|) < \infty$. Further, If $\mathbf{x} \sim H_{\mathbf{x}}^{(n)}$ and $\mathbf{y} \sim H_{\mathbf{y}}^{(n)}$, we still have $\mathbf{x}/\sigma_{1,n} = O_p(1)$ and $\mathbf{y}/\sigma_{2,n} = O_p(1)$ because $H^{(n)}_{\mathbf{x}/\sigma_{1,n}} \sim (\mathbf{0}_p, \mathbf{\Sigma}_1)$ and $H^{(n)}_{\mathbf{y}/\sigma_{2,n}} \sim (\mathbf{0}_q, \mathbf{\Sigma}_2)$. On the basis of these observations, we can obtain $I_1 = O_p\{1/(\sigma_{1,n}\sigma_{2,n})\}$. In a similar fashion, we can obtain $I_2 = O_p\{1/(\sigma_{1,n}\sigma_{2,n})\}$ and $I_3 = O_p\{1/(\sigma_{1,n}\sigma_{2,n})\}$. As a result,

$$n(\sigma_{1,n}\sigma_{2,n})^{-1}\widehat{\mathrm{DC}}(\mathbf{x}, \mathbf{y}) - n\widehat{\mathrm{DC}}\{\mathbf{x}^{(T)}, \mathbf{y}^{(T)}\} = O_p\{n/(\sigma_{1,n}\sigma_{2,n})\}.$$

By the definitions of $\mathbf{x}^{(T)}$ and $\mathbf{y}^{(T)}$, we can show that $E\{\|\mathbf{x}^{(T)}\|^2\} = E(\eta^2)\mathrm{tr}(\mathbf{\Sigma}_1) < \infty$ and $E\{\|\mathbf{y}^{(T)}\|^2\} = E(\eta^2)\mathrm{tr}(\mathbf{\Sigma}_2) < \infty$ where $\eta \sim$ Bernoulli(1,ϵ). Since $\mathbf{x}^{(T)}$ is not independent of $\mathbf{y}^{(T)}$, $\widehat{\mathrm{DC}}\{\mathbf{x}^{(T)}, \mathbf{y}^{(T)}\}$ does not have degeneracy of order one with the finite variance of the kernel. Nevertheless, it is noted that $(\mathbf{x}^{(T)}, \mathbf{y}^{(T)}) \sim (1-\epsilon)\mathrm{pr}(\boldsymbol{\omega}_p = \mathbf{0}_p)\mathrm{pr}(\boldsymbol{\omega}_q = \mathbf{0}_q) + \epsilon H_{\mathbf{x}}^{(1)} H_{\mathbf{y}}^{(2)}$, where $\epsilon = O(n^{-1/2})$. By Le Cam's third lemma (van der Vaart , 2000), $n\widehat{\mathrm{DC}}\{\mathbf{x}^{(T)}, \mathbf{y}^{(T)}\}$ still converges in distribution to an infinite weighted sum of chi-square random variables, i.e., $\sum_{k=1}^{\infty} \lambda_k^{\natural}(Z_k^2 - 1)$, for nonnegative constants $\{\lambda_k^{\natural}\}_{k\geq 1}$ and normal random variables $\{Z_k^{\natural}\}_{k\geq 1}$. Again, combining the arguments mentioned above and the details for dealing with Theorem 1, we can also show that

$$n(\sigma_{1,n}\sigma_{2,n})^{-1}\widehat{\mathrm{DC}}(\mathbf{x}^b, \mathbf{y}^b) - n\widehat{\mathrm{DC}}\{\mathbf{x}^{(T)}, \mathbf{y}^{(T)}\} = O_p^*\{n/(\sigma_{1,n}\sigma_{2,n})\},$$

where O_p^* is defined under the permutation space. By choosing $\sigma_{1,n}$ and $\sigma_{2,n}$ such that $n/(\sigma_{1,n}\sigma_{2,n}) = o(1)$, and using Slutsky's theorem, both $n(\sigma_{1,n}\sigma_{2,n})^{-1}\widehat{\mathrm{DC}}(\mathbf{x}, \mathbf{y})$ and $n(\sigma_{1,n}\sigma_{2,n})^{-1}\,\widehat{\mathrm{DC}}(\mathbf{x}^b, \mathbf{y}^b)$ converge to $\sum_{k=1}^{\infty} \lambda_k^{\natural}(Z_k^2 - 1)$ as well. Thus, when $H_{(\mathbf{x},\mathbf{y})}^{(n)}$ favors H_0, the distance correlation test of asymptotic level-α becomes powerless.

□

S4. Proof of Theorem 3

Given $\mathcal{U}^{PC}(c)$, $\mathcal{U}^{DC}(c)$ and $\mathcal{U}^{mBKR}(c)$ arbitrarily, the family of alternative distributions includes some Gaussian distributions as a subset. Write $\mathbf{w} = (\mathbf{x}, \mathbf{y})$, $\mathrm{cov}(\mathbf{w}) = E\{(\mathbf{w} - E\mathbf{w})(\mathbf{w} - E\mathbf{w})^{\mathrm{T}}\}$ and consider that $\mathbf{w}$ is multivariate Gaussian with pairwise correlation ρ. Under the alternative, we have

$$\begin{aligned} |\mathrm{cov}(\mathbf{w})| &= (1-\rho)^{m-1}\{1+(m-1)\rho\}, \text{ and} \\ |2\mathrm{cov}^{-1}(\mathbf{w}) - \mathbf{I}_m| &= \{(1+\rho)(1-\rho)^{-1}\}^m[1 - 2m\rho\{(1+\rho)(1+(m-1)\rho)\}^{-1}], \end{aligned} \tag{S4.1}$$

where $m = p + q$ and $\mathbf{I}_m \in \mathbb{R}^{m\times m}$ is the identity matrix.

By the Le Cam's lemma (Baraud , 2002) and for sufficiently large n,

$$\inf_{\Phi_\alpha \in \mathcal{T}_\alpha} \sup_{(\mathbf{x},\mathbf{y})\in\mathcal{U}^{PC}(c_0)} E\{\Phi_\alpha \mid H_1\} \geq 1 - \alpha - 2^{-1}\{E(L_n^2 \mid H_0) - 1\}^{1/2},$$

$$\inf_{\Phi_\alpha \in \mathcal{T}_\alpha} \sup_{(\mathbf{x},\mathbf{y})\in\mathcal{U}^{mBKR}(c_0)} E\{\Phi_\alpha \mid H_1\} \geq 1 - \alpha - 2^{-1}\{E(L_n^2 \mid H_0) - 1\}^{1/2},$$

and

$$\inf_{\Phi_\alpha \in \mathcal{T}_\alpha} \sup_{(\mathbf{x},\mathbf{y}) \in \mathcal{U}^{DC}(c_0)} E\{\Phi_\alpha \mid H_1\} \geq 1 - \alpha - 2^{-1}\{E(L_n^2 \mid H_0) - 1\}^{1/2},$$

where $L_n = \prod_{i=1}^{n} d\{\mathrm{pr}(\mathbf{w}_i \mid H_1)\}/d\{\mathrm{pr}(\mathbf{w}_i \mid H_0)\}$ with $\mathbf{w}_i = (\mathbf{x}_i, \mathbf{y}_i)$. Therefore, as long as

$$E(L_n^2 \mid H_0) - 1 = o(1), \tag{S4.2}$$

we can prove the assertion. By definition and (S4.1),

$$\begin{aligned} E(L_n^2 \mid H_0) &= (|\mathbf{\Sigma}|)^{-n} E\left(\prod_{i=1}^{n} e^{\left[-2^{-1}\mathbf{w}_i^{\mathrm{T}}\left\{2\mathrm{cov}^{-1}(\mathbf{w}) - \mathbf{I}_m\right\}\mathbf{w}_i\right]}\right) \\ &= (1+\rho)^{-mn}\{1 - m\rho/(1+\rho)\}^{-n}. \end{aligned}$$

As $\rho \to 0$, $(1+\rho)^{-mn} = e^{-mn\rho}$, and $\{1-m\rho/(1+\rho)\}^{-n} = e^{mn\rho/(1+\rho)}$. Thus, as $\rho \to 0$, $E_0(L_n^2) = e^{-mn\rho^2/(1+\rho)}$, where we use the fact $-mn\rho + mn\rho/(1+\rho) = -mn\rho^2/(1+\rho)$. To get (S4.2), we need to show that as $\rho \to 0$,

$$\mathrm{PC}(\mathbf{x}, \mathbf{y}) \asymp \rho^2, \tag{S4.3}$$

$$\mathrm{mBKR}(\mathbf{x}, \mathbf{y}) \asymp \rho^2, \tag{S4.4}$$

and

$$\mathrm{DC}(\mathbf{x}, \mathbf{y}) \asymp \rho^2, \tag{S4.5}$$

where $a_n \asymp b_n$ means that there exist constants $C_2 \geq C_1 > 0$ such that $C_1|b_n| \leq |a_n| \leq C_2|b_n|$ for all sufficiently large n. We first deal with (S4.3). According to Zhu et al. (2017), we can show

$$\mathrm{PC}(\mathbf{x},\mathbf{y}) = (\gamma_p\gamma_q)^{-1}\int_{\boldsymbol{\alpha}\in\mathcal{S}^{p-1}}\int_{\boldsymbol{\beta}\in\mathcal{S}^{q-1}}(J_1+J_2-2J_3)d\boldsymbol{\alpha}d\boldsymbol{\beta}, \qquad \text{(S4.6)}$$

where

$$\begin{aligned}
J_1 &= \mathrm{pr}(W_1^{(1)}\geq 0, W_2^{(1)}\geq 0, W_3^{(1)}\geq 0, W_4^{(1)}\geq 0),\\
J_2 &= \mathrm{pr}(W_1^{(2)}\geq 0, W_2^{(2)}\geq 0, W_3^{(2)}\geq 0, W_4^{(2)}\geq 0),\\
J_3 &= \mathrm{pr}(W_1^{(3)}\geq 0, W_2^{(3)}\geq 0, W_3^{(3)}\geq 0, W_4^{(3)}\geq 0),\\
W_1^{(1)} &= \{2\boldsymbol{\alpha}^{\mathrm{T}}\mathrm{cov}(\mathbf{x})\boldsymbol{\alpha}\}^{-1/2}\boldsymbol{\alpha}^{\mathrm{T}}(\mathbf{x}_3-\mathbf{x}_1), W_2^{(1)} = \{2\boldsymbol{\alpha}^{\mathrm{T}}\mathrm{cov}(\mathbf{x})\boldsymbol{\alpha}\}^{-1/2}\boldsymbol{\alpha}^{\mathrm{T}}(\mathbf{x}_3-\mathbf{x}_2),\\
W_3^{(1)} &= \{2\boldsymbol{\beta}^{\mathrm{T}}\mathrm{cov}(\mathbf{y})\boldsymbol{\beta}\}^{-1/2}\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{y}_3-\mathbf{y}_1), W_4^{(1)} = \{2\boldsymbol{\beta}^{\mathrm{T}}\mathrm{cov}(\mathbf{y})\boldsymbol{\beta}\}^{-1/2}\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{y}_3-\mathbf{y}_2),\\
W_3^{(2)} &= \{2\boldsymbol{\beta}^{\mathrm{T}}\mathrm{cov}(\mathbf{y})\boldsymbol{\beta}\}^{-1/2}\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{y}_3-\mathbf{y}_4), W_4^{(2)} = \{2\boldsymbol{\beta}^{\mathrm{T}}\mathrm{cov}(\mathbf{y})\boldsymbol{\beta}\}^{-1/2}\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{y}_3-\mathbf{y}_5),\\
W_1^{(1)} &= W_1^{(2)} = W_1^{(3)}, W_2^{(1)} = W_2^{(2)} = W_2^{(3)}, W_3^{(3)} = W_3^{(1)},\\
W_4^{(3)} &= \{2\boldsymbol{\beta}^{\mathrm{T}}\mathrm{cov}(\mathbf{y})\boldsymbol{\beta}\}^{-1/2}\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{y}_3-\mathbf{y}_4).
\end{aligned}$$

We will deal with J_1, J_2 and J_3 by invoking the work by Cheng (1969). Write $a = 1/2$ and $b = \rho\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{1}_p\boldsymbol{\beta}^{\mathrm{T}}\mathbf{1}_q\{\boldsymbol{\alpha}^{\mathrm{T}}\mathrm{cov}(\mathbf{x})\boldsymbol{\alpha}\boldsymbol{\beta}^{\mathrm{T}}\mathrm{cov}(\mathbf{y})\boldsymbol{\beta}\})^{-1/2}$. It is apparent that $(W_1^{(k)},\cdots,W_4^{(k)})$, $k=1,2,3$, follow quadrivariate normal distribution with zero mean, unity variance and correlations $\rho_{rs}^{(k)} = \mathrm{cor}(W_r^{(k)}, W_s^{(k)})$, where

$$\rho_{12}^{(1)} = \rho_{34}^{(1)} = a, \rho_{13}^{(1)} = \rho_{24}^{(1)} = b, \rho_{23}^{(1)} = \rho_{14}^{(1)} = ab,$$

$$\rho_{12}^{(2)} = \rho_{34}^{(2)} = a, \rho_{13}^{(2)} = \rho_{24}^{(2)} = ab, \rho_{14}^{(2)} = \rho_{23}^{(1)} = ab, \text{ and}$$

$$\rho_{12}^{(2)} = \rho_{34}^{(2)} = a, \rho_{13}^{(1)} = b, \rho_{14}^{(1)} = \rho_{23}^{(1)} = \rho_{24}^{(1)} = ab.$$

Following the arguments used by Cheng (1969) to obtain the orthant probabilities of four normal variables with certain specific forms of correlation matrices, we can show

$$\begin{aligned}
J_1 &= 9^{-1} + (4\pi)^{-1}\{\arcsin(b) + \arcsin(b/2)\} + (4\pi^2)^{-1}\{\arcsin(b)^2 - \arcsin(b/2)^2\}, \\
J_2 &= 9^{-1} + (2\pi)^{-1}\arcsin(b/2) + \pi^{-2}\int_0^{\arcsin(b/2)} \arcsin[\sin(x)/\{1 + 2\cos(2x)\}]dx, \\
J_3 &= 9^{-1} + (8\pi)^{-1}\{\arcsin(b) + 3\arcsin(b/2)\} + (4\pi^2)^{-1}\Bigg[\\
&\qquad \int_0^{\arcsin(b/2)} \arcsin\big(\sin(x)\,[\{2\cos(2x)+3\}/\{2\cos(2x)+1\}]\big)dx \\
&\qquad -2\int_0^{\arcsin(b/2)} \arcsin\Big(\sin(x)\,[\{2\cos(2x)-1\}/\{6\cos(2x)+3\}]^{1/2}\Big)\,dx \\
&\qquad + \int_0^{\arcsin(b)} \arcsin\{\sin(x)/3\}dx\Bigg].
\end{aligned}$$

Under the alternative, $\text{cov}(\mathbf{x}) = (1-\rho)\mathbf{I}_p + \rho\mathbf{1}_p\mathbf{1}_p^{\mathrm{T}}$ and $\text{cov}(\mathbf{y}) = (1-\rho)\mathbf{I}_q + \rho\mathbf{1}_q\mathbf{1}_q^{\mathrm{T}}$. Thus, by the fact $\|\boldsymbol{\alpha}\| = \|\boldsymbol{\beta}\| = 1$,

$$b = \{1 + o(1)\}\rho\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{1}_p\boldsymbol{\beta}^{\mathrm{T}}\mathbf{1}_q,$$

as $\rho \to 0$. Using the fact $\arcsin(x) = \{1 + o(1)\}x$ as $x \to 0$, we can show

$$\arcsin(b) + \arcsin(b/2) + 2\arcsin(b/2) - \{\arcsin(b) + 3\arcsin(b/2)\} = 0,$$

$$\arcsin(b)^2 - \arcsin(b/2)^2 = \{3/4 + o(1)\}b^2,$$

$$\int_0^{\arcsin(b/2)} \arcsin[\sin(x)/\{1 + 2\cos(2x)\}]dx = \{1/24 + o(1)\}b^2,$$

$$\int_0^{\arcsin(b/2)} \arcsin\big(\sin(x)\,[\{2\cos(2x) + 3\}/\{2\cos(2x) + 1\}]\big)dx$$
$$-2\int_0^{\arcsin(b/2)} \arcsin\Big(\sin(x)\,[\{2\cos(2x) - 1\}/\{6\cos(2x) + 3\}]^{1/2}\Big)\,dx$$
$$+\int_0^{\arcsin(b)} \arcsin\{\sin(x)/3\}dx = \{7/24 + o(1)\}b^2,$$

as $\rho \to 0$. Therefore, as $\rho \to 0$, $J_1 + J_2 - 2J_3 = \{1 + o(1)\}\rho^2(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{1}_p\boldsymbol{\beta}^{\mathrm{T}}\mathbf{1}_q)^2/(12\pi^2)$, which indicates

$$\begin{aligned}\mathrm{PC}(\mathbf{x},\mathbf{y}) &= \{1/(12\pi^2) + o(1)\}\rho^2(\gamma_p\gamma_q)^{-1}\int_{\boldsymbol{\alpha}\in\mathcal{S}^{p-1}}\int_{\boldsymbol{\beta}\in\mathcal{S}^{q-1}}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{1}_p\boldsymbol{\beta}^{\mathrm{T}}\mathbf{1}_q)^2 d\boldsymbol{\alpha}d\boldsymbol{\beta},\\ &\asymp \rho^2,\end{aligned}$$

which yields (S4.3). Similar to (S4.6), we can show that

$$\mathrm{mBKR}(\mathbf{x},\mathbf{y}) = (\gamma_p\gamma_q)^{-1}\int_{\boldsymbol{\alpha}\in\mathcal{S}^{p-1}}\int_{\boldsymbol{\beta}\in\mathcal{S}^{q-1}}(J_1' + J_2' - 2J_3')d\boldsymbol{\alpha}d\boldsymbol{\beta},$$

where $J_1' = E\{I(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_1 \le \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_3)I(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_2 \le \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_3)I(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_1 \le \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_4)I(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_2 \le \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_4)\}$, $J_2' = E\{I(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_1 \le \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_3)I(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_2 \le \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_3)I(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_4 \le \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_6)I(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_5 \le \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_6)\}$ and $J_3' = E\{I(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_1 \le \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_3)I(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_2 \le \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_3)I(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_1 \le \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_5)I(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_4 \le \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_5)\}$. Using similar arguments to those in the derivation of J_1, J_2 and J_3, we can obtain (S4.4).

In what follows, we consider (S4.5). From the proof of equivalent expression of distance correlation (reported later), we can show

$$\begin{aligned}\mathrm{DC}(\mathbf{x},\mathbf{y}) &\asymp \{\mathrm{var}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})\mathrm{var}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y})\}^{1/2}\int_{\boldsymbol{\beta}\in\mathcal{S}^{q-1}}\int_{\boldsymbol{\alpha}\in\mathcal{S}^{p-1}}\\ &\quad \mathrm{DC}\{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}/\mathrm{var}^{1/2}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}),\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}/\mathrm{var}^{1/2}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y})\}d\mu(\boldsymbol{\alpha})d\mu(\boldsymbol{\beta}),\end{aligned}$$

where $\mu(\cdot)$ will be defined as in (S6.8). Clearly, $\mathrm{var}\{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}/\mathrm{var}^{1/2}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})\} = \mathrm{var}\{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}/\mathrm{var}^{1/2}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y})\} = 1$. Let $\Theta = \mathrm{cov}\{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}/\mathrm{var}^{1/2}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}),\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}/\mathrm{var}^{1/2}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y})\}$. According to Theorem 7 of Székely et al. (2007),

$$\begin{aligned}&\mathrm{DC}\{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}/\mathrm{var}^{1/2}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}),\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}/\mathrm{var}^{1/2}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y})\}\\ =\ & 4\pi(\Theta\arcsin\Theta+\{1-\Theta^2\}^{1/2}-\Theta\arcsin(\Theta/)-\{4-\Theta^2\}^{1/2}+1)\\ \asymp\ & \Theta^2.\end{aligned}$$

It is noted that $\mathrm{var}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})\mathrm{var}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}) = 1+o(1)$ and $\Theta = \{1+o(1)\}\rho\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{1}_p\boldsymbol{\beta}^{\mathrm{T}}\mathbf{1}_q$, as $\rho\to 0$. This, together with the preceding equation, yields (S4.5).

□

S5. Proof of Theorem 4

We start with proving the first statement. By definition, it suffices to show $\lim_{c\to\infty}\inf_{(\mathbf{x},\mathbf{y})\in\mathcal{U}^{PC}(c)}\mathrm{pr}\{n\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y}) \geq q^{PC}_{\alpha,n} \mid H_1\} = 1$, as $n\to\infty$. Define

$\widetilde{\mathrm{PC}}(\mathbf{x},\mathbf{y}) = \widetilde{T}_1 + \widetilde{T}_2 - 2\widetilde{T}_3$, where

$$
\begin{aligned}
\widetilde{T}_1 &= n^{-3} \sum_{i,j,k=1}^{n} \mathrm{ang}(\mathbf{x}_i - \mathbf{x}_k, \mathbf{x}_j - \mathbf{x}_k)\mathrm{ang}(\mathbf{y}_i - \mathbf{y}_k, \mathbf{y}_j - \mathbf{y}_k),\\
\widetilde{T}_2 &= n^{-5} \sum_{i,j,k,l,r=1}^{n} \mathrm{ang}(\mathbf{x}_i - \mathbf{x}_r, \mathbf{x}_j - \mathbf{x}_r)\mathrm{ang}(\mathbf{y}_k - \mathbf{y}_r, \mathbf{y}_l - \mathbf{y}_r),\\
\widetilde{T}_3 &= n^{-4} \sum_{i,j,k,l=1}^{n} \mathrm{ang}(\mathbf{x}_i - \mathbf{x}_l, \mathbf{x}_j - \mathbf{x}_l)\mathrm{ang}(\mathbf{y}_i - \mathbf{y}_l, \mathbf{y}_k - \mathbf{y}_l).
\end{aligned}
$$

Because of $0 \le \mathrm{ang}(\cdot,\cdot) \le \pi$,

$$\mathrm{pr}\{n\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y}) \ge q_{\alpha,n} \mid H_1\} = \{1 + o_p(1)\}\mathrm{pr}\{n\widetilde{\mathrm{PC}}(\mathbf{x},\mathbf{y}) \ge q^{PC}_{\alpha,n} \mid H_1\}.$$

From Theorem 1 of Zhu et al. (2017), $\widetilde{\mathrm{PC}}(\mathbf{x},\mathbf{y})$ equals

$$(n\gamma_p\gamma_q)^{-1} \sum_{i=1}^{n} \left[\int_{\boldsymbol{\alpha}\in\mathcal{S}^{p-1}} \int_{\boldsymbol{\alpha}\in\mathcal{S}^{q-1}} \left\{ \widehat{F}_{1,2}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_i, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_i) - \widehat{F}_1(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_i)\widehat{F}_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_i) \right\}^2 d\boldsymbol{\alpha} d\boldsymbol{\beta} \right],$$

where $\widehat{F}_{1,2}$, $\widehat{F}_1$ and $\widehat{F}_2$ stand for the empirical distributions of $(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y})$, $(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$ and $(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y})$, respectively. For convenience, let us denote by $F_{1,2}$, F_1 and F_2 the distributions of $(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y})$, $(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$ and $(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y})$, respectively. By the fact $\widehat{F}_{1,2}(s,t) - \widehat{F}_1(s)\widehat{F}_2(t) = \{F_{1,2}(s,t) - F_1(s)F_2(t)\} + \{\widehat{F}_{1,2}(s,t) - F_{1,2}(s,t)\} - \{\widehat{F}_1(s) - F_1(s)\}\{\widehat{F}_2(t) - F_2(t)\} - \{\widehat{F}_1(s) - F_1(s)\}\widehat{F}_2(t) - F_1(s)\{\widehat{F}_2(t) - F_2(t)\}$, and Minkowski's inequality, it holds

$$
\begin{aligned}
& \{\widehat{\mathrm{PC}}^{\natural}(\mathbf{x},\mathbf{y})\}^{1/2} \\
\le\ & \{\widetilde{\mathrm{PC}}(\mathbf{x},\mathbf{y})\}^{1/2} + 2\pi\{ \sup_{s,t,\boldsymbol{\alpha},\boldsymbol{\beta}} |\widehat{F}_{1,2}(s,t) - F_{1,2}(s,t)| + \sup_{s,\boldsymbol{\alpha}} |\widehat{F}_1(s) - F_1(s)| \\
& + \sup_{t,\boldsymbol{\beta}} |\widehat{F}_2(t) - F_2(t)|\} \qquad \text{(S5.7)}
\end{aligned}
$$

where $\widehat{\mathrm{PC}}^{\natural}(\mathbf{x},\mathbf{y}) = n^{-1}\sum_{i=1}^{n} N_i$ with

$$N_i \overset{\text{def}}{=} (\gamma_p\gamma_q)^{-1}\int_{\boldsymbol{\alpha}\in\mathcal{S}^{p-1}}\int_{\boldsymbol{\beta}\in\mathcal{S}^{q-1}}\left\{F_{1,2}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_i,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_i) - F_1(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_i)F_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_i)\right\}^2 d\boldsymbol{\alpha}d\boldsymbol{\beta}.$$

As a result,

$$\begin{aligned}
&\mathrm{pr}\{n\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y}) \geq q_{\alpha,n}^{PC} \mid H_1\} \geq \\
&1-\mathrm{pr}\Big[\sup_{s,t,\boldsymbol{\alpha},\boldsymbol{\beta}}|\widehat{F}_{1,2}(s,t)-F_{1,2}(s,t)| + \sup_{s,\boldsymbol{\alpha}}|\widehat{F}_1(s)-F_1(s)| \\
&+\sup_{t,\boldsymbol{\beta}}|\widehat{F}_2(t)-F_2(t)| \geq (\{\widehat{\mathrm{PC}}^{\natural}(\mathbf{x},\mathbf{y})\}^{1/2}-c_{\alpha,n})/(2\pi) \mid H_1\Big] \\
&\geq 1-\mathrm{pr}[\sup_{s,t,\boldsymbol{\alpha},\boldsymbol{\beta}}|\widehat{F}_{1,2}(s,t)-F_{1,2}(s,t)| \geq (\{\widehat{\mathrm{PC}}^{\natural}(\mathbf{x},\mathbf{y})\}^{1/2}-c_{\alpha,n})/(6\pi) \mid H_1] \\
&-\mathrm{pr}\{\sup_{s,\boldsymbol{\alpha}}|\widehat{F}_1(s)-F_1(s)| \geq \{\widehat{\mathrm{PC}}^{\natural}(\mathbf{x},\mathbf{y})-c_{\alpha,n}\}/(6\pi) \mid H_1\} - \mathrm{pr}[\sup_{t,\boldsymbol{\beta}}|\widehat{F}_2(t) \\
&-F_2(t)| \geq (\{\widehat{\mathrm{PC}}^{\natural}(\mathbf{x},\mathbf{y})\}^{1/2}-c_{\alpha,n})/(6\pi) \mid H_1],
\end{aligned}$$

where $c_{\alpha,n} = (q_{\alpha,n}^{PC})^{1/2}/n$. Define the event $\mathcal{A}$ by

$$\mathcal{A} = \left\{\left|\widehat{\mathrm{PC}}^{\natural}(\mathbf{x},\mathbf{y}) - \mathrm{PC}(\mathbf{x},\mathbf{y})\right| \leq 2^{-1}\mathrm{PC}(\mathbf{x},\mathbf{y})\right\}.$$

By noting $\mathrm{PC}(\mathbf{x},\mathbf{y}) \geq cn^{-1}$ and $\widehat{\mathrm{PC}}^{\natural}(\mathbf{x},\mathbf{y}) \geq \mathrm{PC}(\mathbf{x},\mathbf{y})/2$ under $\mathcal{A}$,

$$\begin{aligned}
&\mathrm{pr}\{\sup_{s,t,\boldsymbol{\alpha},\boldsymbol{\beta}}|\widehat{F}_{1,2}(s,t)-F_{1,2}(s,t)| \geq (\{\widehat{\mathrm{PC}}^{\natural}(\mathbf{x},\mathbf{y})\}^{1/2}-c_{\alpha,n})/(6\pi) \mid H_1\} \\
\leq\ &\mathrm{pr}\{\sup_{s,t,\boldsymbol{\alpha},\boldsymbol{\beta}}|\widehat{F}_{1,2}(s,t)-F_{1,2}(s,t)| \geq c^{1/2}n^{-1/2}/(6\pi 2^{1/2}) - c_{\alpha,n}/(6\pi) \mid H_1\} + \mathrm{pr}(\mathcal{A}^c \mid H_1).
\end{aligned}$$

Since $\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y})$ is a degenerate U-statistic with the bounded kernel under H_0, one can find an universal constant C_α such that $C_\alpha/n \geq c_{\alpha,n}^2$. Taking

$c \geq (2^{1/2}C_\alpha^{1/2}+c_0)^2$ with $c_0 \to \infty$ and using the generalization by Wolfowitz (1954) of the Glivenko-Cantelli theorem, we have

$$\lim_{c\to\infty} \sup_{(\mathbf{x},\mathbf{y})\in\mathcal{U}^{PC}(c)} \mathrm{pr}\{ \sup_{s,t,\boldsymbol{\alpha},\boldsymbol{\beta}} |\widehat{F}_{1,2}(s,t)-F_{1,2}(s,t)| \geq c^{1/2}n^{-1/2}/(6\pi 2^{1/2})-c_{\alpha,n}/(6\pi) \mid H_1\} = 0,$$

as $n \to \infty$. On the other hand, an application of Chebyshev's inequality yields

$$\begin{aligned} \mathrm{pr}(\mathcal{A}^c \mid H_1) &\leq 4\{\mathrm{PC}(\mathbf{x},\mathbf{y})\}^{-2}n^{-1}\mathrm{var}(N_1) \\ &\leq n^{-1}\pi^2 E(N_1)\{\mathrm{PC}(\mathbf{x},\mathbf{y})\}^{-2} = \pi^2 n^{-1}\{\mathrm{PC}(\mathbf{x},\mathbf{y})\}^{-1} \\ &\leq \pi^2/c, \end{aligned}$$

where we use the fact $N_1 \leq \pi^2/4$ and $E(N_1) = \mathrm{PC}(\mathbf{x},\mathbf{y})$. This further gives

$$\lim_{c\to\infty} \sup_{(\mathbf{x},\mathbf{y})\in\mathcal{U}^{PC}(c)} \mathrm{pr}(\mathcal{A}^c \mid H_1) = 0.$$

Consequently,

$$\lim_{c\to\infty} \sup_{(\mathbf{x},\mathbf{y})\in\mathcal{U}^{PC}(c)} \mathrm{pr}\{ \sup_{s,t,\boldsymbol{\alpha},\boldsymbol{\beta}} |\widehat{F}_{1,2}(s,t) - F_{1,2}(s,t)| \geq (\{\widehat{\mathrm{PC}}^{\natural}(\mathbf{x},\mathbf{y})\}^{1/2} - c_{\alpha,n})/(6\pi) \mid H_1\} = 0.$$

Similarly, we can obtain

$$\begin{aligned} &\lim_{c\to\infty} \sup_{(\mathbf{x},\mathbf{y})\in\mathcal{U}^{PC}(c)} \mathrm{pr}\{\sup_{s,\boldsymbol{\alpha}} |\widehat{F}_1(s) - F_1(s)| \geq (\{\widehat{\mathrm{PC}}^{\natural}(\mathbf{x},\mathbf{y})\}^{1/2} - c_{\alpha,n})/(6\pi) \mid H_1\} = 0, \text{ and} \\ &\lim_{c\to\infty} \sup_{(\mathbf{x},\mathbf{y})\in\mathcal{U}^{PC}(c)} \mathrm{pr}\{\sup_{t,\boldsymbol{\beta}} |\widehat{F}_2(t) - F_2(t)| \geq (\{\widehat{\mathrm{PC}}^{\natural}(\mathbf{x},\mathbf{y})\}^{1/2} - c_{\alpha,n})/(6\pi) \mid H_1\} = 0. \end{aligned}$$

A summation of these results yields $\lim_{c\to\infty} \inf_{(\mathbf{x},\mathbf{y})\in\mathcal{U}^{PC}(c)} \mathrm{pr}\{\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y}) \geq q_{\alpha,n} \mid H_1\} \geq 1 - o(1)$. In analogy with the notation of $\widetilde{\mathrm{PC}}(\mathbf{x},\mathbf{y})$, we write

$$\widetilde{\mathrm{mBKR}}(\mathbf{x},\mathbf{y}) = \widetilde{W}_1 + \widetilde{W}_2 - 2\widetilde{W}_3,$$

where

$$\begin{aligned}
\widetilde{W}_1 &= n^{-4} \sum_{i,j,k,l=1}^{n} \mathrm{ang}(\mathbf{x}_i - \mathbf{x}_k, \mathbf{x}_j - \mathbf{x}_k)\mathrm{ang}(\mathbf{y}_i - \mathbf{y}_k, \mathbf{y}_j - \mathbf{y}_l),\\
\widetilde{W}_2 &= n^{-6} \sum_{i,j,k,l,r,s=1}^{n} \mathrm{ang}(\mathbf{x}_i - \mathbf{x}_r, \mathbf{x}_j - \mathbf{x}_r)\mathrm{ang}(\mathbf{y}_k - \mathbf{y}_s, \mathbf{y}_l - \mathbf{y}_s),\\
\widetilde{W}_3 &= n^{-5} \sum_{i,j,k,l,r=1}^{n} \mathrm{ang}(\mathbf{x}_i - \mathbf{x}_l, \mathbf{x}_j - \mathbf{x}_l)\mathrm{ang}(\mathbf{y}_i - \mathbf{y}_r, \mathbf{y}_k - \mathbf{y}_r).
\end{aligned}$$

Employing arguments similar to those for proving Theorem 1 of Zhu et al. (2017), $\widetilde{\mathrm{mBKR}}(\mathbf{x},\mathbf{y})$ also has the elegant V-statistic representation. That is, $\widetilde{\mathrm{mBKR}}(\mathbf{x},\mathbf{y})$ is further equal to

$$(n\gamma_p\gamma_q)^{-1} \sum_{i=1}^{n}\sum_{j=1}^{n} \Big[\int_{\boldsymbol{\alpha}\in\mathcal{S}^{p-1}} \int_{\boldsymbol{\alpha}\in\mathcal{S}^{q-1}} \big\{\widehat{F}_{1,2}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_i, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_j) - \widehat{F}_1(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_i)\widehat{F}_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_j)\big\}^2 d\boldsymbol{\alpha} d\boldsymbol{\beta}\Big].$$

By the standard theory of U- and V-statistics, apply techniques parallel to those used in the proof of $\lim_{c\to\infty} \inf_{(\mathbf{x},\mathbf{y})\in\mathcal{U}^{PC}(c)} \mathrm{pr}\{\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y}) \geq q_{\alpha,n} \mid H_1\} \geq 1 - o(1)$ to complete the proof of the first assertion.

On the other hand, let us define $\widetilde{\mathrm{DC}}(\mathbf{x},\mathbf{y}) = \widetilde{R}_1 + \widetilde{R}_2 - 2\widetilde{R}_3$, where

$$\begin{aligned}
\widetilde{R}_1 &= n^{-2}\sum_{i,j=1}^{n} \|\mathbf{x}_i - \mathbf{x}_j\|\|\mathbf{y}_i - \mathbf{y}_j\|, \\
\widetilde{R}_2 &= n^{-4}\sum_{i,j,k,l,=1}^{n} \|\mathbf{x}_i - \mathbf{x}_j\|\|\mathbf{y}_k - \mathbf{y}_l\|, \\
\widetilde{R}_3 &= n^{-3}\sum_{i,j,k=1}^{n} \|\mathbf{x}_i - \mathbf{x}_j\|\|\mathbf{y}_i - \mathbf{y}_k\|.
\end{aligned}$$

By the standard theory of U and V-statistics,, it suffices to show $\lim_{c\to\infty} \inf_{(\mathbf{x},\mathbf{y})\in\mathcal{U}^{DC}(c)}$ $\mathrm{pr}\{n\widetilde{\mathrm{DC}}(\mathbf{x},\mathbf{y}) \geq q_{\alpha,n}^{DC} \mid H_1\} = 1$, as $n\to\infty$. Invoking the following proof for equivalent expression of distance correlation, we can also show that

$$\widetilde{\mathrm{DC}}(\mathbf{x},\mathbf{y}) = \{\gamma_p'\gamma_q'/(\pi^2\gamma_p\gamma_q)\}\int_{\boldsymbol{\beta}\in\mathcal{S}^{q-1}}\int_{\boldsymbol{\alpha}\in\mathcal{S}^{p-1}}\int_{\mathbb{R}^1}\int_{\mathbb{R}^1}\{\widehat{F}_{1,2}(s,t) - \widehat{F}_1(s)\widehat{F}_2(t)\}^2 ds dt d\boldsymbol{\alpha} d\boldsymbol{\beta},$$

where $\gamma_p', \gamma_q', \gamma_p$ and γ_q will be defined as in (S6.8). This, together with Minkowski's inequality and (S6.8), yields

$$\mathrm{DC}(\mathbf{x},\mathbf{y}) \leq \widetilde{\mathrm{DC}}^{1/2}(\mathbf{x},\mathbf{y}) + \left\{\{\gamma_p'\gamma_q'/(\pi^2\gamma_p\gamma_q)\}\int_{\boldsymbol{\beta}\in\mathcal{S}^{q-1}}\int_{\boldsymbol{\alpha}\in\mathcal{S}^{p-1}}\int_{\mathbb{R}^1}\int_{\mathbb{R}^1}\Delta^2(\boldsymbol{\alpha},\boldsymbol{\beta},s,t) ds dt d\boldsymbol{\alpha} d\boldsymbol{\beta}\right\}^{1/2},$$

where $\Delta(\boldsymbol{\alpha},\boldsymbol{\beta},s,t) = \mid \widehat{F}_{1,2}(s,t) - \widehat{F}_1(s)\widehat{F}_2(t) - \{F_{1,2}(s,t) - F_1(s)\ F_2(t)\} \mid$.

By the boundedness of $\widehat{F}_1(\cdot)$, $\widehat{F}_2(\cdot)$, $F_1(\cdot)$ and $F_2(\cdot)$, it is straightforward to show that

$$\begin{aligned}
\Delta(\boldsymbol{\alpha},\boldsymbol{\beta},s,t) \leq\ & \mid \widehat{F}_{1,2}(s,t) - F_{1,2}(s,t) \mid + 2 \mid \widehat{F}_1(s) - F_1(s) \mid \\
& + 2 \mid \widehat{F}_2(t) - F_2(t) \mid .
\end{aligned}$$

Since $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$, we have

$$\begin{aligned} & \sup_{\boldsymbol{\alpha},\boldsymbol{\beta},s,t} \Delta^2(\boldsymbol{\alpha},\boldsymbol{\beta},s,t) \\ \leq\ & 3 \sup_{\boldsymbol{\alpha},\boldsymbol{\beta},s,t} \mid \widehat{F}_{1,2}(s,t) - F_{1,2}(s,t) \mid^2 + 12 \sup_{\boldsymbol{\alpha},\boldsymbol{\beta},s,t} \mid \widehat{F}_1(s) - F_1(s) \mid^2 \\ & + 12 \sup_{\boldsymbol{\alpha},\boldsymbol{\beta},s,t} \mid \widehat{F}_2(t) - F_2(t) \mid^2 . \end{aligned}$$

Since the supports of $\mathbf{x}$ and $\mathbf{y}$ are bounded, i.e., $\max(\|\mathbf{x}\|, \|\mathbf{y}\|) < C$ for some positive constant $C > 0$, and $\|\boldsymbol{\alpha}\| = \|\boldsymbol{\beta}\| = 1$, we have $\max(\mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x} \mid, \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y} \mid) \leq C^{1/2}$, which implies $\max(\Omega_{\boldsymbol{\alpha}}) - \min(\Omega_{\boldsymbol{\alpha}}) \leq 2C^{1/2}$ and $\max(\Omega_{\boldsymbol{\beta}}) - \min(\Omega_{\boldsymbol{\beta}}) \leq 2C^{1/2}$. Here $\Omega_{\boldsymbol{\alpha}}$ and $\Omega_{\boldsymbol{\beta}}$ are the supports of $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}$ and $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}$, respectively. That is, $\int_{\Omega_{\boldsymbol{\alpha}}} \int_{\Omega_{\boldsymbol{\beta}}} 1 ds dt \leq 4C < \infty$. Clearly, $\int_{\boldsymbol{\beta} \in \mathcal{S}^{q-1}} \int_{\boldsymbol{\alpha} \in \mathcal{S}^{p-1}} 1 d\mu(\boldsymbol{\alpha}) d\mu(\boldsymbol{\beta}) = 1 < \infty$. In other words, there exist a positive constant C_1 such that

$$\begin{aligned} & \mathrm{DC}(\mathbf{x},\mathbf{y}) \\ \leq\ & \widetilde{\mathrm{DC}}^{1/2}(\mathbf{x},\mathbf{y}) + C_1 \sup_{\boldsymbol{\alpha},\boldsymbol{\beta},s,t} \mid \widehat{F}_{1,2}(s,t) - F_{1,2}(s,t) \mid + C_1 \sup_{\boldsymbol{\alpha},s,t} \mid \widehat{F}_1(s) - F_1(s) \mid \\ & + C_1 \sup_{\boldsymbol{\beta},s,t} \mid \widehat{F}_2(t) - F_2(t) \mid . \end{aligned}$$

Apply arguments similar to those for dealing with (S5.7) to complete the proof. □

S6. Equivalent Expression of Distance Correlation Given in (1.2)

Simple algebraic calculation yields that

$$\begin{aligned} \mid a-b \mid\mid c-d \mid &= \int_{-\infty}^{+\infty} \Big\{ I(a \leq s < b) I(c \leq s < d) + I(a \leq s < b) I(d \leq s < c) \\ &+ I(b \leq s < a) I(c \leq s < d) + I(b \leq s < a) I(d \leq s < c) \Big\} dsdt. \end{aligned}$$

This, together with Fubini's theorem, entails that

$$\begin{aligned} & E \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_1 - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_2 \mid\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_1 - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_2 \mid = 4 \int F_{1,2}^2(s,t) dsdt + 2 \int F_{1,2}(s,t) dsdt \\ + & \; 2 \int F_1(s) F_2(t) dsdt - 4 \int F_1(s) F_{1,2}(s,t) dsdt - 4 \int F_2(t) F_{1,2}(s,t) dsdt, \end{aligned}$$

Following similar arguments, we can show that

$$\begin{aligned} & E \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_1 - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_2 \mid\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_1 - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_3 \mid = 4 \int F_1(s) F_2(t) F_{1,2}^2(s,t) dsdt \\ - & \; 2 \int F_1(s) F_{1,2}(s,t) dsdt - 2 \int F_1^2(s) F_2(t) dsdt - 2 \int F_2(t) F_{1,2}(s,t) dsdt \\ - & \; 2 \int F_1(s) F_2(t)^2 dsdt + \int F_{1,2}(s,t) dsdt + 3 \int F_1(s) F_2(t) dsdt, \end{aligned}$$

and

$$\begin{aligned} & E \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_1 - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_2 \mid E \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_1 - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_2 \mid = 4 \int F_1(s) F_2(t) dsdt - 4 \int F_1(s) F_2^2(t) dsdt \\ & -4 \int F_1^2(s) F_2(t) dsdt - 4 \int F_2(t) F_{1,2}(s,t) dsdt + 4 \int F_1^2(s) F_2^2(t) dsdt. \end{aligned}$$

Combining the three above results, we obtain that $E \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_1 - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_2 \mid\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_1 - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_2 \mid + E \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_1 - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_2 \mid E \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_1 - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_2 \mid -2E \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_1 - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_2 \mid\mid$

$\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_1 - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_3 \mid$ equals

$$4\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{F_{1,2}(s,t) - F_1(s)F_2(t)\}^2 dsdt.$$

In addition,

$$\int_{\boldsymbol{\alpha}\in\mathcal{S}^{p-1}} \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x} \mid d\mu(\boldsymbol{\alpha}) = \|\mathbf{x}\|/\gamma_p' \text{ and } \int_{\boldsymbol{\beta}\in\mathcal{S}^{q-1}} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y} \mid d\mu(\boldsymbol{\beta}) = \|\mathbf{y}\|/\gamma_q',$$

where $\gamma_p' = \sqrt{\pi}(p-1)\Gamma\{(p-1)/2\}/\{2\Gamma(p/2)\}$ and μ is the uniform distribution on the surface of the unit sphere. Thus, $\mathrm{DC}(\mathbf{x},\mathbf{y})$ can be re-expressed equivalently as

$$\begin{aligned}
&\gamma_p'\gamma_q' \int_{\boldsymbol{\beta}\in\mathcal{S}^{q-1}}\int_{\boldsymbol{\alpha}\in\mathcal{S}^{p-1}} \Big\{E(\mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_1 - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_2 \mid\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_1 - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_2 \mid) + E(\mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_1 - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_2 \mid) \\
&\qquad E(\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_1 - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_2 \mid) - 2E(\mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_1 - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_2 \mid\mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_1 - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}_3 \mid)\Big\} d\mu(\boldsymbol{\alpha})d\mu(\boldsymbol{\beta}) \\
= \ & \{\gamma_p'\gamma_q'/(\pi^2\gamma_p\gamma_q)\}\int_{\boldsymbol{\beta}\in\mathcal{S}^{q-1}}\int_{\boldsymbol{\alpha}\in\mathcal{S}^{p-1}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\{F_{1,2}(s,t) - F_1(s)F_2(t)\}^2 dsdtd\boldsymbol{\alpha}d\boldsymbol{\beta}.
\end{aligned} \tag{S6.8}$$

This completes the proof. □

References

Baraud, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8, 577-606.

Cheng, M. (1969). The orthant probabilities of four Gaussian variates. *Annals of Mathematical Statistics*, 40, 152-161.

Leucht, A. and Neumann, M. (2009). Consistency of general bootstrap methods for degenerate U-type and V-type statistics. *Journal of Multivariate Analysis*, 100, 1622-1633.

Serfling, R. L. (1980). *Approximation Theorems in Mathematical Statistics*. New York, Wiley.

van der Vaart, A. (2000). *Asymptotic statistics*. Cambridge University Press.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

Székely, G., Rizzo, M. and Bakirov, N. (2007). Measuring and testing dependence by correlation of distances. *Annals of Statistics*, 35, 2769-2794.

Wolfowitz, J. (1954). Generalization of the theorem of Glivenko-Cantelli. *Annals of Mathematical Statistics*, 25, 131-138.

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