Expected Conditional Characteristic Function-based Measures for Testing Independence

Supplementary Material

Chenlu Ke and Xiangrong Yin*

Abstract

In this supplementary file, we include detailed proofs of our theorems.

1. **Proof of Theorem 2.**

$$
\begin{aligned}
\psi_\omega^2(\mathbf{y}) &= \int |\varphi_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{u}) - \varphi_{\mathbf{X}}(\mathbf{u})|^2 d\omega(\mathbf{u}) \\
&= \int \Big\{ \varphi_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{u})\bar{\varphi}_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{u}) - \varphi_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{u})\bar{\varphi}_{\mathbf{X}}(\mathbf{u}) \\
&\quad - \bar{\varphi}_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{u})\varphi_{\mathbf{X}}(\mathbf{u}) + \varphi_{\mathbf{X}}(\mathbf{u})\bar{\varphi}_{\mathbf{X}}(\mathbf{u}) \Big\} d\omega(\mathbf{u}) \\
&= \int \Big\{ E\left[e^{i(\mathbf{X}-\mathbf{X}')^T\mathbf{u}} | \mathbf{Y}=\mathbf{y}, \mathbf{Y}'=\mathbf{y} \right] - E\left[e^{i(\mathbf{X}-\mathbf{X}')^T\mathbf{u}} | \mathbf{Y}=\mathbf{y} \right] \\
&\quad - E\left[e^{i(\mathbf{X}-\mathbf{X}')^T\mathbf{u}} | \mathbf{Y}'=\mathbf{y} \right] + E\left[e^{i(\mathbf{X}-\mathbf{X}')^T\mathbf{u}} \right] \Big\} d\omega(\mathbf{u}) \\
&= E_{\mathbf{X}_\mathbf{y},\mathbf{X}'_\mathbf{y}} K(\mathbf{X}-\mathbf{X}') - 2E_{\mathbf{X}_\mathbf{y},\mathbf{X}'} K(\mathbf{X}-\mathbf{X}') + E_{\mathbf{X},\mathbf{X}'} K(\mathbf{X}-\mathbf{X}'). \qquad \square
\end{aligned}
$$

2. **Proof of Theorem 3.** Since K generates ρ, we can write $\rho(\mathbf{x},\mathbf{x}') = K(\mathbf{x},\mathbf{x}) + K(\mathbf{x}',\mathbf{x}') - 2K(\mathbf{x},\mathbf{x}')$. Denote $\nu = \mathbf{P} - \mathbf{Q}$, then

$$
D_\rho(\mathbf{y}) = -\int\int [K(\mathbf{x},\mathbf{x}) + K(\mathbf{x}',\mathbf{x}') - 2K(\mathbf{x},\mathbf{x}')] d\nu(\mathbf{x}) d\nu(\mathbf{x}')
$$

*Department of Statistics, University of Kentucky, 725 Rose St. Lexington, KY 40536-0082. Email: chenlu.ke@uky.edu; yinxiangrong@uky.edu

$$
\begin{aligned}
&= -2\int K(\mathbf{x},\mathbf{x})d\nu(\mathbf{x})\int d\nu(\mathbf{x}') + 2\int\int K(\mathbf{x},\mathbf{x}')d\nu(\mathbf{x})d\nu(\mathbf{x}') \\
&= 2\int\int K(\mathbf{x},\mathbf{x}')d\nu(\mathbf{x})d\nu(\mathbf{x}') \\
&= 2\gamma_K^2(\mathbf{y}). \qquad \square
\end{aligned}
$$

3. **Proof of Theorem 5.**

(1) By definition, $\mathcal{H}_K^2(\mathbf{X}|\mathbf{Y}) \geq 0$.

For arbitrary $\mathbf{x}$ and $\mathbf{x}'$ in $\mathbb{R}^p$, $\begin{pmatrix} K(\mathbf{x},\mathbf{x}) & K(\mathbf{x},\mathbf{x}') \\ K(\mathbf{x},\mathbf{x}') & K(\mathbf{x}',\mathbf{x}') \end{pmatrix}$ is positive semi-definite since K is a symmetric and positive definite kernel. Hence, $K(\mathbf{x},\mathbf{x}') \leq \sqrt{K(\mathbf{x},\mathbf{x})K(\mathbf{x}',\mathbf{x}')}$, $\forall \mathbf{x},\mathbf{x}' \in \mathbb{R}^p$. Note that

$$
\begin{aligned}
E_{\mathbf{Y}}E_{\mathbf{X_Y},\mathbf{X}'_{\mathbf{Y}}}K(\mathbf{X},\mathbf{X}') &\leq E_{\mathbf{Y}}E_{\mathbf{X_Y},\mathbf{X}'_{\mathbf{Y}}}\sqrt{K(\mathbf{X},\mathbf{X})K(\mathbf{X}',\mathbf{X}')} \\
&= E_{\mathbf{Y}}\left[E_{\mathbf{X_Y}}\sqrt{K(\mathbf{X},\mathbf{X})}\right]^2 \\
&\leq E_{\mathbf{Y}}E_{\mathbf{X_Y}}K(\mathbf{X},\mathbf{X}) \\
&= E_{\mathbf{X}}K(\mathbf{X},\mathbf{X}).
\end{aligned}
$$

Therefore, $\mathcal{H}_K^2(\mathbf{X}|\mathbf{Y}) \leq \mathcal{H}_K^2(\mathbf{X}|\mathbf{X})$.

(2) $\rho_K(\mathbf{X}|\mathbf{Y}) = 1$ iff $E_{\mathbf{Y}}E_{\mathbf{X_Y},\mathbf{X}'_{\mathbf{Y}}}K(\mathbf{X},\mathbf{X}') = E_{\mathbf{X}}K(\mathbf{X},\mathbf{X})$. All the inequalities in (1) become equalities iff $\mathbf{X}$ is a function of $\mathbf{Y}$.

4. **Proof of Theorem 6.**

Let $K_{ij} \equiv K(\mathbf{X}_i,\mathbf{X}_j)$ and $I_i^{(l)} \equiv I\{\mathbf{Y}_i = \mathbf{y}^{(l)}\}$ for $l = 1, ..., L$.

$$
\begin{aligned}
\mathcal{H}_{K,n}^2(\mathbf{X}|\mathbf{Y}) &= \sum_{l=1}^{L}\frac{n}{n_l}\frac{1}{n^2}\sum_{i,j=1}^{n_l}K(\mathbf{X}_i^{(l)},\mathbf{X}_j^{(l)}) - \frac{1}{n^2}\sum_{i,j=1}^{n}K(\mathbf{X}_i,\mathbf{X}_j) \\
&= \sum_{l=1}^{L}\frac{n}{n_l}\frac{1}{n^2}\sum_{i,j=1}^{n}K_{ij}I_i^{(l)}I_j^{(l)} - \frac{1}{n^2}\sum_{i,j=1}^{n}K_{ij} \\
&\equiv \sum_{l=1}^{L}\frac{n}{n_l}V_n^{(l)} - V_n^{(0)},
\end{aligned}
$$

where $V_n^{(1)}, ..., V_n^{(L)}$ and $V_n^{(0)}$ are V-statistics. We denote the corresponding U-statistics by $U_n^{(1)}, ..., U_n^{(L)}$ and $U_n^{(0)}$, respectively. Applying the Strong Law of Large Numbers for U-statistic (Hoeffding, 1961), we have

$$\begin{aligned}
&U_n^{(l)} \overset{a.s.}{\to} E\left[K_{12} I_1^{(l)} I_2^{(l)}\right],\ l = 1, ..., L, \\
&U_n^{(0)} \overset{a.s.}{\to} EK_{12}, \\
&\frac{n_l}{n} = \frac{1}{n}\sum_{i=1}^{n} I_i^{(l)} \overset{a.s.}{\to} p_l.
\end{aligned}$$

Note that $E_{\mathbf{X}_{\mathbf{y}^{(l)}}, \mathbf{X}'_{\mathbf{y}^{(l)}}} K(\mathbf{X}, \mathbf{X}') = \frac{1}{p_l^2} E\left[K_{12} I_1^{(l)} I_2^{(l)}\right]$. Therefore,

$$\sum_{l=1}^{L} \frac{n}{n_l} U_n^{(l)} - U_n^{(0)} \overset{a.s.}{\to} \mathcal{H}_K^2(\mathbf{X}|\mathbf{Y}).$$

Since $EK_{12} \le E_{\mathbf{X}} K(\mathbf{X}, \mathbf{X}) < \infty$, we have $E|U_n^{(l)} - V_n^{(l)}| = O(n^{-1})$ by Lemma 5.7.3 in Serfling (1980) and hence, $\mathrm{P}(|U_n^{(l)} - V_n^{(l)}| < \epsilon) \le \frac{E|U_n^{(l)} - V_n^{(l)}|}{\epsilon} \to 0$ as $n \to \infty$ for any $\epsilon > 0$, i.e., $U_n^{(l)} - V_n^{(l)} \overset{P}{\to} 0$ by Markov's inequality for $l = 0, 1, .., L$. Consequently,

$$\mathcal{H}_{K,n}^2(\mathbf{X}|\mathbf{Y}) \overset{P}{\to} \mathcal{H}_K^2(\mathbf{X}|\mathbf{Y}). \qquad \square$$

5. **Proof of Theorem 7.**

Let $K_{ij} \equiv K(\mathbf{X}_i, \mathbf{X}_j)$, $I_i^{(l)} \equiv I\{\mathbf{Y}_i = \mathbf{y}^{(l)}\}$ for $l = 1, ..., L$ and $\widetilde{K}_{ij} \equiv \widetilde{K}(\mathbf{X}_i, \mathbf{X}_j) \equiv K(\mathbf{X}_i, \mathbf{X}_j) - E_{\mathbf{X}} K(\mathbf{X}_i, \mathbf{X}) - E_{\mathbf{X}} K(\mathbf{X}, \mathbf{X}_j) + E_{\mathbf{X}, \mathbf{X}'} K(\mathbf{X}, \mathbf{X}')$.

$$\begin{aligned}
\mathcal{H}_{K,n}^2(\mathbf{X}|\mathbf{Y}) &= \sum_{l=1}^{L} \frac{n}{n_l} \frac{1}{n^2} \sum_{i,j=1}^{n} K_{ij} I_i^{(l)} I_j^{(l)} - \frac{1}{n^2} \sum_{i,j=1}^{n} K_{ij} \\
&= \sum_{l=1}^{L} \frac{n}{n_l} \frac{1}{n^2} \sum_{i,j=1}^{n} \widetilde{K}_{ij} I_i^{(l)} I_j^{(l)} - \frac{1}{n^2} \sum_{i,j=1}^{n} \widetilde{K}_{ij} \\
&\equiv \sum_{l=1}^{L} \frac{n}{n_l} V_n^{(l)} - V_n^{(0)}.
\end{aligned}$$

where $V_n^{(l)}$ $(l = 1, ..., L)$ and $V_n^{(0)}$ are V-statistics. We denote the corresponding U-statistics by $U_n^{(l)} \equiv \frac{1}{n(n-1)} \sum_{i \ne j} h_{ij}^{(l)}$ with kernel $h_{12}^{(l)} \equiv \widetilde{K}_{12} I_1^{(l)} I_2^{(l)}$ for $l =$

$1, ..., L$, and $U_n^{(0)} \equiv \frac{1}{n(n-1)} \sum_{i \neq j} h_{ij}^{(0)}$ with kernel $h_{12}^{(0)} \equiv \widetilde{K}_{12}$, respectively. Let $\widetilde{\mathcal{H}}^2_{K,n}(\mathbf{X}|\mathbf{Y}) \equiv \sum_{l=1}^{L} \frac{n}{n_l} U_n^{(l)} - U_n^{(0)}$, then

$$
\begin{aligned}
& n\mathcal{H}^2_{K,n}(\mathbf{X}|\mathbf{Y}) - n\widetilde{\mathcal{H}}^2_{K,n}(\mathbf{X}|\mathbf{Y}) \\
&= \sum_{l=1}^{L} \frac{n}{n_l} n(V_n^{(l)} - U_n^{(l)}) - n(V_n^{(0)} - U_n^{(0)}) \\
&= \sum_{l=1}^{L} \frac{n}{n_l} \left[\frac{1}{n} \sum_{i=1}^{n} \widetilde{K}_{ii} I_i^{(l)} - U_n^{(l)} \right] - \left[\frac{1}{n} \sum_{i=1}^{n} \widetilde{K}_{ii} - U_n^{(0)} \right] \\
& \xrightarrow{P} (L-1)\mathcal{H}^2_K(\mathbf{X}|\mathbf{X}).
\end{aligned}
$$

The above limit holds due to the null hypothesis. Thus our objective is to show

$$
n\widetilde{\mathcal{H}}^2_{K,n}(\mathbf{X}|\mathbf{Y}) \xrightarrow{d} (L-1)\mathcal{H}^2_K(\mathbf{X}|\mathbf{X})(Q-1), \tag{*}
$$

where $Q = \sum_{i=1}^{\infty} \lambda_i Z_i^2$, $Z_i \overset{i.i.d}{\sim} N(0,1)$ and λ_i are positive constants with $\sum_{i=1}^{\infty} \lambda_i = 1$.

A representation for $h_{12}^{(l)}$, the same as in Serfling (1980, p.196), will be used. Let $\{\phi_m^{(l)}(\cdot)\}$ denote orthonormal eigenfunctions corresponding to the eigenvalues $\{\zeta_m^{(l)}\}$ defined in connection with $h_{12}^{(l)}$, i.e., $\{\phi_m^{(l)}(\cdot)\}$ satisfies the followings for $l = 0, ..., L$:

(i) $E_{(\mathbf{X}_2,\mathbf{Y}_2)} \left[h_{12}^{(l)} \phi_m^{(l)}(\mathbf{X}_2, \mathbf{Y}_2) \right] = \zeta_m^{(l)} \phi_m^{(l)}(\mathbf{X}_1, \mathbf{Y}_1)$

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

(iii) $\lim_{M\to\infty} E\left[h_{12}^{(l)} - \sum_{m=1}^{M} \zeta_m^{(l)} \phi_m^{(l)}(\mathbf{X}_1, \mathbf{Y}_1) \phi_m^{(l)}(\mathbf{X}_2 \mathbf{Y}_2) \right]^2 = 0.$

Then we write $h_{12}^{(l)} = \sum_{m=1}^{\infty} \zeta_m^{(l)} \phi_m^{(l)}(\mathbf{X}_1, \mathbf{Y}_1) \phi_m^{(l)}(\mathbf{X}_2 \mathbf{Y}_2)$. In the same sense, we have $h_1^{(l)}(\mathbf{X}_1, \mathbf{Y}_1) \equiv E_{(\mathbf{X}_2,\mathbf{Y}_2)} h_{12}^{(l)} = \sum_{m=1}^{\infty} \zeta_m^{(l)} \phi_m^{(l)}(\mathbf{X}_1, \mathbf{Y}_1) E\phi_m^{(l)}(\mathbf{X}_2 \mathbf{Y}_2)$. Note that $E_{\mathbf{X}_2} \widetilde{K}_{12} = 0$ so $h_1^{(l)}(\mathbf{X}_1, \mathbf{Y}_1) = 0$. Therefore, $E\phi_m^{(l)} = 0$ as $\mathrm{Var} h_1^{(l)} = 0$, for all m.

Let $\{\phi_m(\cdot)\}$ denote orthonormal eigenfunctions corresponsing to the eigenvalues $\{\zeta_m\}$ defined in connection with $\widetilde{K}_{12}$. Similarly, $E\phi_m = 0$. We can deduce from (i) and (ii) that:

(a) $\phi_m^{(l)}(\mathbf{X}_1, \mathbf{Y}_1) = \frac{1}{\sqrt{p_l}}\phi_m(\mathbf{X}_1)I_1^{(l)}$ for $l = 1, ..., L$

(b) $\phi_m^{(0)}(\mathbf{X}_1, \mathbf{Y}_1) = \phi_m(\mathbf{X}_1)$,

(c) $\frac{\zeta_m^{(l)}}{p_l} = \zeta_m^{(0)} = \zeta_m$, $l = 1, ..., L$.

We explain (a) and (c) only and the rest follows from the same logic. From (i),

$$\begin{aligned}\zeta_m^{(l)}\phi_m^{(l)}(\mathbf{X}_1, \mathbf{Y}_1) &= E_{(\mathbf{X}_2,\mathbf{Y}_2)}\left[h_{12}^{(l)}\phi_m^{(l)}(\mathbf{X}_2, \mathbf{Y}_2)\right] \\ &= \begin{cases} 0, & \text{if } \mathbf{Y}_1 \neq \mathbf{y}^{(l)}, \\ p_l E_{\mathbf{X}_2}\left[\widetilde{K}_{12}\phi_m^{(l)}(\mathbf{X}_2, \mathbf{y}^{(l)})\right], & \text{if } \mathbf{Y}_1 = \mathbf{y}^{(l)}, \end{cases}\end{aligned}$$

for $l = 1, ..., L$. Hence, $\phi_m^{(l)}(\mathbf{X}_1, \mathbf{Y}_1) = c\phi_m(\mathbf{X}_1)I_1^{(l)}$ for some constant c and $\frac{\zeta_m^{(l)}}{p_l} = \zeta_m$, for l=1,...,L. Required by (ii), $c = \frac{1}{\sqrt{p_l}}$.

Let $\widetilde{H} \equiv \sum_{m=1}^{\infty}\zeta_m[\sum_{j=1}^{L-1} Z_{m,j}^2 - (L-1)]$ and $\widetilde{H}_M \equiv \sum_{m=1}^{M}\zeta_m[\sum_{j=1}^{L-1} Z_{m,j}^2 - (L-1)]$, where $Z_{m,j} \overset{i.i.d}{\sim} N(0, 1)$. Putting $T_n^{(l)} \equiv \frac{1}{n}\sum_{i\neq j} h_{ij}^{(l)}$, we have $nU_n^{(l)} = \frac{n}{n-1}T_n^{(l)}$. In terms of the above representation for $h^{(l)}$, $T_n^{(l)} = \frac{1}{n}\sum_{i\neq j}\sum_{m=1}^{\infty}\zeta_m^{(l)}\phi_m^{(l)}(\mathbf{X}_i, \mathbf{Y}_i)\phi_m^{(l)}(\mathbf{X}_j, \mathbf{Y}_j)$ and let $T_{n,M}^{(l)} \equiv \frac{1}{n}\sum_{i\neq j}\sum_{m=1}^{M}\zeta_m^{(l)}\phi_m^{(l)}(\mathbf{X}_i, \mathbf{Y}_i)\phi_m^{(l)}(\mathbf{X}_j, \mathbf{Y}_j)$ for $l = 0, ..., L$. Eventually, we will show that

$$n\widetilde{\mathcal{H}}_{K,n}^2(\mathbf{X}|\mathbf{Y}) = \frac{n}{n-1}\left[\sum_{l=1}^{L}\frac{n}{n_l}T_n^{(l)} - T_n^{(0)}\right] \xrightarrow{d} \widetilde{H} \qquad (**)$$

by using characteristic functions. The proof is decomposed into 3 parts.

(1) Given $\epsilon > 0$ and s, $\left|Ee^{is\left(\sum_{l=1}^{L}\frac{n}{n_l}T_n^{(l)} - T_n^{(0)}\right)} - Ee^{is\left(\sum_{l=1}^{L}\frac{n}{n_l}T_{n,M}^{(l)} - T_{n,M}^{(0)}\right)}\right| < \epsilon$ for M and n sufficiently large.

Using the inequality $|e^{iz} - 1| \leq |z|$, we have

$$\left|Ee^{is\left(\sum_{l=1}^{L}\frac{n}{n_l}T_n^{(l)} - T_n^{(0)}\right)} - Ee^{is\left(\sum_{l=1}^{L}\frac{n}{n_l}T_{n,M}^{(l)} - T_{n,M}^{(0)}\right)}\right|$$

$$\leq |s| E \left| \sum_{l=1}^{L} \frac{n}{n_l} \left(T_n^{(l)} - T_{n,M}^{(l)} \right) - \left(T_n^{(0)} - T_{n,M}^{(0)} \right) \right|$$
$$\leq |s| \left\{ \sum_{l=1}^{L} \frac{n}{n_l} E \left| T_n^{(l)} - T_{n,M}^{(l)} \right| + E \left| T_n^{(0)} - T_{n,M}^{(0)} \right| \right\}$$
$$\leq |s| \left\{ \sum_{l=1}^{L} \frac{n}{n_l} \left[E \left(T_n^{(l)} - T_{n,M}^{(l)} \right)^2 \right]^{1/2} + \left[E \left(T_n^{(0)} - T_{n,M}^{(0)} \right)^2 \right]^{1/2} \right\}.$$

Similar to Serfling (1980, p.197 - p.198), we can show that $\sum_{m=1}^{\infty} \zeta_m^2 = E\widetilde{K}_{12}^2 < \infty$ since $E_{\mathbf{X}} K(\mathbf{X}, \mathbf{X}) < \infty$, and $E \left(T_n^{(l)} - T_{n,M}^{(l)} \right)^2 \leq 2^2 \sum_{m=M+1}^{\infty} \left[\zeta_m^{(l)} \right]^2$ for $l = 0, ..., L$. Combining with the fact that $\frac{n}{n_l} \xrightarrow{a.s.} \frac{1}{p_l}$, the conclusion follows.

(2) $\sum_{l=1}^{L} \frac{n}{n_l} T_{n,M}^{(l)} - T_{n,M}^{(0)} \xrightarrow{d} \widetilde{H}_M$.

We may write

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

where $W_{n,m}^{(l)} \equiv n^{-1/2} \sum_{t=1}^{n} \phi_K^{(l)}(\mathbf{X}_t, \mathbf{Y}_t)$ and $R_{n,m}^{(l)} = n^{-1} \sum_{t=1}^{n} \left[\phi_K^{(l)}(\mathbf{X}_t, \mathbf{Y}_t) \right]^2$.

From the foregoing considerations, it can be seen that

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

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

$$\Sigma = \begin{pmatrix} 1 & & & \sqrt{p_1} \\ & \ddots & & \vdots \\ & & 1 & \sqrt{p_L} \\ \sqrt{p_1} & \cdots & \sqrt{p_l} & 1 \end{pmatrix},$$

and $\text{Cov}(\mathbf{W}_{n,m_1}, \mathbf{W}_{n,m_2}) = \mathbf{0}$ for $m_1 \neq m_2$.

Also, $R_{n,m}^{(l)} \xrightarrow{P} 1$ for $l = 0, ..., L$. Let $\mathbf{i}^2 = 1$.

Let $\mathbf{A}_n \equiv \text{diag} \left(\sqrt{\frac{n}{n_1} p_1}, ..., \sqrt{\frac{n}{n_L} p_L}, \mathbf{i} \right)$, then $\mathbf{A}_n \xrightarrow{P} \mathbf{A} \equiv \text{diag}(1 \ \cdots \ 1 \ \mathbf{i})$ and $\mathbf{A}\Sigma\mathbf{A}^T$ has and only has non-zero eigenvalue 1 with multiplicity $L-1$. Therefore, $(\mathbf{A}\mathbf{W}_{n,m})^T \mathbf{A}\mathbf{W}_{n,m} \xrightarrow{d} \sum_{i=1}^{L-1} Z_i^2 \sim \chi_{L-1}^2$ as $Z_i \overset{iid}{\sim} N(0,1)$ and

$$\sum_{l=1}^{L} \frac{n}{n_l} T_{n,M}^{(l)} - T_{n,M}^{(0)} = \sum_{m=1}^{M} \zeta_m \left[(A\mathbf{W}_{n,m})^T A\mathbf{W}_{n,m} - \left(\sum_{l=1}^{L} p_l \frac{n}{n_l} R_{n,m}^{(l)} - R_{n,m}^{(0)} \right) \right]$$

$$\xrightarrow{d} \widetilde{H}_M.$$

(3) Given $\epsilon > 0$, $\left| Ee^{is\widetilde{H}} - Ee^{is\widetilde{H}_M} \right| < \epsilon$ for M sufficiently large.

This can be seen by Serfling (1980, p.199).

Combining (1) to (3), we establish (**). To finish the proof of (*), note that

$$\begin{aligned}
\widetilde{H} &= \sum_{m=1}^{\infty} \zeta_m [\sum_{j=1}^{L-1} Z_{m,j}^2 - (L-1)] \\
&= \sum_{i=1}^{\infty} \widetilde{\zeta}_i (Z_i^2 - 1), \text{ where } Z_i^2 \overset{i.i.d}{\sim} N(0,1) \text{ and } \sum_{i=1}^{\infty} \widetilde{\zeta}_i = (L-1) \sum_{i=1}^{\infty} \zeta_i, \\
&= [(L-1) \sum_{i=1}^{\infty} \zeta_i](Q-1), \text{ where } Q = \sum_{i=1}^{\infty} \lambda_i Z_i^2 \text{ and } \sum_{i=1}^{\infty} \lambda_i = 1.
\end{aligned}$$

Thus, we need to show that $\sum_{m=1}^{\infty} \zeta_m = \mathcal{H}_K^2(\mathbf{X}|\mathbf{X})$. Indeed,

$$\mathcal{H}_K^2(\mathbf{X}|\mathbf{X}) = E_{\mathbf{X}} \widetilde{K}(\mathbf{X}, \mathbf{X}) = \sum_{m=1}^{\infty} \zeta_m E \phi_m^2(\mathbf{X}) = \sum_{m=1}^{\infty} \zeta_m. \qquad \square$$

6. **Proof of Corollary 1.**

$\mathcal{H}_{K,n}^2(\mathbf{X}|\mathbf{X}) \xrightarrow{P} \mathcal{H}_K^2(\mathbf{X}|\mathbf{X})$ by Theorem 6. If $\mathbf{X}$ and $\mathbf{Y}$ are independent, the conclusion follows from Theorem 6. If $\mathbf{X}$ and $\mathbf{Y}$ are dependent, $\mathcal{H}_{K,n}^2(\mathbf{X}|\mathbf{Y}) \xrightarrow{P} \mathcal{H}_K^2(\mathbf{X}|\mathbf{Y}) > 0$, and therefore, $n\mathcal{H}_{K,n}^2(\mathbf{X}|\mathbf{Y}) \xrightarrow{P} \infty$. $\square$

7. **Proof of Theorem 8.**

Let $\mathbf{W} \equiv (\mathbf{X}, \mathbf{Y})$ and $P_n(\mathbf{W}_1, .., \mathbf{W}_5) \equiv G_{12} G_{13} d_{2345}$. Note that $P_n(\mathbf{W}_1, ..., \mathbf{W}_5)$ is not symmetric in its arguments, we hence re-write Γ_n^U as a U-statistic with a symmetric kernel, i.e.

$$\Gamma_n^U(\mathbf{X}|\mathbf{Y}) = \binom{n}{5} \sum_{t_1 < ... < t_5} \mathcal{P}_n(\mathbf{W}_{t_1}, ..., \mathbf{W}_{t_5}),$$

where $\mathcal{P}_n(\mathbf{W}_{t_1}, ..., \mathbf{W}_{t_5}) \equiv \frac{1}{5!} \sum_{\pi} P_n(\mathbf{W}_{i_1}, ..., \mathbf{W}_{i_5})$ and $\sum_{\pi}$ denotes summation over the 5! permutations $(i_1, ..., i_5)$ of $(t_1, ..., t_5)$. Let $\theta_n = E\mathcal{P}_n(\mathbf{W}_1, ..., \mathbf{W}_5)$ and $\theta = E_{\mathbf{Y}} [\gamma_K^2(\mathbf{Y}) f^2(\mathbf{Y})]$. Our goal is to show that $\Gamma_n^U(\mathbf{X}|\mathbf{Y}) \xrightarrow{P} \theta$. The proof involves two steps.

(1) $\theta_n = \theta + o_p(1)$.

First note that

$$\begin{aligned}\theta_n &= E(G_{12}G_{13}d_{2345}) \\ &= E(G_{12}G_{13}K_{23}) - 2E(G_{12}G_{13}K_{24}) + E(G_{12}G_{13}K_{45}) \\ &= \theta_{n1} + \theta_{n2} + \theta_{n3}\end{aligned}$$

Consider the first term,

$$\begin{aligned}\theta_{n1} &= \int K(\mathbf{x}_2, \mathbf{x}_3) h^{-2q} G(\frac{\mathbf{y}_1 - \mathbf{y}_2}{h}) G(\frac{\mathbf{y}_1 - \mathbf{y}_3}{h}) \\ &\quad f(\mathbf{x}_1, \mathbf{y}_1) f(\mathbf{x}_2, \mathbf{y}_2) f(\mathbf{x}_3, \mathbf{y}_3) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{x}_3 d\mathbf{y}_1 d\mathbf{y}_2 d\mathbf{y}_3 \\ &= \int K(\mathbf{x}_2, \mathbf{x}_3) f(\mathbf{x}_2|\mathbf{y}_1 + h\mathbf{u}) f(\mathbf{x}_3|\mathbf{y}_1 + h\mathbf{u}) d\mathbf{x}_2 d\mathbf{x}_3 \\ &\quad G(\mathbf{u}) G(\mathbf{v}) f(\mathbf{y}_1 + h\mathbf{u}) f(\mathbf{y}_1 + h\mathbf{u}) d\mathbf{u} d\mathbf{v} f(\mathbf{y}_1) d\mathbf{y}_1 \\ &= \int K(\mathbf{x}_2, \mathbf{x}_3) f(\mathbf{x}_2|\mathbf{y}_1) f(\mathbf{x}_3|\mathbf{y}_1) d\mathbf{x}_2 d\mathbf{x}_3 f^3(\mathbf{y}_1) d\mathbf{y}_1 + O_p(h^\nu)\end{aligned}$$

by Taylor expansion and condition (C1), (C3) and (C4). Similarly,

$$\begin{aligned}\theta_{n2} &= \int K(\mathbf{x}_2, \mathbf{x}_4) f(\mathbf{x}_2|\mathbf{y}_1) f(\mathbf{x}_4) d\mathbf{x}_2 d\mathbf{x}_4 f^3(\mathbf{y}_1) d\mathbf{y}_1 + O_p(h^\nu) \\ \theta_{n2} &= \int K(\mathbf{x}_4, \mathbf{x}_5) f(\mathbf{x}_4) f(\mathbf{x}_5) d\mathbf{x}_4 d\mathbf{x}_5 f^3(\mathbf{y}_1) d\mathbf{y}_1 + O_p(h^\nu)\end{aligned}$$

Combining the three terms, we have $\theta_n = \theta + o_p(1)$.

(2) $\Gamma_n^U(\mathbf{X}|\mathbf{Y}) = \theta_n + o_p(1)$.

We adopt the H-decomposition in Lee (1990) and denote

$$\mathcal{P}_{nc}(\mathbf{W}_1, ..., \mathbf{W}_c) \equiv E[\mathcal{P}_n(\mathbf{W}_1, ..., \mathbf{W}_5)|\mathbf{W}_1, ..., \mathbf{W}_c],$$

$$P_{nc}(\mathbf{W}_{i_1}, ..., \mathbf{W}_{i_c}) \equiv E[P_n(\mathbf{W}_1, ..., \mathbf{W}_5)|\mathbf{W}_{i_1}, ..., \mathbf{W}_{i_c}],$$

where $1 \leq i_1 < ... < i_c \leq 5$. Let $\phi_n^{(1)} \equiv \mathcal{P}_{n1}(\mathbf{W}_1) - \theta_n$ and

$$\phi_n^{(c)} \equiv \mathcal{P}_{nc}(\mathbf{W}_1, ..., \mathbf{W}_c) - \sum_{j=1}^{c-1} \sum_{(c,j)} \phi_n^{(j)}(\mathbf{W}_{i_1}, ..., \mathbf{W}_{i_j}) - \theta_n,$$

where the $\sum_{(c,j)}$ is taken over all subsets $1 \leq i_1 < ... < i_j \leq c$ of $\{1, ..., c\}$.

Then

$$\Gamma_n^U(\mathbf{X}|\mathbf{Y}) = E\mathcal{P}_n(\mathbf{W}_1, ..., \mathbf{W}_5) + \sum_{c=1}^{5} \binom{5}{c} \Phi_n^{(c)},$$

where $\Phi_n^{(c)} = \binom{n}{c}^{-1} \sum_{(n,c)} \phi_n^{(c)}(\mathbf{W}_{i_1}, ..., \mathbf{W}_{i_c})$ satisfies the following properties:

(i) $\Phi_n^{(c)}$ are uncorrelated with $E[\Phi_n^{(c)}] = 0$, $c = 1, ..., 5$.

(ii) $Var[\Phi_n^{(c)}] = \binom{n}{c}^{-1} Var[\phi_n^{(c)}(\mathbf{W}_1, ..., \mathbf{W}_c)]$.

(iii) $Var[\phi_n^{(c)}(\mathbf{W}_1, ..., \mathbf{W}_c)] = \sum_{j=1}^{c} (-1)^{c-j} \binom{c}{j} Var[\mathcal{P}_{nj}(\mathbf{W}_1, ..., \mathbf{W}_j)]$.

We first show that $Var[\Phi_n^{(1)}] = \frac{1}{n} Var[\mathcal{P}_{n1}(\mathbf{W}_1)] = O_p(\frac{1}{n})$. Note that $E[\mathcal{P}_{n1}^2(\mathbf{W}_1)]$ can be expanded into several terms as

$$E[\mathcal{P}_{n1}^2(\mathbf{W}_1)] = \frac{1}{5} \sum_{i=1}^{5} E[P_{n1}^2(\mathbf{W}_i)],$$

and each of these terms can be shown to be $O_p(1)$. For example,

$$\begin{aligned}
& E[P_{n1}^2(\mathbf{W}_1)] \\
=& E[E^2(G_{12}G_{13}K_{23}|\mathbf{W}_1)] + 4E[E^2(G_{12}G_{13}K_{24}|\mathbf{W}_1)] + E[E^2(G_{12}G_{13}K_{45}|\mathbf{W}_1)] \\
& - 4E[E(G_{12}G_{13}K_{23}|\mathbf{W}_1)E(G_{12}G_{13}K_{24}|\mathbf{W}_1)] - 4E[E(G_{12}G_{13}K_{45}|\mathbf{W}_1)E(G_{12}G_{13}K_{24}|\mathbf{W}_1)] \\
& + 2E[E(G_{12}G_{13}K_{23}|\mathbf{W}_1)E(G_{12}G_{13}K_{45}|\mathbf{W}_1)] \\
=& E_{11} + 4E_{12} + E_{13} - 4E_{14} - 4E_{15} + E_{16},
\end{aligned}$$

where

$$\begin{aligned}
E_{11} =& \int \Bigg[\int K(\mathbf{x}_2, \mathbf{x}_3) h^{-2q} G(\frac{\mathbf{y}_1 - \mathbf{y}_2}{h}) G(\frac{\mathbf{y}_1 - \mathbf{y}_3}{h}) f(\mathbf{x}_2|\mathbf{y}_2) f(\mathbf{x}_3|\mathbf{y}_3) d\mathbf{x}_2 d\mathbf{x}_3 \\
& f(\mathbf{y}_2) f(\mathbf{y}_3) d\mathbf{y}_2 d\mathbf{y}_3 \Bigg]^2 f(\mathbf{y}_1) d\mathbf{y}_1 \\
=& \int \Bigg[\int K(\mathbf{x}_2, \mathbf{x}_3) G(\mathbf{u}) G(\mathbf{v}) f(\mathbf{x}_2|\mathbf{y}_1 + h\mathbf{u}) f(\mathbf{x}_3|\mathbf{y}_1 + h\mathbf{v}) d\mathbf{x}_2 d\mathbf{x}_3 \\
& f(\mathbf{y}_1 + h\mathbf{u}) f(\mathbf{y}_1 + h\mathbf{v}) d\mathbf{u} d\mathbf{v} \Bigg]^2 f(\mathbf{y}_1) d\mathbf{y}_1
\end{aligned}$$

$$\leq \int \left\{ \int \left[\int \sqrt{K(\mathbf{x}_2, \mathbf{x}_2)} f(\mathbf{x}_2|\mathbf{y}_1 + h\mathbf{u}) d\mathbf{x}_2 \right] \left[\int \sqrt{K(\mathbf{x}_3, \mathbf{x}_3)} f(\mathbf{x}_3|\mathbf{y}_1 + h\mathbf{v}) d\mathbf{x}_3 \right] \right.$$

$$\left. G(\mathbf{u})G(\mathbf{v}) f(\mathbf{y}_1 + h\mathbf{u}) f(\mathbf{y}_1 + h\mathbf{v}) d\mathbf{u} d\mathbf{v} \right\}^2 f(\mathbf{y}_1) d\mathbf{y}_1$$

$$= O_p(1),$$

and $E_{1i} = O_p(1)$ for $i = 2, ..., 6$, which can be shown analogously to above. Therefore, $Var[\Phi_n^{(1)}] = O_p(\frac{1}{n})$. Futhermore, we can obtain that $Var[\Phi_n^{(2)}] = O_p(\frac{1}{n^2h^q})$ and $Var[\Phi_n^{(c)}] = O_p(\frac{1}{n^c h^{2q}})$ for $c \geq 3$ by similar logic. Then by Chebyshev's inequality, $\Phi_n^{(c)} = o_p(1)$ for $c = 1, ..., 5$ and hence, $\Gamma_n^U(\mathbf{X}|\mathbf{Y}) = \theta_n + o_p(1)$.

8. **Proof of Theorem 9.**

We continue to use the notations in the proof of Theorem 8. This proof is built upon Lemma B.4 in Fan and Li (1996). We first examine three prerequisites for $nh^{q/2}\Gamma_n^U(\mathbf{X}|\mathbf{Y})$ to be asymptotically normally distributed.

(1) Under H_0 and assumption $E_{\mathbf{X}} K^2(\mathbf{X}, \mathbf{X}) < \infty$, it is easy to show that $E[\mathcal{P}_{n1}(\mathbf{W}_1)] = 0$ and $E[\mathcal{P}_n^2(\mathbf{W}_1, ..., \mathbf{W}_5)] < \infty$.

(2) When $n \to \infty$,

$$\frac{E[\mathcal{G}_n^2(\mathbf{W}_1, \mathbf{W}_2)] + n^{-1}E[\mathcal{P}_{n2}^4(\mathbf{W}_1, \mathbf{W}_2)]}{E^2[\mathcal{P}_{n2}^2(\mathbf{W}_1, \mathbf{W}_2)]} \to 0,$$

where

$$\mathcal{G}_n(\mathbf{W}_1, \mathbf{W}_2) = E[\mathcal{P}_{n2}(\mathbf{W}_1, \mathbf{W}_3)\mathcal{P}_{n2}(\mathbf{W}_2, \mathbf{W}_3)|\mathbf{W}_1, \mathbf{W}_2].$$

Indeed, we can verify that $E[\mathcal{G}_n^2(\mathbf{W}_1, \mathbf{W}_2)] = O_p(h^{-q})$, $E[\mathcal{P}_{n2}^4(\mathbf{W}_1, \mathbf{W}_2)] = O_p(h^{-3q})$ and $E^2[\mathcal{P}_{n2}^2(\mathbf{W}_1, \mathbf{W}_2)] = O_p(h^{-2q})$. The conclusion follows from the assumptions that $h^q \to 0$ and $nh^q \to \infty$ as $n \to \infty$.

(3) $\dfrac{E[\mathcal{P}_{nc}^2]}{E[\mathcal{P}_{n2}^2]} = \dfrac{O_p(h^{-2q})}{O_p(h^{-q})} = O_p(n^{(c-2)})$ for $c = 3, 4, 5$.

According to Lemma B.4 in Fan and Li (1996), with (1)-(3) verified, it follows that $n\Gamma_n^U(\mathbf{X}|\mathbf{Y})$ is asymptotically distributed as $N(0, \frac{5^2(5-1)^2}{2} E[\mathcal{P}_{n2}^2(\mathbf{W}_1, \mathbf{W}_2)])$. Note that $E[\mathcal{P}_{n2}^2(\mathbf{W}_1, \mathbf{W}_2)] = \frac{1}{100} E[P_{n2}^2(\mathbf{W}_2, \mathbf{W}_3)] + O_p(1)$ and $E[P_{n2}^2(\mathbf{W}_2, \mathbf{W}_3)] = h^{-q}(\sigma^2 + o_p(1))$, where

$$\sigma^2 = C^q \left[EK_{12}^2 - 2EK_{12}K_{13} + E^2 K_{12} \right] Ef^3(\mathbf{Y})$$

and $C = \int_{\mathbb{R}} [\int_{\mathbb{R}} g(\mu + \nu) g(\mu) d\mu]^2 dv$. Therefore,

$$nh^{q/2}\Gamma_n^U(\mathbf{X}|\mathbf{Y}) \xrightarrow{d} N(0, 2\sigma^2).$$

References

Fan, Y., & Li, Q. (1996), "Consistent model specification tests: omitted variables and semiparametric functional forms," *Econometrica: Journal of the econometric society*, pp. 865–890.

Hoeffding, W. (1961), The strong law of large numbers for u-statistics,, Technical report, North Carolina State University. Dept. of Statistics.

Lee, A. J. (1990), *U-Statistics: Theory and Practice*, New York: Marcel Dekker.

Serfling, R. J. (1980), *Approximation theorems of mathematical statistics*, Hoboken: John Wiley & Sons.