Supplementary material

A. TECHNICAL DETAILS

The Generalized Banach-Mazur theorem is a theorem of functional analysis, which states that most well-behaved metric space can be subspaces of the space of continuous paths. It was given by Banach (1932) and Kleiber and Pervin (1969).

Lemma A.1 (Banach-Mazur theorem) *A separable metric space is isometric to a subspace of C_1, the space of all real-valued continuous functions on the closed unit interval with the supmetric.*

According to this lemma, independence-zero equivalence property of Ball Covariance on the metric space can be studied in Banach spaces since the isometric mapping preserves all of the topological properties of the original space. To prove Theorem 2.1.1, we need the following Lemma which is analogous to Corollary 5.8.2 in Bogachev (2007) to complete the proof of case (a). The proof of this Lemma depends on the covering theorem (See Theorem 3.1 in Jackson and Mauldin (1999)).

Lemma A.2 *Let θ be a probability measure on a finite dimensional Banach space* Q*, $\mathcal{C}$ a collection of non-degenerate closed balls, and Q_C the set of their centers such that for every $v \in Q_C$ and every $\varepsilon > 0$, $\mathcal{C}$ contains a ball $\bar{B}(v, r)$ with $r < \varepsilon$. Then, for every nonempty open set $Q_0 \subset Q$, there is at most a countable collection of disjoint balls $\bar{B}_j \in \mathcal{C}$ such that*

$$\bigcup_{j=1}^{\infty} \bar{B}_j \subset Q_0 \quad \textit{and} \quad \theta((Q_C \cap Q_0) \backslash \bigcup_{j=1}^{\infty} \bar{B}_j) = 0.$$

Proof of Theorem 2.1.1: Without loss of generality, we prove the results only for $\omega_1 = \omega_2 \equiv 1$.

Case (a): Let $\mathcal{F}_{\mathscr{X}}$, $\mathcal{F}_{\mathscr{Y}}$ be the Sigma algebras on $\mathscr{X}$ and $\mathscr{Y}$, and $\mathcal{C}_{\mathscr{X}}, \mathcal{C}_{\mathscr{Y}}$ be the collections of open sets on $\mathscr{X}$ and $\mathscr{Y}$, respectively. In order to prove that $\mathbf{BCov}(X, Y) = 0$ implies $\theta = \mu \otimes \nu$, we need to prove that $\theta(A \times B) = \mu \otimes \nu(A \times B)$ for $A \in \mathcal{F}_{\mathscr{X}}$,$B \in \mathcal{F}_{\mathscr{Y}}$. We complete this proof by three steps:

Step 1: $A = \bar{B}_\rho(x_1, x_2)$ for $x_1 \in S_\mu, x_2 \in \mathscr{X}$ and $B = \bar{B}_\zeta(y_1, y_2)$ for $y_1 \in S_\nu, y_2 \in \mathscr{Y}$.

For $(x_1, y_1), (x_2, y_2) \in S_\theta$, we have $\theta(\bar{B}_\rho(x_1, x_2) \times \bar{B}_\zeta(y_1, y_2)) = \mu(\bar{B}_\rho(x_1, x_2)) \times \nu(\bar{B}_\zeta(y_1, y_2))$ when $\mathbf{BCov}(X, Y) = 0$. Because $S_\theta = S_\mu \times S_\nu$, we have $\theta(\bar{B}_\rho(x_1, x_2) \times \bar{B}_\zeta(y_1, y_2)) = \mu(\bar{B}_\rho(x_1, x_2)) \times \nu(\bar{B}_\zeta(y_1, y_2))$ for any $x_1, x_2 \in S_\mu$ and $y_1, y_2 \in S_\nu$.

Next, let $r_\mu = \sup\{\rho(x_1, x') : x' \in S_\mu \cap \bar{B}_\rho(x_1, x_2)\}$. Since $S_\mu \cap \bar{B}_\rho(x_1, x_2)$ is a closed set and $\rho(x_1, x')$ is a continuous function, there exists $x_\mu \in S_\mu \cap \bar{B}_\rho(x_1, x_2)$ such that $r_\mu = \rho(x_1, x_\mu)$. Thus, we have

$$\theta(\bar{B}_\rho(x_1, x_\mu) \times \bar{B}_\zeta(y_1, y_2)) = \mu(\bar{B}_\rho(x_1, x_\mu)) \times \nu(\bar{B}_\zeta(y_1, y_2)).$$

Since

$$\begin{aligned} &\bar{B}_\rho(x_1, x_2) \times \bar{B}_\zeta(y_1, y_2) \\ =&(\bar{B}_\rho(x_1, x_\mu) \times \bar{B}_\zeta(y_1, y_2)) \cup ((\bar{B}_\rho(x_1, x_2) \backslash \bar{B}_\rho(x_1, x_\mu)) \times \bar{B}_\zeta(y_1, y_2)) \end{aligned}$$

and

$$0 \le \theta((\bar{B}_\rho(x_1, x_2) \backslash \bar{B}_\rho(x_1, x_\mu)) \times \bar{B}_\zeta(y_1, y_2)) \le \mu(\bar{B}_\rho(x_1, x_2) \backslash \bar{B}_\rho(x_1, x_\mu)) = 0,$$

we obtain that

$$\begin{aligned} \theta(\bar{B}_\rho(x_1, x_2) \times \bar{B}_\zeta(y_1, y_2)) &= \theta(\bar{B}_\rho(x_1, x_\mu) \times \bar{B}_\zeta(y_1, y_2)) \\ &= \mu(\bar{B}_\rho(x_1, x_\mu))\nu(\bar{B}_\zeta(y_1, y_2)) \\ &= \mu(\bar{B}_\rho(x_1, x_2))\nu(\bar{B}_\zeta(y_1, y_2)). \end{aligned}$$

Thus, for $x_1 \in S_\mu, x_2 \in \mathscr{X}, y_1, y_2 \in S_\nu$, we have

$$\theta(\bar{B}_\rho(x_1, x_2) \times \bar{B}_\zeta(y_2, y_2)) = \mu(\bar{B}_\rho(x_1, x_2))\nu(\bar{B}_\zeta(y_1, y_2)). \tag{A.1}$$

Repeat the above steps for ν, (A.1) also holds for $x_1 \in S_\mu, x_2 \in \mathscr{X}, y_1 \in S_\nu, y_2 \in \mathscr{Y}$.

Step 2: $A \in \mathcal{C}_{\mathscr{X}}$ and $B \in \mathcal{C}_{\mathscr{Y}}$.

For any $A \in \mathcal{C}_{\mathscr{X}}$, according to Lemma A.2, there exist as many as countable disjoint closed balls $\bar{B}_j^{\mathscr{X}}$ with centers located in S_μ, such that

$$\bigcup_{j=1}^{\infty} \bar{B}_j^{\mathscr{X}} \subset A, \quad \mu(S_\mu \cap A \backslash \bigcup_{j=1}^{\infty} \bar{B}_j^{\mathscr{X}}) = 0.$$

By the results in Step 2, we have

$$\begin{aligned}
\theta(\bigcup_{j=1}^{\infty} \bar{B}_j^{\mathscr{X}} \times \bar{B}_\zeta(y_1, y_2)) &= \sum_{j=1}^{\infty} \theta(\bar{B}_j^{\mathscr{X}} \times \bar{B}_\zeta(y_1, y_2)) \\
&= \sum_{j=1}^{\infty} \mu(\bar{B}_j^{\mathscr{X}})\nu(\bar{B}_\zeta(y_1, y_2)) \\
&= \mu(\bigcup_{j=1}^{\infty} \bar{B}_j^{\mathscr{X}})\nu(\bar{B}_\zeta(y_1, y_2)).
\end{aligned}$$

Thus, we obtain that

$$\begin{aligned}
\theta(A \times \bar{B}_\zeta(y_1, y_2)) &= \theta((A \cap S_\mu) \times \bar{B}_\zeta(y_1, y_2)) \\
&= \theta(\bigcup_{j=1}^{\infty} \bar{B}_j^{\mathscr{X}} \times \bar{B}_\zeta(y_1, y_2)) + \theta((A \cap S_\mu \backslash \bigcup_{j=1}^{\infty} \bar{B}_j^{\mathscr{X}}) \times \bar{B}_\zeta(y_1, y_2)) \\
&= \theta(\bigcup_{j=1}^{\infty} \bar{B}_j^{\mathscr{X}} \times \bar{B}_\zeta(y_1, y_2)) \\
&= \mu(\bigcup_{j=1}^{\infty} \bar{B}_j^{\mathscr{X}})\nu(\bar{B}_\zeta(y_1, y_2)) \\
&= \mu(\bigcup_{j=1}^{\infty} \bar{B}_j^{\mathscr{X}})\nu(\bar{B}_\zeta(y_1, y_2)) + \mu(S_\mu \cap A \backslash \bigcup_{j=1}^{\infty} \bar{B}_j^{\mathscr{X}})\nu(\bar{B}_\zeta(y_1, y_2)) \\
&= \mu(S_\mu \cap A)\nu(\bar{B}_\zeta(y_1, y_2)) \\
&= \mu(A)\nu(\bar{B}_\zeta(y_1, y_2)).
\end{aligned}$$

Therefore, for any $A \in \mathcal{C}_{\mathscr{X}}, y_1 \in S_\nu, y_2 \in \mathscr{Y}$, we have

$$\theta(A \times \bar{B}_\zeta(y_1, y_2)) = \mu(A)\nu(\bar{B}_\zeta(y_1, y_2)).$$

Repeat the above steps for ν, we obtain that $\theta(A \times B) = \mu(A)\nu(B)$ holds for all $A \in \mathcal{C}_{\mathscr{X}}$ and $B \in \mathcal{C}_{\mathscr{Y}}$.

Step 3: $A \in \mathcal{F}_{\mathscr{X}}$ and $B \in \mathcal{F}_{\mathscr{Y}}$.

Let

$$M(\mathcal{C}_{\mathscr{X}}) := \{A \in \mathcal{F}_{\mathscr{X}} : \theta(A \times B) = \mu(A)\nu(B), B \in \mathcal{C}_{\mathscr{Y}}\}.$$

Since $M(\mathcal{C}_{\mathscr{X}})$ is a Dynkin system and $\mathcal{C}_{\mathscr{X}}$ is a π-system, we have $M(\mathcal{C}_{\mathscr{X}}) \supseteq \mathcal{F}_{\mathscr{X}}$. This means that $\theta(A \times B) = \mu(A)\nu(B)$ holds for all $A \in \mathcal{F}_{\mathscr{X}}, B \in \mathcal{C}_{\mathscr{Y}}$.

Similarly, let

$$M(\mathcal{C}_{\mathscr{Y}}) := \{B \in \mathcal{F}_{\mathscr{Y}} : \theta(A \times B) = \mu(A)\nu(B), A \in \mathcal{F}_{\mathscr{X}}\}.$$

Since $M(\mathcal{C}_{\mathscr{Y}})$ is a Dynkin system and $\mathcal{C}_{\mathscr{Y}}$ is a π-system, $M(\mathcal{C}_{\mathscr{Y}}) \supseteq \mathcal{F}_{\mathscr{Y}}$, which means that $\theta(A \times B) = \mu(A)\nu(B)$ holds for all $A \in \mathcal{F}_{\mathscr{X}}$, $B \in \mathcal{F}_{\mathscr{Y}}$. Therefore, we have $\theta = \mu \otimes \nu$.

Case (b): Let $P_{X,Y}(x, y)$ be the function of the discrete measures $\theta((X, Y) = (x, y))$ and $P_X(x), P_Y(y)$ be the discrete measures of the corresponding marginal distribution when (x, y) is the discrete point of θ. Similarly, let $h(x, y)$ be the Radon-Nikodym derivative of θ with respect to Gaussian measure and $f(x), g(y)$ be the Radon-Nikodym derivatives of the corresponding marginal distribution when (x, y) is the absolutely continuous point of θ.

If $\theta \neq \mu \otimes \nu$, then there exists one point $(x, y) \in \mathscr{X} \times \mathscr{Y}$ such that $P_{X,Y}(x, y) \neq P_X(x)P_Y(y)$ or $h(x, y) \neq f(x)g(y)$. First, we will show that this point (x, y) can be chosen in S_θ. If (x, y) is the discrete point of θ and $(x, y) \in S_\theta^c$, which means that $0 = P_{X,Y}(x, y) < P_X(x)P_Y(y)$, then we have $(x, y) \in S_\mu \times S_\nu \backslash S_\theta$. Denote G as the Gaussian measure. By the equality of

$$\begin{aligned}
&\sum_{(x,y)\in S_\mu\times S_\nu} P_{X,Y}(x,y) + \int_{S_\mu\times S_\nu} h(x,y)G(dx)G(dy) \\
=&1 \\
=&\sum_{(x,y)\in S_\theta} P_X(x)P_Y(y) + \sum_{(x,y)\in S_\mu\times S_\nu\setminus S_\theta} P_X(x)P_Y(y) \\
&+\int_{S_\theta} f(x)g(y)G(dx)G(dy) + \int_{S_\mu\times S_\nu\setminus S_\theta} f(x)g(y)G(dx)G(dy),
\end{aligned}$$

there exists another point $(x^*, y^*) \in S_\theta$ such that $P_{X,Y}(x^*, y^*) > P_X(x^*)P_Y(y^*)$ or $h(x^*, y^*) > f(x^*)g(y^*)$. Similar discussions can be done for (x, y) being absolutely case. Thus, we can assume that $P_{X,Y}(x, y) > P_X(x)P_Y(y)$ or $h(x, y) > f(x)g(y)$.

We first consider the case where (x, y) is the discrete point of θ,

$$\begin{aligned}
&\mathbf{BCov}^2(X,Y) \\
\geq &\sum_{\{(x_1,y_1):P_{X,Y}(x_1,y_1)>0\}} \sum_{\{(x_2,y_2):P_{X,Y}(x_2,y_2)>0\}} \\
&(\theta-\mu\otimes\nu)^2(\bar{B}_\rho(x_1,x_2)\times\bar{B}_\zeta(y_1,y_2))P_{X,Y}(x_1,y_1)P_{X,Y}(x_2,y_2) \\
\geq&(P_{X,Y}(x,y)-P_X(x)P_Y(y))^2 P_{X,Y}(x,y)P_{X,Y}(x,y) \\
>&0.
\end{aligned}$$

Second, we consider the case where (x, y) is the continuous point of θ. According to the assumption that the Radon-Nikodym derivative $h(x, y)$ is continuous, we can find an area $\bar{B}_\rho(x, x') \times \bar{B}_\zeta(y, y')$ $(x \neq x', y \neq y')$ such that $h(x_1, y_1) > f(x_1)g(y_1)$ for $(x_1, y_1) \in \bar{B}_\rho(x, x') \times \bar{B}_\zeta(y, y')$ if $h(x, y) > f(x)g(y)$. Thus, we have

$$\begin{aligned}
\theta(\bar{B}_\rho(x_1,x_2)\times\bar{B}_\zeta(y_1,y_2)) &= \int_{\bar{B}_\rho(x_1,x_2)\times\bar{B}_\zeta(y_1,y_2)} h(x,y)G(dx)G(dy) \\
&> \int_{\bar{B}_\rho(x_1,x_2)} f(x)G(dx)\int_{\bar{B}_\zeta(y_1,y_2)} g(y)G(dy) \\
&= \mu(\bar{B}_\rho(x_1,x_2))\nu(\bar{B}_\zeta(y_1,y_2))
\end{aligned}$$

for every $\bar{B}_\rho(x_1,x_2) \times \bar{B}_\zeta(y_1,y_2) \subseteq \bar{B}_\rho(x,x') \times \bar{B}_\zeta(y,y')$. Let $A = \{(x_1,y_1)\times(x_2,y_2) : \bar{B}_\rho(x_1,x_2) \times \bar{B}_\zeta(y_1,y_2) \subseteq \bar{B}_\rho(x,x') \times \bar{B}_\zeta(y,y')\}$, we have

$$
\begin{aligned}
&\mathbf{BCov}^2(X,Y) \\
\geq &\int_A [\theta - \mu\otimes\nu]^2(\bar{B}_\rho(x_1,x_2)\times\bar{B}_\zeta(y_1,y_2))\theta(dx_1,dy_1)\theta(dx_2,dy_2) \\
= &\int_A \Big(\theta(\bar{B}_\rho(x_1,x_2)\times\bar{B}_\zeta(y_1,y_2)) - \mu(\bar{B}_\rho(x_1,x_2))\nu(\bar{B}_\zeta(y_1,y_2))\Big)^2 \\
&\cdot h(x_1,y_1)h(x_2,y_2)G(dx_1)G(dx_2)G(dy_1)G(dy_2) \\
>&0.
\end{aligned}
$$

Combining the above conclusions, we obtain the result that $\mathbf{BCov}(X,Y)=0$ implies $\theta = \mu\otimes\nu$.

Proof of Proposition 2.2.1: Let $W_1 = (X_1,Y_1)$ and $W_2 = (X_2,Y_2)$. We begin with the right-hand side of the equation (1) in Proposition 2.2.1.

$$
\begin{aligned}
&E\xi^X_{12,3456}\xi^Y_{12,3456}\omega_1(X_1,X_2)\omega_2(Y_1,Y_2) \\
=&E(\delta^X_{12,34}\delta^Y_{12,34}\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)) + E(\delta^X_{12,34}\delta^Y_{12,56}\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)) \\
&-2E(\delta^X_{12,34}\delta^Y_{12,35}\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)).
\end{aligned}
$$

Since

$$
\begin{aligned}
&E(\delta^X_{12,34}\delta^Y_{12,34}\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)) \\
=&E[\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)E(\delta^X_{12,34}\delta^Y_{12,34}|W_1,W_2)] \\
=&E[\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)E(\delta^X_{12,3}\delta^X_{12,4}\delta^Y_{12,3}\delta^Y_{12,4}|W_1,W_2)] \\
=&E[\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)E^2(\delta^X_{12,3}\delta^Y_{12,3}|W_1,W_2)],
\end{aligned}
$$

$$
\begin{aligned}
&E(\delta^X_{12,34}\delta^Y_{12,56}\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)) \\
=&E[\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)E(\delta^X_{12,34}\delta^Y_{12,56}|W_1,W_2)] \\
=&E[\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)E(\delta^X_{12,3}\delta^X_{12,4}\delta^Y_{12,5}\delta^Y_{12,6}|W_1,W_2)] \\
=&E[\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)E^2(\delta^X_{12,3}|W_1,W_2)E^2(\delta^Y_{12,3}|W_1,W_2)],
\end{aligned}
$$

and

$$
\begin{aligned}
&E(\delta^X_{12,34}\delta^Y_{12,35}\omega_1(X_1,X_2)\omega_2(Y_1,Y_2))\\
=&E[\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)E(\delta^X_{12,34}\delta^Y_{12,35}|W_1,W_2)]\\
=&E[\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)E(\delta^X_{12,3}\delta^X_{12,4}\delta^Y_{12,3}\delta^Y_{12,5}|W_1,W_2)]\\
=&E[\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)E(\delta^X_{12,3}\delta^Y_{12,3}|W_1,W_2)E(\delta^X_{12,3}|W_1,W_2)E(\delta^Y_{12,3}|W_1,W_2)],
\end{aligned}
$$

we have

$$
\begin{aligned}
&E\xi^X_{12,3456}\xi^Y_{12,3456}\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)\\
=&E\{\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)[E(\delta^X_{12,3}\delta^Y_{12,3}|W_1,W_2)-E(\delta^X_{12,3}|W_1,W_2)E(\delta^Y_{12,3}|W_1,W_2)]^2\}\\
=&E\{\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)[P(X_3\in\bar{B}_\rho(X_1,X_2),Y_3\in\bar{B}_\zeta(Y_1,Y_2)|W_1,W_2)\\
&\quad -P(X_3\in\bar{B}_\rho(X_1,X_2)|W_1,W_2)P(Y_3\in\bar{B}_\zeta(Y_1,Y_2))|W_1,W_2)]^2\}\\
=&\int[\theta-\mu\otimes\nu]^2(\bar{B}_\rho(x_1,x_2)\times\bar{B}_\zeta(y_1,y_2))\omega_1(x_1,x_2)\omega_2(y_1,y_2)\theta(dx_1,dy_1)\theta(dx_2,dy_2)\\
=&\mathbf{BCov}_\omega(X,Y).
\end{aligned}
$$

Proof of Proposition 2.2.2: With the conclusion of Proposition 2.2.1, applying the Cauchy-Schwarz inequality to the left-hand side of the equation (2) in Proposition 2.2.2, we have

$$
\begin{aligned}
\mathbf{BCov}^2_\omega(X,Y)=&E\xi^X_{12,3456}\omega_1(X_1,X_2)\xi^Y_{12,3456}\omega_2(Y_1,Y_2)\\
\le&\sqrt{E(\xi^X_{12,3456}\omega_1(X_1,X_2))^2E(\xi^Y_{12,3456}\omega_2(Y_1,Y_2))^2}\\
=&\mathbf{BCov}_\omega(X)\mathbf{BCov}_\omega(Y).
\end{aligned}
$$

Obviously, we have $\xi^X_{12,3456}=\xi^Y_{12,3456}$ when $\delta^X_{12,3}=\delta^Y_{12,3}$. Thus, the equality holds when $\delta^X_{12,3}=\delta^Y_{12,3}$ and $\omega_1(X_1,X_2)=\omega_2(Y_1,Y_2)$.

Proof of Proposition 2.3.1: Because

$$
\begin{aligned}
&\mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{Y})\\
=&\frac{1}{n^2}\sum_{i,j=1}^{n}(\Delta^{XY}_{ij,n}-\Delta^{X}_{ij,n}\Delta^{Y}_{ij,n})^2\hat{\omega}_{1,n}(X_i,X_j)\hat{\omega}_{2,n}(Y_i,Y_j)\\
=&\frac{1}{n^2}\sum_{i,j=1}^{n}\Big(\frac{1}{n}\sum_k\delta^X_{ij,k}\delta^Y_{ij,k}-\frac{1}{n^2}\sum_k\delta^X_{ij,k}\sum_l\delta^Y_{ij,l}\Big)^2\hat{\omega}_{1,n}(X_i,X_j)\hat{\omega}_{2,n}(Y_i,Y_j)\\
=&\frac{1}{n^6}\sum_{i,j=1}^{n}\Big(n\sum_k\delta^X_{ij,k}\delta^Y_{ij,k}-\sum_k\delta^X_{ij,k}\sum_l\delta^Y_{ij,l}\Big)^2\hat{\omega}_{1,n}(X_i,X_j)\hat{\omega}_{2,n}(Y_i,Y_j)\\
=&\frac{1}{n^6}\sum_{i,j=1}^{n}\Big(n^2\sum_k\delta^X_{ij,k}\delta^Y_{ij,k}\sum_l\delta^X_{ij,l}\delta^Y_{ij,l}+\sum_k\delta^X_{ij,k}\sum_l\delta^Y_{ij,l}\sum_u\delta^X_{ij,u}\sum_v\delta^Y_{ij,v}\\
&-2n\sum_k\delta^X_{ij,k}\delta^Y_{ij,k}\sum_u\delta^X_{ij,u}\sum_v\delta^Y_{ij,v}\Big)\hat{\omega}_{1,n}(X_i,X_j)\hat{\omega}_{2,n}(Y_i,Y_j)\\
=&\frac{1}{n^6}\sum_{i,j,k,l,u,v=1}^{n}\Big[\delta^X_{ij,k}\delta^Y_{ij,k}\delta^X_{ij,l}\delta^Y_{ij,l}+\delta^X_{ij,k}\delta^Y_{ij,l}\delta^X_{ij,u}\delta^Y_{ij,v}-2\delta^X_{ij,k}\delta^Y_{ij,k}\delta^X_{ij,u}\delta^Y_{ij,v}\Big]\\
&\qquad\qquad\hat{\omega}_{1,n}(X_i,X_j)\hat{\omega}_{2,n}(Y_i,Y_j)\\
=&\frac{1}{n^6}\sum_{i,j,k,l,u,v=1}^{n}\Big[\delta^X_{ij,kl}\delta^Y_{ij,kl}+\delta^X_{ij,ku}\delta^Y_{ij,lv}-2\delta^X_{ij,ku}\delta^Y_{ij,kv}\Big]\hat{\omega}_{1,n}(X_i,X_j)\hat{\omega}_{2,n}(Y_i,Y_j)\\
=&\frac{1}{n^6}(S_1+S_2-2S_3),
\end{aligned}
$$

where

$$
\begin{aligned}
S_1&=\sum_{i,j,k,l,u,v=1}^{n}\delta^X_{ij,kl}\delta^Y_{ij,kl}\hat{\omega}_{1,n}(X_i,X_j)\hat{\omega}_{2,n}(Y_i,Y_j),\\
S_2&=\sum_{i,j,k,l,u,v=1}^{n}\delta^X_{ij,ku}\delta^Y_{ij,lv}\hat{\omega}_{1,n}(X_i,X_j)\hat{\omega}_{2,n}(Y_i,Y_j),\\
S_3&=\sum_{i,j,k,l,u,v=1}^{n}\delta^X_{ij,ku}\delta^Y_{ij,kv}\hat{\omega}_{1,n}(X_i,X_j)\hat{\omega}_{2,n}(Y_i,Y_j).
\end{aligned}
$$

On the other hand, we verify that

$$
\begin{aligned}
& V_n(\mathbf{X},\mathbf{Y}) \\
=& \frac{1}{n^6}\sum_{i,j,k,l,u,v=1}^{n} \frac{1}{4}\Big(\delta^X_{ij,kl}+\delta^X_{ij,uv}-\delta^X_{ij,ku}-\delta^X_{ij,lv}\Big)\Big(\delta^Y_{ij,kl}+\delta^Y_{ij,uv}-\delta^Y_{ij,ku}-\delta^Y_{ij,lv}\Big) \\
& \qquad\qquad\qquad \hat{\omega}_{1,n}(X_i,X_j)\hat{\omega}_{2,n}(Y_i,Y_j) \\
=& \frac{1}{n^6}\sum_{i,j,k,l,u,v=1}^{n} \frac{1}{4}\Big(4\delta^X_{ij,kl}\delta^Y_{ij,kl}+4\delta^X_{ij,kl}\delta^Y_{ij,uv}-8\delta^X_{ij,ku}\delta^Y_{ij,kv}\Big)\hat{\omega}_{1,n}(X_i,X_j)\hat{\omega}_{2,n}(Y_i,Y_j) \\
=& \frac{1}{n^6}(S_1+S_2-2S_3).
\end{aligned}
$$

Thus, the proof is completed.

Proof of Theorem 2.3.1: Since $\hat{\omega}_{1,n}(x_1,x_2)$ and $\hat{\omega}_{2,n}(y_1,y_2)$ are the uniform estimators of $\omega_1(x_1,x_2)$ and $\omega_2(y_1,y_2)$, we have

$$
\sup_{x_1,x_2,y_1,y_2} |\hat{\omega}_{1,n}(x_1,x_2)\hat{\omega}_{2,n}(y_1,y_2)-\omega_1(x_1,x_2)\omega_2(y_1,y_2)| \xrightarrow[n\to\infty]{a.s.} 0.
$$

Then

$$
\begin{aligned}
& \mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{Y}) \\
=& \frac{1}{n^6}\sum_{i,j,k,l,u,v=1} \xi^X_{ij,kluv}\xi^Y_{ij,kluv}\hat{\omega}_{1,n}(X_i,X_j)\hat{\omega}_{2,n}(Y_i,Y_j) \\
=& \frac{1}{n^6}\sum_{i,j,k,l,u,v=1} \xi^X_{ij,kluv}\xi^Y_{ij,kluv}\omega_1(X_i,X_j)\omega_2(Y_i,Y_j) \\
& + \frac{1}{n^6}\sum_{i,j,k,l,u,v=1} \xi^X_{ij,kluv}\xi^Y_{ij,kluv}(\hat{\omega}_{1,n}(X_i,X_j)\hat{\omega}_{2,n}(Y_i,Y_j)-\omega_1(X_i,X_j)\omega_2(Y_i,Y_j)).
\end{aligned}
$$

Observe that

$$
\begin{aligned}
E(|\xi^X_{12,3456}\xi^Y_{12,3456}\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)|) &\le 4E\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)<\infty, \\
E(\xi^X_{12,3456}\xi^Y_{12,3456}\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)) &= \mathbf{BCov}^2_{\omega}(X,Y)
\end{aligned}
$$

and

$$
\frac{1}{n^6}\sum_{i,j,k,l,u,v=1} \xi^X_{ij,kluv}\xi^Y_{ij,kluv}(\hat{\omega}_{1,n}(X_i,X_j)\hat{\omega}_{2,n}(Y_i,Y_j)-\omega_1(X_i,X_j)\omega_2(Y_i,Y_j)) \xrightarrow[n\to\infty]{a.s.} 0.
$$

According to Theorem 3 in Chapter 3 and Theorem 1 in Chapter 4 of Lee (1990), we obtain that

$$\mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{Y}) \xrightarrow[n\to\infty]{a.s.} \mathbf{BCov}^2_{\omega}(X,Y),\ \mathbf{BCor}^2_{\omega,n}(\mathbf{X},\mathbf{Y}) \xrightarrow[n\to\infty]{a.s.} \mathbf{BCor}^2_{\omega}(X,Y).$$

Proof of Theorem 3.1.1: According to the definition of $\mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{Y})$ and the assumptions of Theorem 3.1.1, we have

$$\begin{aligned}\mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{Y}) =& \frac{1}{n^6}\sum_{i,j,k,l,u,v=1}\xi^X_{ij,kluv}\xi^Y_{ij,kluv}\hat{\omega}_{1,n}(X_i,X_j)\hat{\omega}_{2,n}(Y_i,Y_j)\\ =&\frac{1}{n^6}\sum_{i,j,k,l,u,v=1}\xi^X_{ij,kluv}\xi^Y_{ij,kluv}\omega_1(X_i,X_j)\omega_2(Y_i,Y_j)\\ &+\frac{o_p(1)}{n^6}\sum_{i,j,k,l,u,v=1}\xi^X_{ij,kluv}\xi^Y_{ij,kluv}.\end{aligned}$$

$$\begin{aligned}&E(\xi^X_{12,3456}\xi^Y_{12,3456}\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)|(X_1,Y_1))\\ =&\frac{1}{4}E[(\delta^X_{12,34}+\delta^X_{12,56}-\delta^X_{12,35}-\delta^X_{12,46})\omega_1(X_1,X_2)|X_1]\\ &\cdot E[(\delta^Y_{12,34}+\delta^Y_{12,56}-\delta^Y_{12,35}-\delta^Y_{12,46})\omega_2(Y_1,Y_2)|Y_1]\\ =&0.\end{aligned}$$

Similarly, we can verify that $E(\xi^X_{12,3456}\xi^Y_{12,3456}\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)|(X_i,Y_i)) = 0$, where $i = 2,\ldots,6$. This means that

$$E[\bar{\psi}(W_1,\ldots,W_6)|(X_1,Y_1)=(x_1,y_1)] = 0.$$

On the other hand, we can verify that

$$\begin{aligned}&E[\bar{\psi}(W_1,\ldots,W_6)|(X_3,Y_3)=(x_3,y_3),(X_4,Y_4)=(x_4,y_4)]\\ =&\frac{1}{6!}\times 4\times 4!\times 2E[\xi^X_{12,3456}\xi^Y_{12,3456}\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)|(X_3,Y_3)=(x_3,y_3),(X_4,Y_4)=(x_4,y_4)]\\ =&\frac{1}{15}E\{(\delta^X_{12,34}+\delta^X_{12,56}-\delta^X_{12,35}-\delta^X_{12,46})\omega_1(X_1,X_2)|(X_3,Y_3)=(x_3,y_3),(X_4,Y_4)=(x_4,y_4)\}\\ &\cdot E\{(\delta^Y_{12,34}+\delta^Y_{12,56}-\delta^Y_{12,35}-\delta^Y_{12,46})\omega_2(Y_1,Y_2)|(X_3,Y_3)=(x_3,y_3),(X_4,Y_4)=(x_4,y_4)\}\end{aligned}$$

Thus, $\frac{1}{n^6}\sum_{i,j,k,l,u,v=1}\xi^X_{ij,kluv}\xi^Y_{ij,kluv}\omega_1(X_i,X_j)\omega_2(Y_i,Y_j)$ is a degenerate V statistic of order 1. Then we have

$$n\cdot\frac{1}{n^6}\sum_{i,j,k,l,u,v=1}\xi^X_{ij,kluv}\xi^Y_{ij,kluv}\omega_1(X_i,X_j)\omega_2(Y_i,Y_j)\xrightarrow[n\to\infty]{d}\sum_{v=1}^{\infty}\lambda_v Z_v^2,$$

where Z_v are independent standard normal random variables, λ_v are the eigenvalues of the symmetric function $E[\bar{\psi}(W_1,\dots,W_6)|W_1,W_2]$.

Similarly, we have

$$n\cdot\frac{o_p(1)}{n^6}\sum_{i,j,k,l,u,v=1}\xi^X_{ij,kluv}\xi^Y_{ij,kluv}\xrightarrow[n\to\infty]{P}0.$$

Therefore,

$$n\mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{Y})\xrightarrow[n\to\infty]{d}\sum_{v=1}^{\infty}\lambda_v Z_v^2.$$

Proof of Theorem 3.1.2: Suppose that X and Y are dependent and satisfy the assumptions of Theorem 2.1.1, we have $\mathbf{BCov}^2_{\omega}(X,Y)>0$ and

$$\begin{aligned}\mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{Y})=&\frac{1}{n^6}\sum_{i,j,k,l,u,v=1}\xi^X_{ij,kluv}\xi^Y_{ij,kluv}\omega_1(X_i,X_j)\omega_2(Y_i,Y_j)\\&+\frac{o_p(1)}{n^6}\sum_{i,j,k,l,u,v=1}\xi^X_{ij,kluv}\xi^Y_{ij,kluv}.\end{aligned}\tag{A.2}$$

Follow the notations in the proof of Theorem 3.1.1 and let $\bar{\psi}_1(w)=E(\bar{\psi}(W_1,\dots,W_6)|W_1=w)-\mathbf{BCov}^2_{\omega}(X,Y)$. The first term on the right-hand side of the equation (A.2) is a V statistic with non-degenerate kernel $\bar{\psi}(W_1,\dots,W_6)$, thus we have

$$\sqrt{n}(\frac{1}{n^6}\sum_{i,j,k,l,u,v=1}\xi^X_{ij,kluv}\xi^Y_{ij,kluv}\omega_1(X_i,X_j)\omega_2(Y_i,Y_j)-\mathbf{BCov}^2_{\omega}(X,Y))\xrightarrow[n\to\infty]{d}N(0,\Sigma),$$

where $\Sigma=36Var(\bar{\psi}_1(W_1))$. Similarly, $\sqrt{n}\cdot\frac{o_p(1)}{n^6}\sum_{i,j,k,l,u,v=1}\xi^X_{ij,kluv}\xi^Y_{ij,kluv}$ converges to 0 in probability. By Slutsky's theorem, we obtain

$$\sqrt{n}(\mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{Y})-\mathbf{BCov}^2_{\omega}(X,Y))\xrightarrow[n\to\infty]{d}N(0,\Sigma).$$

Proof of Theorem 3.1.3: Suppose that X and Y are dependent and satisfy the assumptions of Theorem 2.1.1. We have $\lim_{n\to\infty}\mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{Y})=c>0$ almost surely according to Theorem 2.1.1 and Theorem 2.3.1. Under the null hypothesis, Theorem 3.1.1 shows that for the significance level α, there exists a constant c_α such that

$$\lim_{n\to\infty} P(n\mathbf{BCov}_{\omega,n}(\mathbf{X},\mathbf{Y}) > c_\alpha) = \alpha.$$

Moreover, we have

$$\lim_{n\to\infty} P(n\mathbf{BCov}_{\omega,n}(\mathbf{X},\mathbf{Y}) > c_\alpha) = 1$$

under the alternative hypotheses.

B. SOME PROPERTIES FOR BALL COVARIANCE (CORRELATION)

This section adds some properties of Ball Covariance (Correlation) which are not listed in the main body of the paper. If the conditions of Theorem 2.1.1 are not satisfied, we can obtain the equivalence of independence in the equivalent distances. The equivalent distances are deduced by the equivalent norms of the corresponding Banach spaces. In Banach space, the equivalent norms of $\|\cdot\|_1$ and $\|\cdot\|_2$ means that there exist some constants c_1 and c_2 such that $c_1\|\cdot\|_1 \le \|\cdot\|_2 \le c_2\|\cdot\|_1$.

Theorem B.1 *Given two Banach spaces $(\mathscr{X},\rho)$ and $(\mathscr{Y},\zeta)$ with the Schauder bases. Denote their support sets by S_μ and S_ν respectively. Let θ be a Borel probability measure on $\mathscr{X}\times\mathscr{Y}$ and (X,Y) be a B-valued random variable defined on a probability space such that $(X,Y)\sim\theta$, $X\sim\mu$, and $Y\sim\nu$.*

(1) If $S_\theta = \mathscr{X}\times\mathscr{Y}$, then $\mathbf{BCov}_\omega(X,Y)=0$ if and only if $\theta=\mu\otimes\nu$.

(2) If $S_\theta \subsetneq \mathscr{X}\times\mathscr{Y}$, we can correct Ball Covariance as

$$\mathbf{BCov}^2_{\omega,\delta}(X,Y) := \int [\theta-\mu\otimes\nu]^2(\bar{B}_\rho(x_1,x_2)\times\bar{B}_\zeta(y_1,y_2))$$
$$\omega_1(x_1,x_2)\omega_2(y_1,y_2)\theta_\delta(dx_1,dy_1)\theta_\delta(dx_2,dy_2),$$

where μ_δ and ν_δ are two probability measures such that $TV(\mu_\delta,\mu)<\delta, TV(\nu_\delta,\nu)<\delta$ and $S_{\mu_\delta}=\mathscr{X}, S_{\nu_\delta}=\mathscr{Y}$ (TV denotes the total variation distance of probability *measures). Then $\mathbf{BCov}^2_{\omega,\delta}(X,Y)$ has the property of the equivalence of independence and $|\mathbf{BCov}_{\omega,\delta} - \mathbf{BCov}_{\omega}| < \varepsilon$.*

Proof of theorem B.1: (1) It is worth noting that we can not obtain all Borel sets from balls by using the operations of taking complements and at most countable disjoint unions in arbitrary infinite-dimensional Banach space. However, it is proved that an arbitrary infinite-dimensional Banach space with basis admits an equivalent norm such that any Borel set can be obtained from balls by taking complements and countable disjoint unions in Riss (2006). With the condition that $S_\theta = \mathscr{X} \times \mathscr{Y}$, we have $\theta(\bar{B}_\rho(x_1,x_2) \times \bar{B}_\zeta(y_1,y_2)) = \mu(\bar{B}_\rho(x_1,x_2))\nu(\bar{B}_\zeta(y_1,y_2))$ for all $x_1, x_2 \in \mathscr{X}, y_1, y_2 \in \mathscr{Y}$. This implies that $\theta(A \times B) = \mu(A)\nu(B)$ holds for all $A \in \mathcal{Q}_{\mathscr{X}}, B \in \mathcal{Q}_{\mathscr{Y}}$, where $\mathcal{Q}_{\mathscr{X}}, \mathcal{Q}_{\mathscr{Y}}$ denote the collection of closed balls on $\mathscr{X}$ and $\mathscr{Y}$, respectively. By repeating the last step of case (a) in Theorem 2.1.1 and Theorem 1 in Riss (2006), we obtain the desired result.
(2) We can find a measure θ_δ such that $TV(\theta_\delta, \theta) < \delta$ and $S_{\theta_\delta} = \mathscr{X} \times \mathscr{Y}$. Specifically, we can construct θ_δ that satisfies $TV(\theta_\delta, \theta) < \delta$ by this way:

Let $\theta_\delta = (\theta + \delta\theta_W)/(1+\delta)$, where θ_W denotes Winner measure. Then we have

$$\begin{aligned}|\theta - \theta_\delta| &= |\theta - (\theta + \delta\theta_W)/(1+\delta)| \\ &= \frac{\delta}{1+\delta}|\theta - \theta_W| \\ &\leq \delta.\end{aligned}$$

Thus, we can obtain that $TV(\theta, \theta_\delta) < \delta$ and $S_{\theta_\delta} = \mathscr{X} \times \mathscr{Y}$.

Then Ball Covariance can be corrected as

$$\begin{aligned}\mathbf{BCov}^2_{\omega,\delta}(X,Y) := \int &[\theta - \mu \otimes \nu]^2(\bar{B}_\rho(x_1,x_2) \times \bar{B}_\zeta(y_3,y_4)) \\ &\omega_1(x_1,x_2)\omega_2(y_1,y_2)\theta_\delta(dx_1,dy_1)\theta_\delta(dx_2,dy_2),\end{aligned}$$

Following the arguments of case (a) in Theorem 2.1.1 and Theorem 1 in Riss (2006), we can also verify that the new definition of $\mathbf{BCov}_{\omega,\delta}(X,Y)$ still satisfies the property that

$\mathbf{BCov}_{\omega,\delta}(X,Y) = 0$ if and only if $\theta = \mu \otimes \nu$ for the equivalent distances of ρ and ζ (We still denote them as ρ and ζ). Moreover, according to the weak convergence of probability measure, $|\mathbf{BCov}_{\omega,\delta}(X,Y) - \mathbf{BCov}_{\omega}(X,Y)| < \varepsilon$ for $\forall\ \varepsilon > 0$ when there exists a constant δ such that $|\theta - \theta_\delta| < \delta$.

Next, we show that HHG can be derived from Ball Covariance with special weight in form. Following the notations in Heller et al. (2013), we let

$$A_{1,1}(i,j) = \sum_{k=1,k\neq i,j}^{n} \delta^X_{ij,k}\delta^Y_{ij,k}, \quad A_{2,2}(i,j) = \sum_{k=1,k\neq i,j}^{n} (1-\delta^X_{ij,k})(1-\delta^Y_{ij,k}),$$

$$A_{1,2}(i,j) = \sum_{k=1,k\neq i,j}^{n} \delta^X_{ij,k}(1-\delta^Y_{ij,k}), \quad A_{2,1}(i,j) = \sum_{k=1,k\neq i,j}^{n} (1-\delta^X_{ij,k})\delta^Y_{ij,k},$$

$$A_{1,\cdot}(i,j) = A_{1,1}(i,j) + A_{1,2}(i,j), \quad A_{2,\cdot}(i,j) = A_{2,1}(i,j) + A_{2,2}(i,j),$$

$$A_{\cdot,1}(i,j) = A_{1,1}(i,j) + A_{2,1}(i,j), \quad A_{\cdot,2}(i,j) = A_{1,2}(i,j) + A_{2,2}(i,j),$$

and the HHG statistic is defined as follows

$$T = \sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} (n-2)\frac{[A_{1,2}(i,j)A_{2,1}(i,j) - A_{1,1}(i,j)A_{2,2}(i,j)]^2}{A_{1,\cdot}(i,j)A_{2,\cdot}(i,j)A_{\cdot,1}(i,j)A_{\cdot,2}(i,j)}.$$

Proposition B.1 *If $\hat{\omega}_{1,n} = \{\Delta^X_{ij,n}(1-\Delta^X_{ij,n})\}^{-1}$ and $\hat{\omega}_{2,n} = \{\Delta^Y_{ij,n}(1-\Delta^Y_{ij,n})\}^{-1}$,* $\mathbf{BCov}_{\omega,n}$ *is asymptotically equivalent to HHG.*

Proof of Proposition B.1: Since

$$\begin{aligned}
&A_{1,2}(i,j)A_{2,1}(i,j) - A_{1,1}(i,j)A_{2,2}(i,j)\\
=&A_{2,1}(i,j)[A_{1,\cdot}(i,j) - A_{1,1}(i,j)] - A_{1,1}(i,j)[A_{2,\cdot}(i,j) - A_{2,1}(i,j)]\\
=&A_{2,1}(i,j)A_{1,\cdot}(i,j) - A_{1,1}(i,j)A_{2,\cdot}(i,j)\\
=&[A_{\cdot,1}(i,j) - A_{1,1}(i,j)]A_{1,\cdot}(i,j) - A_{1,1}(i,j)[n-2-A_{1,\cdot}(i,j)]\\
=&A_{\cdot,1}(i,j)A_{1,\cdot}(i,j) - A_{1,1}(i,j)(n-2),
\end{aligned}$$

we have

$$T = \sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}(n-2)[\frac{A_{1,1}(i,j)}{n-2} - \frac{A_{1,\cdot}(i,j)}{n-2}\frac{A_{\cdot,1}(i,j)}{n-2}]^2/[\frac{A_{\cdot,1}(i,j)}{n-2}\frac{A_{\cdot,2}(i,j)}{n-2}\frac{A_{1,\cdot}(i,j)}{n-2}\frac{A_{2,\cdot}(i,j)}{n-2}].$$

Observed that $A_{1,1}(i,j)/(n-2)$, $A_{1,\cdot}(i,j)/(n-2)$, $A_{\cdot,1}(i,j)/(n-2)$, $A_{2,\cdot}(i,j)/(n-2)$, $A_{\cdot,1}(i,j)/(n-2)$ are actually $\Delta_{i,j}^{X,Y}$, $\Delta_{i,j}^{X}$, $\Delta_{i,j}^{Y}$, $1-\Delta_{i,j}^{X}$ and $1-\Delta_{i,j}^{Y}$ except that two points i,j are not considered. Therefore, HHG and Ball Covariance is asymptotically equivalent.

For HHG, we can use Taylor expansion of the weight

$$\frac{1}{\Delta_{ij,n}^X(1-\Delta_{ij,n}^X)} = \frac{1}{\mu_{ij}(1-\mu_{ij})} - \frac{1}{\mu_{ij}^2(1-\mu_{ij})^2}(\Delta_{ij,n}^X(1-\Delta_{ij,n}^X)-\mu_{ij}(1-\mu_{ij}))+R_X,$$
$$\frac{1}{\Delta_{ij,n}^Y(1-\Delta_{ij,n}^Y)} = \frac{1}{\nu_{ij}(1-\nu_{ij})} - \frac{1}{\nu_{ij}^2(1-\nu_{ij})^2}(\Delta_{ij,n}^Y(1-\Delta_{ij,n}^Y)-\nu_{ij}(1-\nu_{ij}))+R_Y,$$

where $\mu_{ij}=\mu(X\in\bar{B}_\rho(X_i,X_j))$ and $\nu_{ij}=\nu(Y\in\bar{B}_\zeta(Y_i,Y_j))$. According to Elker et al. (1979), $\Delta_{ij,n}^X$ and $\Delta_{ij,n}^Y$ uniformly converge to μ_{ij} and ν_{ij}, respectively. Thus, Theorem 2.3.1 and Theorems 3.1.1-3.1.3 can also hold for HHG. Similarly, the argument can hold for $\mathbf{BCov}_{\Delta,n}(\mathbf{X},\mathbf{Y})$.

The following proposition ensures that Ball Covariance remains invariance when the data are under the appropriate transformation.

Proposition B.2 (Transformation invariance) *If there exist maps $g:\mathscr{X}\to\mathscr{X}$ and $h:\mathscr{Y}\to\mathscr{Y}$ such that for all (x,y) and (x',y') in the support of θ, we have*

$$\rho(g(x),g(x'))=a\rho(x,x'),\zeta(h(y),h(y'))=b\zeta(y,y')$$

and

$$\omega_1(g(x),g(x'))=\omega_1(x,x'),\omega_2(h(y),h(y'))=\omega_2(y,y')$$

for some positive numbers a and b, then

$$\mathbf{BCov}_\omega^2(g(X),h(Y))=\mathbf{BCov}_\omega^2(X,Y).$$

Proof of Proposition B.2: Let $\{(X_i,Y_i),i=1,2,\ldots,6\}$ be i.i.d samples from θ. By the condition of g and h, we have

$$\begin{aligned} I(\rho(g(X_1),g(X_3))\le\rho(g(X_1),g(X_2)) =& I(a\rho(X_1,X_3)\le a\rho(X_1,X_2)) \\ =& I(\rho(X_1,X_3)\le\rho(X_1,X_2)) \end{aligned}$$

and

$$I(\zeta(h(Y_1), h(Y_3)) \leq \zeta(h(Y_1), h(Y_2)) = I(\zeta(Y_1, Y_3) \leq \zeta(Y_1, Y_2)),$$

where $I(\cdot)$ is the indicator function. Furthermore,

$$\delta_{12,3}^{g(X)} = \delta_{12,3}^{X}, \quad \delta_{12,3}^{h(Y)} = \delta_{12,3}^{Y}, \quad \delta_{12,34}^{g(X)} = \delta_{12,34}^{X}, \quad \delta_{12,34}^{h(Y)} = \delta_{12,34}^{Y},$$

and

$$\xi_{12,3456}^{g(X)} = \xi_{12,3456}^{X}, \quad \xi_{12,3456}^{h(Y)} = \xi_{12,3456}^{Y}.$$

According to Proposition 2.2.1, we can obtain that

$$\begin{aligned}\mathbf{BCov}_{\omega}^{2}(g(X), h(Y)) =& E\xi_{12,3456}^{g(X)}\xi_{12,3456}^{h(Y)}\omega_1(g(X), g(X'))\omega_2(h(Y), h(Y')) \\ =& E\xi_{12,3456}^{X}\xi_{12,3456}^{Y}\omega_1(X, X')\omega_2(Y, Y') \\ =& \mathbf{BCov}_{\omega}^{2}(X, Y).\end{aligned}$$

Remark B.1 *The following are two examples of the appropriate transformations.*

(1) *when ρ,ζ are Euclidean distances, and $\omega_1(x_1, x_2) = \omega_2(y_1, y_2) \equiv 1$, then for all constant vectors a_1, a_2, scalars b_1, b_2 and orthonormal matrices C_1, C_2 , we have*

$$\mathbf{BCov}_{\omega}^{2}(a_1 + b_1C_1X, a_2 + b_2C_2Y) = \mathbf{BCov}_{\omega}^{2}(X, Y),$$

because

$$\rho(a_1 + b_1C_1X_1, a_1 + b_1C_1X_2) = b_1^2(X_1 - X_2)'C_1'C_1(X_1 - X_2) = b_1^2\rho(X_1, X_2),$$
$$\zeta(a_2 + b_2C_2Y_1, a_2 + b_2C_2Y_2) = b_2^2(Y_1 - Y_2)'C_2'C_2(Y_1 - Y_2) = b_2^2\zeta(Y_1, Y_2).$$

(2) *when ρ,ζ are Mahalanobis distances and $\omega_1(x_1, x_2) = \omega_2(y_1, y_2) \equiv 1$, then for all constant vectors a_1, a_2, scalars b_1, b_2 and invertible matrices C_1, C_2 , we also have:*

$$\mathbf{BCov}_{\omega}^{2}(a_1 + b_1C_1X, a_2 + b_2C_2Y) = \mathbf{BCov}_{\omega}^{2}(X, Y),$$

because

$$
\begin{aligned}
&\rho(a_1 + b_1 C_1 X_1, a_1 + b_1 C_1 X_2) \\
=&b_1^2 (C_1(X_1 - X_2))'(C_1 \Sigma_X C_1')^{-1}(C_1(X_1 - X_2)) \\
=&b_1^2 \rho(X_1, X_2),
\end{aligned}
$$

where Σ_X is the covariance matrix of X. Similarly,

$$\zeta(a_2 + b_2 C_2 Y_1, a_2 + b_2 C_2 Y_2) = b_2^2 \zeta(Y_1, Y_2).$$

We can see that the $\mathbf{BCov}_\omega(X, Y)$ is rotation invariant when both ρ and ζ are Euclidean distances, and affine invariant when both ρ and ζ are Mahalanobis distances.

The next proposition provides the analogous properties of $\mathbf{BCov}_{\omega,n}(\mathbf{X}, \mathbf{Y})$ and $\mathbf{BCor}_{\omega,n}(\mathbf{X}, \mathbf{Y})$ to Proposition B.2.

Proposition B.3 *If there exist maps $g : \mathscr{X} \to \mathscr{X}$ and $h : \mathscr{Y} \to \mathscr{Y}$ such that for all (x, y) and (x', y') in the support of θ, we have*

$$\rho(g(x), g(x')) = a\rho(x, x'), \zeta(h(y), h(y')) = b\zeta(y, y')$$

and

$$\hat{\omega}_{1,n}(g(x), g(x')) = \hat{\omega}_{1,n}(x, x'), \hat{\omega}_{2,n}(h(y), h(y')) = \hat{\omega}_{2,n}(y, y')$$

for some positive numbers a and b, then we have the following properties

(1) $\mathbf{BCov}^2_{\omega,n}(g(\mathbf{X}), h(\mathbf{Y})) = \mathbf{BCov}^2_{\omega,n}(\mathbf{X}, \mathbf{Y})$.

(2) $\mathbf{BCor}^2_{\omega,n}(g(\mathbf{X}), h(\mathbf{Y})) = \mathbf{BCor}^2_{\omega,n}(\mathbf{X}, \mathbf{Y})$.

(3) $0 \le \mathbf{BCor}_{\omega,n}(\mathbf{X}, \mathbf{Y}) \le 1$.

Proof of Proposition B.3: Properties of (1), (2) and (3) can be derived from the fact that

$$\mathbf{BCov}^2_{\omega,n}(g(\mathbf{X}), h(\mathbf{Y})) = \frac{1}{n^6} \sum_{i,j,k,l,u,v=1}^{n} \xi^{g(X)}_{ij,kluv} \xi^{h(Y)}_{ij,kluv} \hat{\omega}_{1,n}(X_i, X_j) \hat{\omega}_{2,n}(Y_i, Y_j)$$

and the equations $\xi^{g(X)}_{ij,kluv} = \xi^{X}_{ij,kluv}, \xi^{h(Y)}_{ij,kluv} = \xi^{Y}_{ij,kluv}$.

In the following, we discuss some properties of $\mathbf{BCor}(X,Y)$.

Proposition B.4 $\mathbf{BCor}(X,Y)$ *is an non-decreasing function of* $|\gamma|$ *when* X *and* Y *follow the standard normal distributions with* $Cov(X,Y) = \gamma$.

Proof of Proposition B.4: Define the bivariate normal density function as

$$f(\gamma; x, y) = \frac{1}{2\pi\sqrt{1-\gamma^2}} \exp(-\frac{x^2 - 2\gamma xy + y^2}{2(1-\gamma^2)}).$$

Thus, the Ball Covariance of binary normal distribution (X,Y) is given by

$$\mathbf{BCov}^2(X,Y) = \int_{-\infty}^{\infty} [\int_{x_1-|x_2-x_1|}^{x_1+|x_2-x_1|} \int_{y_1-|y_2-y_1|}^{y_1+|y_2-y_1|} \{f(\gamma; x, y) - f(0; x, y)\} dxdy]^2$$
$$f(\gamma; x_1, y_1) f(\gamma; x_2, y_2) dx_1 dy_1 dx_2 dy_2.$$

The derivative of $\mathbf{BCov}(X,Y)$ is given by

$$\begin{aligned}
&\frac{\partial \mathbf{BCov}^2(X,Y)}{\partial \gamma} \\
=&2\int_{-\infty}^{\infty} \int_{x_1-|x_2-x_1|}^{x_1+|x_2-x_1|} \int_{y_1-|y_2-y_1|}^{y_1+|y_2-y_1|} \{f(\gamma; x, y) - f(0; x, y)\} f(\gamma; x_1, y_1) f(\gamma; x_2, y_2) \\
&f(\gamma; x, y) \frac{xy(1+\gamma^2) - (x^2+y^2)\gamma + \gamma(1-\gamma^2)}{(1-\gamma^2)^2} dxdydx_1dy_1dx_2dy_2 \\
&+ \int_{-\infty}^{\infty} [\int_{x_1-|x_2-x_1|}^{x_1+|x_2-x_1|} \int_{y_1-|y_2-y_1|}^{y_1+|y_2-y_1|} \{f(\gamma; x, y) - f(0; x, y)\} dxdy]^2 \\
&f(\gamma; x_1, y_1) f(\gamma; x_2, y_2) \frac{x_1y_1(1+\gamma^2) - (x_1^2+y_1^2)\gamma + \gamma(1-\gamma^2)}{(1-\gamma^2)^2} dx_1dy_1dx_2dy_2 \\
&+ \int_{-\infty}^{\infty} [\int_{x_1-|x_2-x_1|}^{x_1+|x_2-x_1|} \int_{y_1-|y_2-y_1|}^{y_1+|y_2-y_1|} \{f(\gamma; x, y) - f(0; x, y)\} dxdy]^2 \\
&f(\gamma; x_1, y_1) f(\gamma; x_2, y_2) \frac{x_2y_2(1+\gamma^2) - (x_2^2+y_2^2)\gamma + \gamma(1-\gamma^2)}{(1-\gamma^2)^2} dx_1dy_1dx_2dy_2.
\end{aligned}$$

For any $\gamma > 0$, we have

$$
\begin{aligned}
&\frac{\partial \mathbf{BCov}^2(X,Y)}{\partial \gamma} \\
\geq & 2\int_{-\infty}^{\infty}\int_{B_1}\{f(\gamma;x,y)-f(0;x,y)\}f(\gamma;x_1,y_1)f(\gamma;x_2,y_2) \\
& f(\gamma;x,y)\frac{xy(1+\gamma^2)-(x^2+y^2)\gamma+\gamma(1-\gamma^2)}{(1-\gamma^2)^2}dxdydx_1dy_1dx_2dy_2 \\
& + \int_{B_2}[\int_{x_1-|x_2-x_1|}^{x_1+|x_2-x_1|}\int_{y_1-|y_2-y_1|}^{y_1+|y_2-y_1|}\{f(\gamma;x,y)-f(0;x,y)\}dxdy]^2 f(\gamma;x_1,y_1)f(\gamma;x_2,y_2) \\
& \frac{x_1y_1(1+\gamma^2)-(x_1^2+y_1^2)\gamma+\gamma(1-\gamma^2)}{(1-\gamma^2)^2}dx_1dy_1dx_2dy_2 \\
& + \int_{B_3}[\int_{x_1-|x_2-x_1|}^{x_1+|x_2-x_1|}\int_{y_1-|y_2-y_1|}^{y_1+|y_2-y_1|}\{f(\gamma;x,y)-f(0;x,y)\}dxdy]^2 f(\gamma;x_1,y_1)f(\gamma;x_2,y_2) \\
& \frac{x_2y_2(1+\gamma^2)-(x_2^2+y_2^2)\gamma+\gamma(1-\gamma^2)}{(1-\gamma^2)^2}dx_1dy_1dx_2dy_2 \geq 0.
\end{aligned}
$$

where B_1, B_2 and B_3 are, respectively, given by

$$
\begin{aligned}
B_1 =& \{x_1-|x_2-x_1|<x<x_1+|x_2-x_1|, y_1-|y_2-y_1|<y<y_1+|y_2-y_1|\} \\
& \cap \{\gamma^2(x+y)^2 \leq \min(-(1-\gamma^2)\log(1-\gamma^2)-2\gamma xy, \gamma xy(1+\gamma^2)+\gamma^2(1-\gamma^2))\}, \\
B_2 =& \{4\gamma^2(x_1-\frac{1+\gamma^2}{2\gamma}y_1)^2 \leq (1-\gamma^2)^2y_1^2+4\gamma^2(1-\gamma^2)\}, \\
B_3 =& \{4\gamma^2(x_2-\frac{1+\gamma^2}{2\gamma}y_2)^2 \leq (1-\gamma^2)^2y_2^2+4\gamma^2(1-\gamma^2)\}.
\end{aligned}
$$

Thus, $\mathbf{BCov}(X,Y)$ is a monotonically non-decreasing function of γ for $\gamma > 0$. Similarly, we can prove that $\mathbf{BCov}(X,Y)$ is the monotonically non-increasing function of γ for $\gamma < 0$. The result also holds for $\mathbf{BCor}(X,Y)$ since $\mathbf{BCov}(X,X)$ and $\mathbf{BCov}(Y,Y)$ are not functions of γ. Thus, $\mathbf{BCor}(X,Y)$ is an non-decreasing function of $|\gamma|$. The relationship of Ball Correlation, distance correlation and Pearson correlation in binary normal case is shown in Figure 1.

From Proposition 2.2.2, we have $\mathbf{BCor}(X,Y) = 1$ when $\delta_{12,3}^X = \delta_{12,3}^Y$. The following proposition tells us that the opposite result can also be concluded.

Proposition B.5 *If* $\mathbf{BCor}(X,Y) = 1$ *holds, then there exists a map* $f : \mathscr{X} \longmapsto \mathscr{Y}$ *which satisfies that for all* $(X,Y),(X_1,Y_1),\ldots,(X_6,Y_6)$ *in the support of* θ*, we have* $\delta_{12,3}^X = \delta_{12,3}^Y$ *and* $Y = f(X)$ *for* θ*-a.e.*

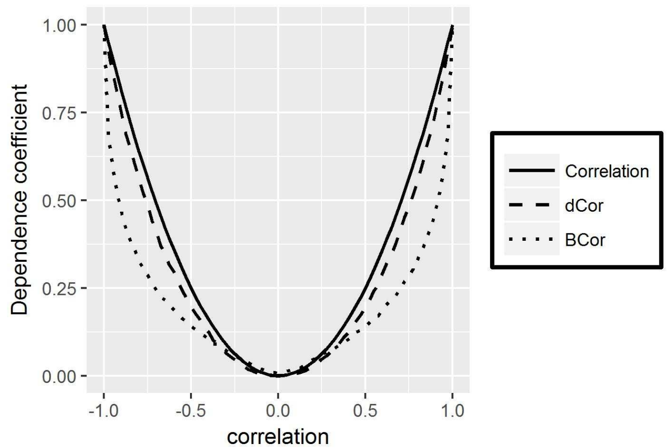

Figure 1: Ball Correlation, distance correlation and Pearson correlation γ^2 in the binary normal case

Proof of Proposition B.5: If the equality of (2) in Proposition 2.2.2 holds, then there exists a nonzero real number c, such that

$$\xi^X_{12,3456} = c\xi^Y_{12,3456} \qquad a.e.$$

That is

$$\delta^X_{12,34} + \delta^X_{12,56} - \delta^X_{12,35} - \delta^X_{12,46} = c(\delta^Y_{12,34} + \delta^Y_{12,56} - \delta^Y_{12,35} - \delta^Y_{12,46})$$

for all $(X_i, Y_i), i = 1, \ldots, 6$ in the support of θ. Because θ is non-degenerate, we can put $(X_1, Y_1) = (X_3, Y_3) = (X_5, Y_5), (X_4, Y_4) = (X_6, Y_6)$ to deduce that

$$\delta^X_{12,4} + \delta^X_{12,6} - 1 - \delta^X_{12,4} = c(\delta^Y_{12,4} + \delta^Y_{12,6} - 1 - \delta^Y_{12,4}).$$

Thus, we can obtain that $\delta^X_{12,6} - 1 = c(\delta^Y_{12,6} - 1)$ for all $(X_1, Y_1), (X_2, Y_2), (X_6, Y_6)$ in the support of θ. Because both $\delta^X_{12,6}$ and $\delta^Y_{12,6}$ only take values on $\{0, 1\}$, we obtain that if $\delta^X_{12,6} = 1$ then $\delta^Y_{12,6} = 1$; if $\delta^X_{12,6} = 0$ then $\delta^Y_{12,6} = 0$, and we can also obtain that $c = 1$ at the same time. Therefore, we have $\delta^X_{12,6} = \delta^Y_{12,6}$ for all $(X_1, Y_1), (X_2, Y_2), (X_6, Y_6)$ in the support of θ. That is

$$X_6 \in \bar{B}_\rho(X_1, X_2) \Longleftrightarrow Y_6 \in \bar{B}_\zeta(Y_1, Y_2)$$

for all $(X_1, Y_1), (X_2, Y_2), (X_6, Y_6)$ in the support of θ. If there exist two different points $(X, Y), (X, Y')$ in the support of θ, then $X \in \bar{B}_\rho(X, X)$ leads to $Y' \in \bar{B}_\zeta(Y, Y)$. That means $Y' = Y$. Similarly, we can obtain that if $Y_1 = Y_2$ then $X_1 = X_2$. Thus, we can reach the conclusion that if the equality of (2) in Proposition 2.2.2 holds, then there exists a one-to-one map $f : \mathscr{X} \to \mathscr{Y}$, such that $Y = f(X)$ for θ-a.e.

Remark B.2 *According to Proposition 2.5 in Lyons (2013),* $\mathrm{dcor}(X, Y) = 1$ *iff* $\rho(X_1, X_2) = c\zeta(Y_1, Y_2)$ *a.e. for some* $c > 0$. *The condition* $\delta^X_{12,3} = \delta^Y_{12,3}$ *is weaker than* $\rho(X_1, X_2) = c\zeta(Y_1, Y_2)$ *if the former can be derived from the latter condition.*

If $Y = f(X)$ *a.e, where the function* f *satisfies that* $\zeta(f(X), f(X')) = c\rho(X, X')$ *holds for all* (X, Y) *in the support of* θ *and some positive number* c*, then we have*

$$
\begin{aligned}
I(\zeta(Y_1, Y_3) \leq \zeta(Y_1, Y_2)) =& I(\zeta(f(X_1), f(X_3)) \leq \zeta(f(X_1), f(X_2))) \\
=& I(c\rho(X_1, X_3) \leq c\rho(X_1, X_2)) \\
=& I(\rho(X_1, X_3) \leq \rho(X_1, X_2)).
\end{aligned}
$$

According to Proposition B.5 and Remark B.2, the condition of $\mathbf{BCor}(X, Y) = 1$ *is weaker than that of* $\mathrm{dcor}(X, Y) = 1$. *To a certain extent, this implies that Ball Correlation may have better performance in detecting the strong relationships compared to distance correlation.*

REFERENCES

Banach, S. (1932), “Théorie des opérations linéaires,” , .

Bogachev, V. I. (2007), *Measure Theory Volume I* Springer.

Elker, J., Pollard, D., and Stute, W. (1979), “Glivenko-Cantelli theorems for classes of convex sets,” *Advances in Applied Probability*, 11, 820–833.

Heller, R., Heller, Y., and Gorfine, M. (2013), “A consistent multivariate test of association based on ranks of distances,” *Biometrika*, 100, 503–510.

Jackson, S., and Mauldin, R. D. (1999), “On the sigma-class generated by open balls,” *Proc. Cambridge Phil. Soc.*, 127, 99–108.

Kleiber, M., and Pervin, W. J. (1969), “A generalized Banach-Mazur theorem,” *Bulletin of The Australian Mathematical Society*, 1, 169–173.

Lee, A. (1990), *U-Statistics: Theory and Practice*, Statistics: Textbooks and Monographs M. Dekker.

Lyons, R. (2013), “Distance covariance in metric spaces,” *The Annals of Probability*, 41, 3284–3305.

Riss, E. (2006), “Generating Borel sets by balls,” *St. Petersburg Mathematical Journal*, 17, 683–698.