PNAS
www.pnas.org

Supplementary Information for

Jackknife Approach to the Estimation of Mutual Information

Xianli Zeng, Yingcun Xia and Howell Tong

Yingcun Xia.
E-mail: staxyc@nus.edu.sg

This PDF file includes:

Supplementary text
Figs. S1 to S2
Table S1
References for SI reference citations

www.pnas.org/cgi/doi/10.1073/pnas.1715593115

Supporting Information Text

In this document, we provide supplementary information, including the estimation procedure, regularity conditions for theorems, models considered in simulation, simulation results not listed in the paper, details of notation and proofs of all theorems. This document is organized as follows.

In section A, we list several notations that are used frequently in the main text and this supplementary document.

In section B, we describe the estimation procedure for JMI.

In section C, we state the regularity conditions required for our theoretical results, and provide details for the constants appearing in the theorems.

In section D, we list all the models we have considered in our simulation study.

In section E, we provide simulation results not shown in the main text.

In section F, we show the extraordinary local power for independence test based on JMI for the case that both X and Y are univariate random variables.

In section G, we extend the JMI estimator to discrete random vector case.

In section H, we list technical lemmas for proving theoretical results.

In section I, we present proofs for theorems in the main text and this supplementary document.

In section J, we prove all auxiliary lemmas in section H.

A. Useful notations.

Table S1. Notations and their meanings

Notation	Meaning
C	universal constant (Without ambiguity, we use C denote all unknown constants)
C',C''	universal constant (when there are more than one unknown constants)
C_k	constant that changes with subscript k
$\frac{df}{dx}$	first order derivative of univariate function f
$\frac{d^2f}{dx^2}$	second order derivative of univariate function f
$\frac{\partial \mathbf{f}}{\partial u_p}$	first order partial derivatives of P-variate function $\mathbf{f}$ with respect to the p-th element, $p = 1, 2, ..., P$
$\frac{\partial^2 \mathbf{f}}{\partial u_{p_1} \partial u_{p_2}}$	second order partial derivatives of P-variate function $\mathbf{f}$ with respect to the p_1-th and p_2-th elements, $p_1, p_2 = 1, 2, ..., P$
$\mathbf{X}$	$= (X_1, X_2, ..., X_P)'$, a P dimensional random vector
$\mathbf{Y}$	$= (Y_1, Y_2, ..., Y_Q)'$, a Q dimensional random vector
$\mathbf{x}_i$	$= (x_{i1}, x_{i2}, ..., x_{iP})'$, sample from $\mathbf{X}$
$\mathbf{y}_i$	$= (y_{i1}, y_{i2}, ..., y_{iQ})'$, sample from $\mathbf{Y}$
$F_{\mathbf{X}}$	cumulative distribution function of $\mathbf{X}$
F_{X_p}	cumulative distribution function of $X_p, p = 1, 2, ..., P$
$F_{X_p,n}$	empirical cumulative distribution function of $X_p, p = 1, 2, ..., P$
$F_{\mathbf{Y}}$	cumulative distribution function of $\mathbf{Y}$
F_{Y_p}	cumulative distribution function of $Y_q, q = 1, 2, ..., Q$
$F_{Y_p,n}$	empirical cumulative distribution function of $Y_q, q = 1, 2, ..., Q$
$\mathbf{U}$	$= (U_1, U_2, ..., U_P) = (F_{X_1}(X_1), F_{X_2}(X_2), ..., F_{X_P}(X_P))'$
$\mathbf{U}^\star$	$= (U_1^\star, U_2^\star, ..., U_P^\star) = (F_{X_1,n}(X_1), F_{X_2,n}(X_2), ..., F_{X_P,n}(X_P))'$
$\mathbf{V}$	$= (V_1, V_2, ..., V_P) = (F_{Y_1}(Y_1), F_{Y_2}(Y_2), ..., F_{Y_Q}(Y_Q))'$
$\mathbf{V}^\star$	$= (V_1^\star, V_2^\star, ..., V_P^\star) = (F_{Y_1,n}(Y_1), F_{Y_2,n}(Y_2), ..., F_{Y_Q,n}(Y_Q))'$
$c_{\mathbf{U}}$	copula density function of random vector $(\mathbf{U})$(or $(\mathbf{X})$)
$c_{\mathbf{V}}$	copula density function of random vector $(\mathbf{V})$(or $(\mathbf{Y})$)
$c_{\mathbf{UV}}$	copula density function of random vector $(\mathbf{U}, \mathbf{V})$(or $(\mathbf{X}, \mathbf{Y})$)
$\mathbf{u}_i$	$= (u_{i1}, u_{i2}, ..., u_{iP}) = (F_{X_1}(x_{i1}), F_{X_2}(x_{i2}), ..., F_{X_P}(x_{iP}))'$
$\mathbf{v}_i$	$= (v_{i1}, v_{i2}, ..., v_{iQ}) = (F_{Y_1}(y_{i1}), F_{Y_2}(y_{i2}), ..., F_{Y_Q}(y_{iQ}))'$
$\mathbf{u}_i^\star$	$= (u_{i1}^\star, u_{i2}^\star, ..., u_{iP}^\star) = (F_{X_1,n}(x_{i1}), F_{X_2,n}(x_{i2}), ..., F_{X_P,n}(x_{iP}))'$
$\mathbf{v}_i^\star$	$= (v_{i1}^\star, v_{i2}^\star, ..., v_{iQ}^\star) = (F_{Y_1,n}(y_{i1}), F_{Y_2,n}(y_{i2}), ..., F_{Y_Q,n}(y_{iQ}))'$
$\Delta u_{ij,p}$	$= u_{jp}^\star - u_{ip}^\star - u_{jp} + u_{ip}$
$\Delta v_{ij,q}$	$= v_{jq}^\star - v_{iq}^\star - v_{jq} + v_{iq}$
$\frac{\mathbf{u}}{h}$	$= (\frac{u_1}{h}, \frac{u_2}{h}, ..., \frac{u_P}{h})'$
$\frac{\mathbf{v}}{h}$	$= (\frac{v_1}{h}, \frac{v_2}{h}, ..., \frac{v_Q}{h})'$
$K(u)$	kernel function for univariate kernel density estimation, a symmetric density function
$K_h(u)$	$= \frac{1}{h}K(\frac{u}{h})$
$\mathbf{K}^P(\mathbf{u})$	$= \prod_{p=1}^{P} K(u_p)$, kernel function of P-dimensional kernel density estimation
$\mathbf{K}^Q(\mathbf{v})$	$= \prod_{q=1}^{P} K(v_q)$, kernel function of Q-dimensional kernel density estimation
$\mathbf{K}^P_{\mathbf{H}}(\mathbf{u})$	$= \lvert\mathbf{H}\rvert^{-1/2}\mathbf{K}^\mathbf{P}(\mathbf{H}^{-1/2}\mathbf{u})$

Table S1. Continued from previous page

Notation	Meaning
$\mathbf{B_U}$	$=diag(b^2_{\mathrm{U}_1}, b^2_{\mathrm{U}_2}, ..., b^2_{\mathrm{U}_P})$
$\mathbf{B_V}$	$=diag(b^2_{\mathrm{V}_1}, b^2_{\mathrm{V}_2}, ..., b^2_{\mathrm{V}_Q})$
$\mathbf{H_U}$	$=diag(h^2_{\mathrm{U}_1}, h^2_{\mathrm{U}_2}, ..., h^2_{\mathrm{U}_P})$
$\mathbf{H_V}$	$=diag(h^2_{\mathrm{V}_1}, h^2_{\mathrm{V}_2}, ..., h^2_{\mathrm{V}_Q})$
$\mathbf{I}_k$	$k \times k$ identity matrix
$\mathbf{1}_A$	indicator function of set A
$\vert A\vert$	determinant of matrix A
$\Vert f\Vert_{\sup}$	supremum norm of function f
$\Vert \mathbf{t}\Vert$	Euclidean norm of vector $\mathbf{t}$
$\hat{c}^{\setminus i}_{\mathbf{U},\mathbf{H_U}}(\mathbf{u})$	$= \frac{1}{n-1}\sum_{j\neq i} \mathbf{K}^P_{\mathbf{H_U}}(\mathbf{u}^\star_j - \mathbf{u})$, leave-one-out kernel density estimation for $c_\mathbf{U}$
$\hat{c}^{\setminus i}_{\mathbf{V},\mathbf{H_V}}(\mathbf{v}))$	$= \frac{1}{n-1}\sum_{j\neq i} \mathbf{K}^Q_{\mathbf{H_V}}(\mathbf{v}^\star_j - \mathbf{v})$, leave-one-out kernel density estimation for $c_\mathbf{V}$
$\hat{c}^{\setminus i}_{\mathbf{UV},\mathbf{B_U},\mathbf{B_V}}(\mathbf{u},\mathbf{v})$	$= \frac{1}{n-1}\sum_{j\neq i} \mathbf{K}^P_{\mathbf{B_U}}(\mathbf{u}^\star_j - \mathbf{u})\mathbf{K}^Q_{\mathbf{B_V}}(\mathbf{v}^\star_j - \mathbf{v})$, leave one out kernel density estimation for $c_\mathbf{UV}$
$\hat{I}_0$	$= \frac{1}{n}\sum_{i=1}^n \log \frac{c_\mathbf{UV}(\mathbf{u}_i,\mathbf{v}_i)}{c_\mathbf{U}(\mathbf{u}_i)c_\mathbf{V}(\mathbf{v}_i)}$
$\hat{I}_2(\mathbf{B_U},\mathbf{B_V},\mathbf{H_U},\mathbf{H_V})$	$= \frac{1}{n}\sum_{i=1}^n \log \frac{\hat{c}^{\setminus i}_{\mathbf{UV},\mathbf{B_U},\mathbf{B_V}}(\mathbf{u}^\star_i,\mathbf{v}^\star_i)}{\hat{c}^{\setminus i}_{\mathbf{U},\mathbf{H_U}}(\mathbf{u}^\star_i)\hat{c}^{\setminus i}_{\mathbf{V},\mathbf{H_V}}(\mathbf{v}^\star_i)}$
$\hat{I}_3(h)$	$= \frac{1}{n}\sum_{i=1}^n \log \frac{\hat{c}^{\setminus i}_{\mathbf{UV},h^2\mathbf{I}_P,h^2\mathbf{I}_Q}(\mathbf{u}^\star_i,\mathbf{v}^\star_i)}{\hat{c}^{\setminus i}_{\mathbf{U},h^2\mathbf{I}_P}(\mathbf{u}^\star_i)\hat{c}^{\setminus i}_{\mathbf{V},h^2\mathbf{I}_Q}(\mathbf{v}^\star_i)}$
$\tilde{I}_3(h)$	$= \frac{1}{n}\sum_{i=1}^n \log \frac{\hat{c}^{\setminus i}_{\mathbf{UV},h^2\mathbf{I}_P,h^2\mathbf{I}_Q}(\mathbf{u}_i,\mathbf{v}_i)}{\hat{c}^{\setminus i}_{\mathbf{U},h^2\mathbf{I}_P}(\mathbf{u}_i)\hat{c}^{\setminus i}_{\mathbf{V},h^2\mathbf{I}_Q}(\mathbf{v}_i)}$
$A_i(u,h)$	$= \int_{-\frac{u}{h}}^{\frac{1-u}{h}} t^i K(t)dt, i = 0, 1, 2, ...$
$\mathbf{A}^P_0(\mathbf{u},h)$	$= \prod_{p=1}^{P} A_0(u_p, h)$
$\mathbf{A}^Q_0(\mathbf{v},h)$	$= \prod_{q=1}^{Q} A_0(v_q, h)$
$L(u,h)$	$= \int_0^1 \frac{K_h(s-u)}{A_0(s,h)}ds$
$\mathbf{L}^P(\mathbf{u},h)$	$= \prod_{p=1}^{P} L(u_p, h)$
$T_1(h)$	$= \frac{1}{n}\sum_{i=1}^n \left[\frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}^\star_j-\mathbf{u}^\star_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}^\star_j-\mathbf{v}^\star_i)}{c_\mathbf{UV}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)} - \frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}^\star_j-\mathbf{u}^\star_i)}{c_\mathbf{U}(\mathbf{u}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)} - \frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}^\star_j-\mathbf{v}^\star_i)}{c_\mathbf{V}(\mathbf{v}_i)\mathbf{A}^Q_0(\mathbf{v}_i,h)} + 1\right]$
$\tilde{T}_1(h)$	$= \frac{1}{n}\sum_{i=1}^n \left[\frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{c_\mathbf{UV}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)} - \frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)}{c_\mathbf{U}(\mathbf{u}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)} - \frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{c_\mathbf{V}(\mathbf{v}_i)\mathbf{A}^Q_0(\mathbf{v}_i,h)} + 1\right]$
$T_2(h)$	$= \frac{1}{n}\sum_{i=1}^n \left\{\left[\frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}^\star_j-\mathbf{u}^\star_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}^\star_j-\mathbf{v}^\star_i)}{c_\mathbf{UV}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)} - 1\right]^2 - \left[\frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}^\star_j-\mathbf{u}^\star_i)}{c_\mathbf{U}(\mathbf{u}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)} - 1\right]^2 - \left[\frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}^\star_j-\mathbf{v}^\star_i)}{c_\mathbf{V}(\mathbf{v}_i)\mathbf{A}^Q_0(\mathbf{v}_i,h)}\right]^2\right\}$
$\tilde{T}_2(h)$	$= \frac{1}{n}\sum_{i=1}^n \left\{\left[\frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{c_\mathbf{UV}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)} - 1\right]^2 - \left[\frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)}{c_\mathbf{U}(\mathbf{u}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)} - 1\right]^2 - \left[\frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{c_\mathbf{V}(\mathbf{v}_i)\mathbf{A}^Q_0(\mathbf{v}_i,h)}\right]^2\right\}$
$T_3(\mathbf{B_U},\mathbf{B_V},\mathbf{H_U},\mathbf{H_V})$	$= \frac{1}{n}\sum_{i=1}^n \left[\frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{\mathbf{B_U}}(\mathbf{u}^\star_j-\mathbf{u}^\star_i)\mathbf{K}^Q_{\mathbf{B_V}}(\mathbf{v}^\star_j-\mathbf{v}^\star_i)}{c_\mathbf{UV}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,\mathbf{B_U})\mathbf{A}^Q_0(\mathbf{v}_i,\mathbf{B_V})} - 1\right]$
$\tilde{T}_3(\mathbf{B_U},\mathbf{B_V},\mathbf{H_U},\mathbf{H_V})$	$= \frac{1}{n}\sum_{i=1}^n \left[\frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{\mathbf{B_U}}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}^Q_{\mathbf{B_V}}(\mathbf{v}_j-\mathbf{v}_i)}{c_\mathbf{UV}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,\mathbf{B_U})\mathbf{A}^Q_0(\mathbf{v}_i,\mathbf{B_V})} - 1\right]$
$T_4(\mathbf{B_U},\mathbf{B_V},\mathbf{H_U},\mathbf{H_V})$	$= \frac{1}{n}\sum_{i=1}^n \left[\frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{\mathbf{B_U}}(\mathbf{u}^\star_j-\mathbf{u}^\star_i)\mathbf{K}^Q_{\mathbf{B_V}}(\mathbf{v}^\star_j-\mathbf{v}^\star_i)}{c_\mathbf{UV}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,\mathbf{B_U})\mathbf{A}^Q_0(\mathbf{v}_i,\mathbf{B_V})} - 1\right]^2$
$\tilde{T}_4(\mathbf{B_U},\mathbf{B_V},\mathbf{H_U},\mathbf{H_V})$	$= \frac{1}{n}\sum_{i=1}^n \left[\frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{\mathbf{B_U}}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}^Q_{\mathbf{B_V}}(\mathbf{v}_j-\mathbf{v}_i)}{c_\mathbf{UV}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,\mathbf{B_U})\mathbf{A}^Q_0(\mathbf{v}_i,\mathbf{B_V})} - 1\right]^2$

B. Estimation procedure.

The calculation of JMI involves a maximization problem with respect to common bandwidth h. As the objective function shows a unimodal shape, the maximization can be easily solved by most existing methods such as the Newton-Raphson method or grid-search. In our calculation, we consider the latter as it is easy to implement, especially when the permutation test is carried out, and is more convenient for parallel computation. Let

$$\mathcal{H}_n = \Big\{c_0 < h_1 < h_2 < \cdots < h_m < c_1\Big\},$$

where $c_0, c_1 > 0$, and $m \to \infty$ with n. As suggested by Theorem 2, theoretically c_0 and c_1 can be proportional to $n^{-1/(P+Q+3)}$. In our calculation, we set $m = 50$ in Eq. (B) and $\mathcal{H}_n = \{h_k = \frac{1}{50}k^2 n^{-1/(P+Q+3)}\hat{\sigma},\ \ k = 1, 2, ..., 50\}$ with $\hat{\sigma}$ being the sample standard deviation of $\frac{1}{n+1}, \frac{2}{n+1}, ..., \frac{n}{n+1}$. i.e. $\hat{\sigma} = \sqrt{\frac{n}{12(n+1)}}$. Then, JMI is calculated as

$$\widehat{JMI}(X, Y) = \max_{h \in \mathcal{H}_n} \hat{I}_3(h).$$

To counteract the negative bias indicated by Theorem 2, we use the average of JMI and the KDE estimator that uses the full data and the same bandwidth of JMI, to calculate MI in the simulation study of estimation efficiency. We find that the choice of the kernel function has little impact on the results. In our calculation, we use kernel function

$$K(t) = \frac{3}{2(1+|t|)^4}$$

and

$$\mathbf{K}^P(\mathbf{t}) = \mathbf{K}^P(t_1, ..., t_p) = \prod_{p=1}^{P} K(t_p).$$

C. Regularity conditions and Constants in Theorems.

C.1. Regularity conditions.

Assumption A.1

(i) $c_{\mathbf{U}}(\mathbf{u})$ *and* $c_{\mathbf{V}}(\mathbf{v})$ *are smooth on* $[0,1]^P$ *and* $[0,1]^Q$*, respectively;*

(ii) $c_{\mathbf{U}}(\mathbf{u})$ *and* $c_{\mathbf{V}}(\mathbf{v})$ *are bounded away from* 0 *and infinity on their supports.*

Assumption A.2

(i) $c_{\mathbf{UV}}(\mathbf{u}, \mathbf{v})$ *is smooth on* $[0,1]^{P+Q}$*;*

(ii) $c_{\mathbf{UV}}(\mathbf{u}, \mathbf{v})$ *is bounded away from* 0 *and infinity on their supports.*

Assumptions A.1 and A.2 are typically used in studying the consistency rate of mutual information estimators (1, 2).

Assumption A.3

When $Q \le P$, $\mathbf{V} = (V_1, V_2, ..., V_Q)' = \mathbf{\Phi}(\mathbf{U}) = (\Phi_1(\mathbf{U}), \Phi_2(\mathbf{U}), ..., \Phi_Q(\mathbf{U}))'$*, where*

(i) $\Phi_q(\mathbf{U})$ *is smooth on* $[0,1]^P$, $q = 1, 2, ..., Q$*;*

(ii) $\Phi_q(\mathbf{U})$ *is bounded away from infinity and has bounded partial derivatives on* $[0,1]^P$, $q = 1, 2, ..., Q$.

When $Q > P$, $\mathbf{U} = (U_1, U_2, ..., U_P)' = \mathbf{\Phi}(\mathbf{V}) = (\Phi_1(\mathbf{V}), \Phi_2(\mathbf{V}), ..., \Phi_P(\mathbf{V}))'$ *where*

(i) $\Phi_p(\mathbf{V})$ *is smooth on* $[0,1]^Q$, $p = 1, 2, ..., P$*;*

(ii) $\Phi_p(\mathbf{V})$ *is bounded away from infinity and has bounded partial derivatives on* $[0,1]^Q$, $p = 1, 2, ..., P$.

Assumption A.3 is the condition for theoretical analysis when $\mathbf{X}$ and $\mathbf{Y}$ are functionally dependent, and thus $\mathbf{U}$ and $\mathbf{V}$ are functionally dependent.

Assumption A.4

(i) $c_{\mathbf{UV}}(\mathbf{u}, \mathbf{v})$ *is bounded away from* 0*;*

(ii) for any $M > 0$, $c^M_{\mathbf{UV}}(\mathbf{u}, \mathbf{v}) = \min[c_{\mathbf{UV}}(\mathbf{u}, \mathbf{v}), M]$ *is absolutely continuous on* $[0,1]^{P+Q}$.

Assumption A.5

(i) There exists a constant K_0 *such that* $||K||_{\sup}$, $||\frac{dK}{dt}||_{\sup}$, $||\frac{d^2K}{dt^2}||_{\sup} < K_0$.

(ii) $\int_{-\infty}^{\infty} K(t)dt = 1$; *and* $\int_{-\infty}^{\infty} tK(t)dt = 0$;

(iii) There exists an $\epsilon_1 > 0$ *such that,* $\int_{-\infty}^{\infty} |t|^{\frac{1}{2}+\epsilon_1} |dK(t)| < \infty$; *and* $\int_{-\infty}^{\infty} |t|^{2+\epsilon_1} K(t)dt) < \infty$;

(iv) when $h = o(1)$, $\int_{-\infty}^{-\frac{C}{h}} tK(t)dt = o(h)$.

Assumption A.5 is commonly used in the analysis of consistency for kernel density estimation, for example (3). Many commonly used kernels, such as the Gaussian kernel and the kernel above, satisfy the assumption.

Assumption A.6

As $n \to \infty$,

(i) $\max(b_{\mathrm{U_p}}, b_{\mathrm{V_q}}, h_{\mathrm{U_p}}, h_{\mathrm{V_q}}) \to 0$ *for* $p = 1, 2, ..., P$ *and* $q = 1, 2, ..., Q$*;*

(ii) $\min\{\frac{nb^{P+Q}_{\mathrm{U_p}}}{\log n}, \frac{nb^{P+Q}_{\mathrm{V_q}}}{\log n}, \frac{nh^{P+Q}_{\mathrm{U_p}}}{\log n}, \frac{nh^{P+Q}_{\mathrm{V_p}}}{\log n}\} \to \infty$ *for* $p = 1, 2, ..., P$ *and* $q = 1, 2, ..., Q$.

Assumption A.7

As $n \to \infty$,

(i) $h \to 0$;

(ii) $\frac{nh^{P+Q}}{\log(n)} \to 0$.

Assumption A.8 *There exists an $\epsilon_2 > 0$ such that*

$$h \in \mathcal{H}_n = \left\{ c_1 n^{-\frac{1}{P+Q+3}} = h_1 < h_2 < ... < h_m = c_2 n^{-\frac{1}{P+Q+3}} \right\}$$

with $m = O(n^{\frac{P+Q}{P+Q+3} - \epsilon_2})$.

Assumptions A.6 to A.8 are conditions on bandwidths to guarantee the consistency of kernel density estimation and our JMI estimation. They are also commonly assumed for kernel estimation; see for example (4, 5) and (6).

C.2. Constants in Theorem 1 and Theorem 2.

C_1 in theorem 1.

$$C_1 = 2\int_0^\infty \frac{tK(t)}{\int_{-t}^\infty K(s)ds} dt. \quad [1]$$

C_2 and C_3 in theorem 2.

$$C_2 = \lim_{h\to 0} \frac{1}{h} \left[\int_{[0,1]^{P+Q}} \left[\sum_{p=1}^{P} \frac{A_1^2(u_p,h)(\frac{\partial c_{\mathbf{UV}}}{\partial u_p}(\mathbf{u},\mathbf{v}))^2}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})A_0^2(\mathbf{u}_p,h)} + \sum_{q=1}^{Q} \frac{A_1^2(v_q,h)(\frac{\partial c_{\mathbf{UV}}}{\partial v_q}(\mathbf{u},\mathbf{v}))^2}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})A_0^2(v_q,h)} \right] d\mathbf{u}d\mathbf{v} \right], \quad [2]$$

and

$$C_3 = \left[\int_{-\infty}^{\infty} K^2(s)ds \right]^{(P+Q)}. \quad [3]$$

D. Models considered in simulation studies.

D.1. Estimation Efficiency. For estimation efficiency, we consider the same experiments as in (7), which include the following nine different models:

Model I: Both X and Y are mixtures of continuous and discrete distributions. The continuous part follows the bivariate normal distribution with mean $(0,0)$ and covariance matrix $\begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$. For the discrete part, $P(X=1,Y=1) = P(X=-1,Y=-1) = 0.45$ and $P(X=1,Y=-1) = P(X=-1,Y=1) = 0.05$. The two distributions are mixed with equal probabilities.

Model II: X is uniformly distributed over integers $\{0,1,2,3,4\}$ and $Y \sim U(X, X+2)$.

Model III: $X = (X_1,X_2,X_3,X_4,X_5)'$ and $Y = (Y_1,Y_2,Y_3,Y_4,Y_5)'$, where (X_1,Y_1), (X_2,Y_2), (X_3,Y_3), (X_4,Y_4) and (X_5,Y_5) follow the same distribution as in model II.

Model IV: $X \sim Exp(1)$, the standard exponential random variable; $Y \sim Possion(x)$ given $X = x$.

Model V: $X \sim Exp(1)$, the standard exponential random variable; Y takes the value 0 with probability 0.15, and $Y \sim Possion(x)$ given $X = x$ with probability $1 - 0.15$.

Model VI: $X = (X_1,X_2,...,X_{20})'$ and $Y = (Y_1,Y_2,...,Y_{20})'$, where (X_1,Y_1), (X_2,Y_2),..., (X_{20},Y_{20}) follow the same distribution as in model I.

Model VII: $X = (X_1,X_2,...,X_{20})'$ and $Y = (Y_1,Y_2,...,Y_{20})'$, where (X_1,Y_1), (X_2,Y_2),..., (X_{20},Y_{20}) follow the same distribution as in model II.

Model VIII: $X = (X_1,X_2,...,X_{20})'$ and $Y = (Y_1,Y_2,...,Y_{20})'$, where (X_1,Y_1), (X_2,Y_2),..., (X_{20},Y_{20}) follow the same distribution as in model IV.

Model IX: $X = (X_1,X_2,...,X_{20})'$ and $Y = (Y_1,Y_2,...,Y_{20})'$, where (X_1,Y_1), (X_2,Y_2),..., (X_{20},Y_{20}) follow the same distribution as in model V.

D.2. Statistical power of independence test.

For independence test, we consider 18 different models used in (8–10) with either additive noise or contaminated noise. For additive noise, the normal noise is additive to Y, with a parameter to control the noise ratio amplitude NR. As for linear models, we consider both normal noise (symmetric) and standardize chi-square noise . Contaminated noise observations are randomly generated from $[a, b] \times [c, d]$ with a, b, c, d being selected to cover the range of (X, Y), and we increase the noise level by increasing the number of contaminated observations. These models are listed bellow.

(a1) linear with symmetric additive noise

$$\begin{aligned} &y = \frac{2}{3}x + NR \times \epsilon; \\ &x \sim N(0,1);\ \epsilon \sim N(0,1); \\ &x \text{ and } \epsilon \text{ are independent;} \end{aligned}$$

(a2) linear with asymmetric additive noise

$$\begin{aligned} &y = 0.05x + NR \times \epsilon; \\ &x \sim N(0,1);\ \epsilon \sim N^2(0,1) - 1; \\ &x \text{ and } \epsilon \text{ are independent;} \end{aligned}$$

(a3) quadratic with additive noise

$$\begin{aligned} &y = \frac{2}{3}x^2 + NR \times \epsilon; \\ &x \sim N(0,1);\ \epsilon \sim N(0,1); \\ &x \text{ and } \epsilon \text{ are independent;} \end{aligned}$$

(a4) circle with additive noise

$$\begin{aligned} &x = 10\cos(2\pi\theta) + NR \times \epsilon; \\ &y = 10\sin(2\pi\theta) + NR \times \xi; \\ &\theta \sim U(0,1);\ \epsilon, \xi \sim N(0,1); \\ &\theta, \epsilon \text{ and } \xi \text{ are independent;} \end{aligned}$$

(a5) spiral circle with additive noise

$$\begin{aligned} &x = 3u\sin(u\pi) + NR \times \epsilon; \\ &y = 3u\cos(u\pi) + NR \times \xi; \\ &u \sim U(0,4);\quad \epsilon, \xi \sim N(0,1); \\ &u, \epsilon \text{ and } \xi \text{ are independent;} \end{aligned}$$

(a6) cloud with additive noise

$$\begin{aligned} &x = 10 \times (floor(5U_1) + U_2) + NR \times \epsilon; \\ &y = 10 \times (2floor(2U_3) + mod(x,2) + U_4) + NR \times \xi; \\ &U_1, U_2, U_3, U_4 \sim U(0,1);\ \epsilon, \xi \sim N(0,1); \\ &U_1, U_2, U_3, U_4, \epsilon \text{ and } \xi \text{ are independent;} \end{aligned}$$

(a7) sine with high frequency

$$y = \sin[(20 + 7NR)\pi x];\ x \sim U(0,1);$$

(a8) diamond with additive noise

$$\begin{aligned} &x = U_1\cos(\frac{\pi}{4}) + U_2 \sim (\frac{\pi}{4}) + NR \times \epsilon; \\ &y = -U_1\sin(\frac{\pi}{4}) + U_2\cos(\frac{\pi}{4}) + NR \times \xi; \\ &U_1, U_2 \sim U(-1,1);\quad \epsilon, \xi \sim N(0,1); \\ &U_1, U_2, \epsilon \text{ and } \xi \text{ are independent;} \end{aligned}$$

(a9) x-para with additive noise

$$y = \begin{cases} 4x^2 + NR \times \epsilon; & \text{if } U \geq \frac{1}{2} \\ -4x^2 + NR \times \epsilon; & \text{otherwise} \end{cases}$$
$$x \sim U(-1,1); \quad U \sim U(0,1); \quad \epsilon \sim N(0,1);$$
$$x, U \text{ and } \epsilon \text{ are independent;}$$

(a10) step function with additive noise

$$y = \begin{cases} 1; & \text{if } (x+\epsilon)_{(\frac{n}{4})} \leq x + \epsilon \leq (x+\epsilon)_{(\frac{3n}{4})} \\ 0; & \text{otherwise} \end{cases}$$
$$x \sim U(-4,4); \quad \epsilon \sim N(0,1);$$
$$x \text{ and } \epsilon \text{ are independent;}$$
$$(x+\epsilon)_{(i)} \text{ is the } i\text{-th order statistics of } (x_i + \epsilon_i)_{i=1}^n;$$

(a11) 5 dimensional conditional variance

$$\mathbf{y} = \sqrt{0.5\mathbf{x}^2 + NR} \times \mathbf{e};$$
$$\mathbf{x} \sim N(\mathbf{0}, \mathbf{I}_5); \quad \mathbf{e} \sim N(\mathbf{0}, \mathbf{I_5});$$
$$\mathbf{x} \text{ and } \mathbf{e} \text{ are independent;}$$

(a12) 5 dimensional logarithm with additive noise

$$\mathbf{y} = \log(\mathbf{x}^2) + NR \times \mathbf{e};$$
$$\mathbf{x} \sim N(\mathbf{0}, \mathbf{I}_5); \quad \mathbf{e} \sim N(\mathbf{0}, \mathbf{I_5});$$
$$\mathbf{x} \text{ and } \mathbf{e} \text{ are independent;}$$

(a13) 5 dimensional polynomial with additive noise 1

$$\mathbf{y} = (y_1, y_2, y_3, y_4, y_5)'$$
$$y_i = x_i + 4x_i^2 + NR \times e_i, \; i = 1, 2$$
$$y_i = NR \times e_i, \; i = 3, 4, 5$$
$$\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)' \sim N(\mathbf{0}, \mathbf{I}_5);$$
$$\mathbf{e} = (e_1, e_2, e_3, e_4, e_5)' \sim N(\mathbf{0}, \mathbf{I_5});$$
$$\mathbf{x} \text{ and } \mathbf{e} \text{ are independent;}$$

(a14) 5 dimensional polynomial with additive noise 2

$$\mathbf{y} = (y_1, y_2, y_3, y_4, y_5)'$$
$$y_i = 3x_i + 2.5x_i^2 + NR \times e_i, \; i = 1, 2$$
$$y_i = NR \times e_i, \; i = 3, 4, 5$$
$$\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)' \sim N(\mathbf{0}, \mathbf{I}_5);$$
$$\mathbf{e} = (e_1, e_2, e_3, e_4, e_5)' \sim N(\mathbf{0}, \mathbf{I}_5);$$
$$\mathbf{x} \text{ and } \mathbf{e} \text{ are independent;}$$

(a15) 20 dimensional multivariate logarithm with additive noise

$$\mathbf{y} = \log(\mathbf{x}^2) + NR \times \mathbf{e};$$
$$\mathbf{x} \sim N(\mathbf{0}, \mathbf{I}_{20}); \quad \mathbf{e} \sim N(\mathbf{0}, \mathbf{I_{20}});$$
$$\mathbf{x} \text{ and } \mathbf{e} \text{ are independent;}$$

(a16) multivariate linear with additive noise

$$\mathbf{y} = (y_1, y_2, y_3, y_4, y_5)'$$
$$y_i = \sum_{j=10i-9}^{10i} x_j + NR \times e_i, \; i = 1, 2, 3, 4, 5$$
$$\mathbf{x} = (x_1, x_2, ..., x_{50})' \sim N(\mathbf{0}, \mathbf{I}_{50});$$
$$\mathbf{e} = (e_1, e_2, e_3, e_4, e_5)' \sim N(\mathbf{0}, \mathbf{I}_5);$$
$$\mathbf{x} \text{ and } \mathbf{e} \text{ are independent;}$$

(c1) linear with contaminated noise

$$\begin{aligned}&\text{sample}: y_0 = \frac{2}{3}x_0; \quad x_0 \sim U(0,1);\\&\text{noise}: \quad x_1 \sim U(-\pi, 6\pi); \quad y_1 \sim U(-3,3);\\&x_0, x_1 \text{ and } y_1 \text{ are independent;}\end{aligned}$$

(c2) quadratic with contaminated noise

$$\begin{aligned}&\text{sample}: y_0 = \frac{2}{3}x_0^2; \quad x_0 \sim N(0,1);\\&\text{noise}: \quad x_1 \sim U(-3,3); \quad y_1 \sim U(-1,7);\\&x_0, x_1 \text{ and } y_1 \text{ are independent;}\end{aligned}$$

(c3) circle with contaminated noise

$$\begin{aligned}\text{sample}: x_0 &= 10\sin(2\pi\theta);\\y_0 &= 10\cos(2\pi\theta); \quad \theta \sim U(0,1);\\\text{noise}: &\quad x_1 \sim U(-12,12); \quad y_1 \sim U(-12,12);\\\theta, x_1 &\text{ and } y_1 \text{ are independent;}\end{aligned}$$

(c4) spiral circle with contaminated noise

$$\begin{aligned}\text{sample}: x_0 &= 3u\sin(u\pi);\\y_0 &= 3u\cos(u\pi);\\u &\sim U(0,4);\\\text{noise}: &\quad x_1 \sim U(-12,10); \quad y_1 \sim U(-12,14);\\u, x_1 &\text{ and } y_1 \text{ are independent;}\end{aligned}$$

(c5) cloud with contaminated noise

$$\begin{aligned}\text{sample}: x_0 &= 10 \times (floor(5U_1));\\y_0 &= 10 \times (2 \times floor(2U_2) + mod(x,2));\\U_1, U_2 &\sim U(0,1);\\floor(x) &\text{ is the maximum integer that does not larger than } x\\\text{noise}: &\quad x_1 \sim U(-5,45); \quad y_1 \sim U(-5,35);\\U_1, U_2, x_1 &\text{ and } y_1 \text{ are independent;}\end{aligned}$$

(c6) sine with contaminated noise

$$\begin{aligned}&\text{sample}: y_0 = \sin(5\pi x_0); \quad x_0 \sim N(0,1);\\&\text{noise}: \quad x_1 \sim U(-3,3); \quad y_1 \sim U(-1,7);\\&x_0, x_1 \text{ and } y_1 \text{ are independent;}\end{aligned}$$

(c7) diamond with contaminated noise

$$\begin{aligned}\text{sample}: x_0 &= U_1\cos(\frac{\pi}{4}) + U_2\sin(\frac{\pi}{4});\\y_0 &= -U_1\sin(\frac{\pi}{4}) + U_2\cos(\frac{\pi}{4});\\U_1, U_2 &\sim U(-1,1); \quad ;\\\text{noise}: &\quad x_1 \sim U(-15,15); \quad y_1 \sim U(-15,15);\\U_1, U_2, x_1 &\text{ and } y_1 \text{ are independent;}\end{aligned}$$

(c8) x-para with contaminated noise

$$\begin{aligned}\text{sample}: y_0 &= \begin{cases} x_0^2; & \text{if } U \geq \frac{1}{2}\\ -x_0^2; & \text{otherwise}\end{cases}\\x_0 &\sim U(-1,1); \quad U \sim U(0,1);\\\text{noise}: &\quad x_1 \sim U(-3,3); \quad y_1 \sim U(-0.5,1.5);\\x_0, x_1 &\text{ and } y_1 \text{ are independent;}\end{aligned}$$

(c9) step function with contaminated noise

$$\begin{aligned}
\text{sample} : y_0 &= \begin{cases} 1; & \text{if } x_{0(\frac{n}{4})} \le x_0 \le x_{0(\frac{3n}{4})} \\ 0; & \text{otherwise} \end{cases} \\
& x_0 \sim U(-4, 4); \\
& x_{0(i)} \text{ is the } i\text{-th order statistics of } (x_{0i})_{i=1}^{n}; \\
\text{noise} : \quad & x_1 \sim U(-5, 5); \quad y_1 \sim U(-0.5, 1.5); \\
& x_0, x_1 \text{ and } y_1 \text{ are independent;}
\end{aligned}$$

(c10) multivariate logarithm with contaminated noise

$$\begin{aligned}
\text{sample} : \mathbf{y}_0 &= \log(\mathbf{x}_0^2); \quad \mathbf{x}_0 \sim N(\mathbf{0}, \mathbf{I}_5); \\
\text{noise} : \quad & \mathbf{x}_1 = (x_{11}, x_{12}, x_{13}, x_{14}, x_{15})'; \\
& \mathbf{y}_1 = (y_{11}, y_{12}, y_{13}, y_{14}, y_{15})' \\
& x_{1i} \sim U(-3, 3); \quad i = 1, 2, 3, 4, 5; \\
& y_{1i} \sim U(-10, 4); \quad i = 1, 2, 3, 4, 5; \\
& \mathbf{x}_0, \mathbf{x}_1 \text{ and } \mathbf{y}_1 \text{ are independent;}
\end{aligned}$$

(c11) 5 dimensional polynomial with contaminated noise 1

$$\begin{aligned}
\text{sample} : \mathbf{y}_0 &= (y_{01}, y_{02}, y_{03}, y_{04}, y_{05})' \\
& y_{0i} = x_{0i} + 4x_{0i}^2, \ i = 1, 2 \\
& y_{0i} = 0, \ i = 3, 4, 5 \\
& \mathbf{x}_0 = (x_{01}, x_{02}, x_{03}, x_{04}, x_{05})' \sim N(\mathbf{0}, \mathbf{I}_5); \\
\text{noise} : \quad & \mathbf{x}_1 = (x_{11}, x_{12}, x_{13}, x_{14}, x_{15})'; \\
& \mathbf{y}_1 = (y_{11}, y_{12}, y_{13}, y_{14}, y_{15})' \\
& x_{1i} \sim U(-2, 2); \quad i = 1, 2, 3, 4, 5; \\
& y_{1i} \sim U(-2, 18); \quad i = 1, 2, 3, 4, 5; \\
& \mathbf{x}_0, \mathbf{x}_1 \text{ and } \mathbf{y}_1 \text{ are independent;}
\end{aligned}$$

(c12) 5 dimensional polynomial with contaminated noise 2

$$\begin{aligned}
\text{sample} : \mathbf{y}_0 &= (y_{01}, y_{02}, y_{03}, y_{04}, y_{05})' \\
& y_{0i} = 3x_{0i} + 2.5x_{0i}^2, \ i = 1, 2 \\
& y_{0i} = 0, \ i = 3, 4, 5 \\
& \mathbf{x}_0 = (x_{01}, x_{02}, x_{03}, x_{04}, x_{05})' \sim N(\mathbf{0}, \mathbf{I}_5); \\
\text{noise} : \quad & \mathbf{x}_1 = (x_{11}, x_{12}, x_{13}, x_{14}, x_{15})'; \\
& \mathbf{y}_1 = (y_{11}, y_{12}, y_{13}, y_{14}, y_{15})' \\
& x_{1i} \sim U(-2, 2); \quad i = 1, 2, 3, 4, 5; \\
& y_{1i} \sim U(-2, 18); \quad i = 1, 2, 3, 4, 5; \\
& \mathbf{x}_0, \mathbf{x}_1 \text{ and } \mathbf{y}_1 \text{ are independent;}
\end{aligned}$$

E. Simulation results that are not reported in the paper.

E.1. estimation efficiency. Here we plot the averaged MSEs against sample sizes plots for model VI to model IX in the figure S1 below. Again, due to the excessive computational complexity of other KDE based methods, only JMI, mixed-KSG and copula-KSG are calculated and plotted.

(*Fig S1 is about here*)

E.2. Statistical power of independence test. In the main text, we only provide the power curves for (a7)-(a16) with additive noises, and models (c4)-c(9) with contaminated noises. The plots for the remaining models are presented in figure S2 below.

(*Fig S2 is about here*)

F. Local power of independence test.

We only consider the case that X and Y are both univariate continuous random variables to justify the extraordinary power. The marginal transformations are then taken to be $U = F_X(X)$ and $V = F_Y(Y)$, where F_X and F_Y are the CDF of X and Y. Let $c_{UV}(u, v)$ be the copula density function of (U, V)(or X, Y). By the properties of mutual information, testing the independence is equivalent to the following test,

$$\mathbf{H_0}: \quad MI = 0 \qquad \text{versus} \qquad \mathbf{H_1}: \quad MI > 0.$$

The proposed JMI estimator is a natural test statistic. In statistical inference, the power is usually evaluated by the local power to investigate the probability of rejecting the null hypothesis under a sequence of local alternatives. We consider the following local tests for JMI,

$$\mathbf{H_0}: \quad MI = 0 \qquad \text{versus} \qquad \mathbf{H_1}: \quad MI \geq \rho_n,$$

where ρ_n is a sequence of positive values.

Theorem A.1 *Let t_α be the critical value for rejecting the null hypothesis at significance level α, i.e. $P(\widehat{JMI} \leq t_\alpha | H_0) \leq \alpha$. Suppose that general regularity conditions A.1, A.5 and A.8 on the kernel, copula and bandwidths hold. Suppose further that the partial derivatives of copula density $c_{UV}(u, v)$ are bounded. Let $\rho_n = Cn^{-3/5}$ for some constant $C > 0$, then, there exists a constant C_α such that, for $0 < \alpha < 1$ and $C > C_\alpha$,*

$$\lim_{n\to\infty} P(\widehat{JMI} > t_\alpha) = 1.$$

It is interesting to note that the local power of testing the dependence is higher than the usual power of the root-n local alternative.

G. JMI for discrete random vectors. Suppose $\mathbf{X} = (X_1, X_2, ..., X_P)'$ and $\mathbf{Y} = (Y_1, Y_2, ..., Y_Q)'$ be two discrete random vectors. Let $\mathcal{X} = \{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_K\}$ and $\mathcal{Y} = \{\mathbf{t}_1, \mathbf{t}_2 ..., \mathbf{t}_L\}$be the regions of $\mathbf{X}$ and $\mathbf{Y}$, respectively. $P_{\mathbf{X}}(\mathbf{s}) = P(\mathbf{X} = \mathbf{s})$, $P_{\mathbf{Y}}(\mathbf{t}) = P(\mathbf{Y} = \mathbf{t})$ and $P_{\mathbf{XY}}(\mathbf{s}, \mathbf{t}) = P(\mathbf{X} = \mathbf{s}, \mathbf{Y} = \mathbf{t})$ be the marginal and joint probability mass functions. Let $n_k = \#(\mathbf{x}_i | \mathbf{x}_i = \mathbf{s}_k)$, $n_l = \#(\mathbf{y}_i | \mathbf{y}_i = \mathbf{t}_l)$ and $n_{kl} = \#((\mathbf{x}_i, \mathbf{y}_i) | \mathbf{x}_i = \mathbf{s}_k, \mathbf{y}_i = \mathbf{t}_l)$ with # represents the number of elements in a set. Denote

$$\begin{aligned} \hat{P}_{\mathbf{XY}}(\mathbf{s}_k, \mathbf{t}_l) \quad &= \quad \frac{n_{kl}}{n}; \\ \hat{P}_{\mathbf{X}}(\mathbf{s}_k) = \frac{n_k}{n} \quad &\text{and} \quad \hat{P}_{\mathbf{Y}}(\mathbf{t}_l) = \frac{n_l}{n}. \end{aligned}$$

Then, under general regularity conditions,

$$\begin{aligned} \hat{I}_3(h) &= \frac{1}{n}\sum_{i=1}^{n} \log\left\{ \frac{\frac{1}{n-1}\sum_{j\neq i} \mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{x}_j - \mathbf{x}_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{y}_j - \mathbf{y}_i)}{\frac{1}{n-1}\sum_{j\neq i} \mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{x}_j - \mathbf{x}_i)\frac{1}{n-1}\sum_{j\neq i} \mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{y}_j - \mathbf{y}_i)} \right\} \\ &= \frac{1}{n}\sum_{k=1}^{K}\sum_{l=1}^{L} n_{kl} \log\left\{ \frac{\frac{n_{kl}-1}{(n-1)h^{P+Q}} + \sum_{(i,j)\neq(k,l)} \frac{n_{ij}}{n-1}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{s}_i - \mathbf{s}_k)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{t}_j - \mathbf{t}_l)}{\left[\frac{n_k-1}{(n-1)h^P} + \sum_{i\neq k}\frac{n_i}{n-1}\mathbf{K}_{h^2\mathbf{I}_P}(\mathbf{s}_i - \mathbf{s}_k)\right]\left[\frac{n_l-1}{(n-1)h^Q} + \sum_{j\neq l}\frac{n_j}{n-1}\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{t}_j - \mathbf{t}_l)\right]} \right\} \\ &= \frac{1}{n}\sum_{k=1}^{K}\sum_{l=1}^{L} n_{kl}\log\left\{ \frac{\frac{n_{kl}-1}{n-1}}{\frac{n_k-1}{n-1}\frac{n_l-1}{n-1}} \right\} + o(n^{-1/2}) \\ &= \sum_{k=1}^{K}\sum_{l=1}^{L} \hat{P}_{\mathbf{XY}}(\mathbf{s}_k, \mathbf{y}_l)\log\left\{ \frac{\hat{P}_{\mathbf{XY}}(\mathbf{s}_k, \mathbf{y}_l)}{\hat{P}_{\mathbf{X}}(\mathbf{s}_k)\hat{P}_{\mathbf{Y}}(\mathbf{y}_l)} \right\} + o(n^{-1/2}). \end{aligned}$$

This implies that, when h is small enough, the estimated mutual information remains unchanged and that our method can achieve the same efficiency as the commonly considered discrete MI estimation:

$$\hat{I_0}'' = \sum_{k=1}^{K}\sum_{l=1}^{L} \hat{P}_{\mathbf{XY}}(\mathbf{s}_k, \mathbf{y}_l)\log\left\{ \frac{\hat{P}_{\mathbf{XY}}(\mathbf{s}_k, \mathbf{y}_l)}{\hat{P}_{\mathbf{X}}(\mathbf{s}_k)\hat{P}_{\mathbf{Y}}(\mathbf{y}_l)} \right\}.$$

In fact, the JMI is also applicable to mixture random vectors.

H. Technical lemmas.

Lemma 1 *Let χ_n^2 be the Chi-squared distribution with degree of freedom n. Then, for $0 \le t \le 1$,*

$$P(|\frac{1}{n}\chi_n^2 - 1| > t) \le 2e^{-\frac{nt^2}{8}}.$$

Lemma 2 *Let $u_i, i = 1, 2, ..., n$ be n IID samples from standard uniform distribution $U(0,1)$. Denote by $u_{(i)}, i = 1, 2, ..., n$ the order statistics. Then, for any $1 \le i, j \le n$ and $0 \le t \le 1$, we have*

$$P\left(\left|u_{(j)} - u_{(i)} - \frac{j-i}{n+1}\right| > \frac{t\sqrt{|j-i|}\log(n)}{n}\right) \le 3e^{-\frac{t^2(\log(n))^2}{25}}.$$

Lemma 3 *For any $p : 1 \le p \le P$ and $0 \le t \le 1$, we have*

$$P\left(\left|u_{jp} - u_{ip} - u_{jp}^\star + u_{ip}^\star\right| > \frac{t\sqrt{|u_{jp}^\star - u_{ip}^\star|}\log(n)}{\sqrt{n}}\right) \le 3e^{-\frac{t^2(\log(n))^2}{25}}.$$

Moreover, when $|u_{jp}^\star - u_{ip}^\star| > n^{-9/10}$,

$$P\left(\left|u_{jp} - u_{ip} - u_{jp}^\star + u_{ip}^\star\right| > \frac{t\sqrt{|u_{jp} - u_{ip}|}(\log(n))}{\sqrt{n}}\right) \le 3e^{-\frac{t^2(\log(n))^2}{36}}.$$

Lemma 4 *Under Assumptions A.2, A.5 and A.8 ,*

$$(I) \qquad \sup_{h\in\mathcal{H}_n}\left[\frac{1}{(n-1)h^{P+1}}\sum_{j\neq i}\left|\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j - \mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j - \mathbf{v}_i)\right|\right] = O(n^{\frac{1}{P+Q+3}}) + o(n^{-\frac{1}{2(P+Q+3)}}\log(n)) \ a.s.;$$

$$(II) \qquad \sup_{h\in\mathcal{H}_n}\left[\frac{1}{(n-1)h^{P+1}}\sum_{j\neq i}\left|\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j - \mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j - \mathbf{v}_i)\right|\sqrt{|u_{jp} - u_{ip}|}\right] = O(n^{\frac{1}{2(P+Q+3)}}) + o(n^{-\frac{1}{P+Q+3}}\log(n)) \ a.s..$$

Lemma 5 *Let $1 \le p, p_1, p_2 \le P$ and $1 \le q \le Q$. Then, under Assumptions A.1, A.2, A.5 and A.8,*

$$\begin{aligned}
(I) \qquad & \sup_{h\in\mathcal{H}_n}\left[\frac{1}{(n-1)h^{P+1}}\sum_{j\neq i}\left|\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j - \mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j - \mathbf{v}_i)\Delta u_{ij,p}\right|\right] \\
= \ & O(n^{-\frac{P+Q+2}{2(P+Q+3)}}\log(n)) + o(n^{-\frac{P+Q+2}{2(P+Q+3)}}\log(n)) \ a.s.; \\
(II) \qquad & \sup_{h\in\mathcal{H}_n}\left[\frac{1}{(n-1)h^{P+2}}\sum_{j\neq i}\left|\frac{\partial^2 \mathbf{K}^P}{\partial u_{p_1}\partial u_{p_2}}(\frac{\mathbf{u}_j - \mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j - \mathbf{v}_i)\Delta u_{ij,p_1}\Delta u_{ij,p_2}\right|\right] \\
= \ & O(n^{-\frac{P+Q+2}{P+Q+3}}(\log(n))^2) + o(n^{-\frac{P+Q+2}{P+Q+3}}(\log(n))^2) \ a.s.; \\
(III) \qquad & \sup_{h\in\mathcal{H}_n}\left[\frac{1}{(n-1)h^{P+Q+2}}\sum_{j\neq i}\left|\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j - \mathbf{u}_i}{h})\frac{\partial \mathbf{K}^Q}{\partial v_q}(\frac{\mathbf{v}_j - \mathbf{v}_i}{h})\Delta u_{ij,p}\Delta v_{ij,q}\right|\right] \\
= \ & O(n^{-\frac{P+Q+2}{P+Q+3}}(\log(n))^2) + o(n^{-\frac{P+Q+2}{P+Q+3}}(\log(n))^2) \ a.s.; \\
(IV) \qquad & \sup_{h\in\mathcal{H}_n}\left[\frac{1}{(n-1)h^{P+Q+1}}\sum_{j\neq i}\left|\mathbf{K}^P(\frac{\mathbf{u}_j - \mathbf{u}_i}{h})\Delta u_{ij,p}\right|\right] \\
= \ & O(n^{-\frac{P-Q+2}{2(P+Q+3)}}\log(n)) + o(n^{-\frac{P-Q+2}{2(P+Q+3)}}\log(n)) \ a.s.; \\
(V) \qquad & \sup_{h\in\mathcal{H}_n}\left[\frac{1}{(n-1)h^{P+Q+1}}\sum_{j\neq i}\left|\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j - \mathbf{u}_i}{h})\Delta u_{ij,p}\right|\right] \\
= \ & O(n^{-\frac{P-Q+2}{2(P+Q+3)}}\log(n)) + o(n^{-\frac{P-Q+2}{2(P+Q+3)}}\log(n)) \ a.s..
\end{aligned}$$

Lemma 6 *Under Assumptions A.1, A.5 and A.8,*

$$(I) \qquad \sup_{h\in\mathcal{H}_n}\left|\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j^\star - \mathbf{u}_i^\star)\mathbf{K}_{h^2\mathbf{I}_Q}(\mathbf{v}_j^\star - \mathbf{v}_i^\star) - \frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j - \mathbf{u}_i)\mathbf{K}_{h^2\mathbf{I}_Q}(\mathbf{v}_j - \mathbf{v}_i)\right| = o(1) \ a.s.;$$

$$(II) \qquad \sup_{h\in\mathcal{H}_n}\left|\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j^\star - \mathbf{u}_i^\star) - \frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j - \mathbf{u}_i)\right| = o(1) \ a.s.;$$

$$(III) \qquad \sup_{h\in\mathcal{H}_n}\left|\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j^\star - \mathbf{v}_i^\star) - \frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j - \mathbf{v}_i)\right| = o(1) \ a.s..$$

Lemma 7 *Under Assumptions A.1, A.2, A.5 and A.7,*

$$(I) \qquad \sup_{\mathbf{u}\in[0,1]^P}\Big|\frac{1}{n}\sum_{j=1}^{n}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u})-E(\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U}-\mathbf{u}))\Big| = o(1)\ a.s..$$

$$(II) \qquad \sup_{\mathbf{v}\in[0,1]^Q}\Big|\frac{1}{n}\sum_{j=1}^{n}\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v})-E(\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{V}-\mathbf{v}))\Big| = o(1)\ a.s..$$

$$(III) \qquad \sup_{\mathbf{u},\mathbf{v}\in[0,1]^{P+Q}}\Big|\frac{1}{n}\sum_{j=1}^{n}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v})-E\big(\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U}-\mathbf{u})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{V}-\mathbf{v})\big)\Big| = o(1)\ a.s..$$

Lemma 8 *Let* $\mathbf{U}$ *and* $\mathbf{V}$ *be two continuous random vectors with absolutely continuous joint and marginal copula densities* $c_{\mathbf{U},\mathbf{V}}(\mathbf{u},\mathbf{v})$, $c_{\mathbf{U}}(\mathbf{u})$ *and* $c_{\mathbf{V}}(\mathbf{v})$. *Let* $\hat{c}^{\setminus i}_{\mathbf{UV},\mathbf{B}_{\mathrm{U}},\mathbf{B}_{\mathrm{V}}}(\mathbf{u},\mathbf{v})$, $\hat{c}^{\setminus i}_{\mathbf{U},\mathbf{H}_{\mathrm{U}}}(\mathbf{u})$ *and* $\hat{c}^{\setminus i}_{\mathbf{V},\mathbf{H}_{\mathrm{V}}}(\mathbf{v})$ *be the leave-one-out kernel density estimators with* $\mathbf{B}_{\mathrm{U}}$, $\mathbf{B}_{\mathrm{V}}$, $\mathbf{H}_{\mathrm{U}}$ *and* $\mathbf{H}_{\mathrm{V}}$ *defined in Table S1. Under Assumptions A.1, A.2, A.5 and A.6, we have*

$$(I) \qquad \sup_{\mathbf{u}\in[0,1]^P}|\hat{c}^{\setminus i}_{\mathbf{U},\mathbf{H}_{\mathrm{U}}}(\mathbf{u})-c_{\mathbf{U}}(\mathbf{u})\prod_{p=1}^{P}A_0(u_p,h_{\mathrm{U_p}})|\to 0,$$

$$(II) \qquad \sup_{\mathbf{v}\in[0,1]^Q}|\hat{c}^{\setminus i}_{\mathbf{V},\mathbf{H}_{\mathrm{V}}}(\mathbf{v})-c_{\mathbf{V}}(\mathbf{v})\prod_{q=1}^{Q}A_0(v_q,h_{\mathrm{V_p}})|\to 0,$$

$$(III) \qquad \sup_{\mathbf{u},\mathbf{v}\in[0,1]^{P+Q}}|\hat{c}^{\setminus i}_{\mathbf{UV},\mathbf{B}_{\mathrm{U}},\mathbf{B}_{\mathrm{V}}}(\mathbf{u},\mathbf{v})-c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})\prod_{p=1}^{P}A_0(u_p,b_{\mathrm{U_p}})\prod_{q=1}^{Q}A_0(v_q,b_{\mathrm{V_p}})|\to 0.$$

Lemma 9 *Let* $1\le p,p_1,p_2\le P$. *Then, under Assumptions A.1, A.2, A.5 and A.7 ,*

$$(I) \qquad E\Big|\frac{\partial\mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p}\Big| = O(n^{-1/2}h^{P+1/2}\log(n));$$

$$(II) \qquad E\Big|\frac{\partial\mathbf{K}^P}{\partial u_{p_1}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p_1}\frac{\partial\mathbf{K}^P}{\partial u_{p_2}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p_2}\Big| = O(n^{-1}h^{P+1}(\log(n))^2);$$

$$(III) \qquad E\Big|\frac{\partial\mathbf{K}^P}{\partial u_{p_1}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p_1}\frac{\partial\mathbf{K}^P}{\partial u_{p_2}}(\frac{\mathbf{u}_k-\mathbf{u}_i}{h})\Delta u_{ik,p_2}\Big| = O(n^{-1}h^{2P+1}(\log(n))^2);$$

$$(IV) \qquad E\Big|\frac{\partial\mathbf{K}^P}{\partial u_{p_1}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p_1}\frac{\partial\mathbf{K}^P}{\partial u_{p_2}}(\frac{\mathbf{u}_l-\mathbf{u}_k}{h})\Delta u_{kl,p_2}\Big| = O(n^{-1}h^{2P+1}(\log(n))^2).$$

Lemma 10 *Let* $1\le p_1,p_2\le P$. *Then, under Assumptions A.1, A.2, A.5, A.7 and that* $w(\mathbf{u}_i)=O(h^2), i=1,2,...,n$, *we have*

$$(I) \qquad Cov\Big[\frac{\frac{\partial\mathbf{K}^P}{\partial u_{p_1}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p_1}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)}w(\mathbf{u}_j),\frac{\frac{\partial\mathbf{K}^P}{\partial u_{p_2}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p_2}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)}w(\mathbf{u}_j)\Big] = O(n^{-1}h^{P+5}(\log(n))^2);$$

$$(II) \qquad Cov\Big[\frac{\frac{\partial\mathbf{K}^P}{\partial u_{p_1}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p_1}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)}w(\mathbf{u}_j),\frac{\frac{\partial\mathbf{K}^P}{\partial u_{p_2}}(\frac{\mathbf{u}_i-\mathbf{u}_k}{h})\Delta u_{ki,p_2}}{c_{\mathbf{U}}(\mathbf{u}_k)\mathbf{A}^P_0(\mathbf{u}_k,h)}w(\mathbf{u}_i)\Big] = O(n^{-1}h^{2P+5}(\log(n))^2);$$

$$(III) \qquad Cov\Big[\frac{\frac{\partial\mathbf{K}^P}{\partial u_{p_1}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p_1}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)}w(\mathbf{u}_j),\frac{\frac{\partial\mathbf{K}^P}{\partial u_{p_2}}(\frac{\mathbf{u}_l-\mathbf{u}_k}{h})\Delta u_{kl,p_2}}{c_{\mathbf{U}}(\mathbf{u}_k)\mathbf{A}^P_0(\mathbf{u}_k,h)}w(\mathbf{u}_l)\Big] = O(n^{-1}h^{2P+5}(\log(n))^2).$$

Lemma 11 *Under Assumptions A.1, A.2, A.5 and A.8 ,*

$$\sup_{h\in\mathcal{H}_n}|T_1(h)-\tilde{T}_1(h)| = o_p(n^{-\frac{3}{P+Q+3}}).$$

Lemma 12 *Under Assumptions A.1, A.2, A.5 and A.8 ,*

$$\sup_{h\in\mathcal{H}_n}|T_2(h)-\tilde{T}_2(h)| = o_p(n^{-\frac{3}{P+Q+3}}).$$

Lemma 13 *Let* Q *be a given positive number,*

$$W_1 = \int_{-1}^{Q} tK(t)dt;$$

$$\begin{aligned} W_2 &= \int_{-\infty}^{\infty} t^2 K(t)dt; \\ W_3 &= \int_{0}^{\infty} tK(t)dt. \end{aligned}$$

Then, as $h \to 0$,

$$\begin{aligned} &(i) \quad W_1 h < \int_0^h \left|\frac{A_1(u,h)}{A_0(u,h)}\right| du \int_0^1 \left|\frac{A_1(u,h)}{A_0(u,h)}\right| du < (2W_2+1)h; \\ &(ii) \quad W_1^2 h < \int_0^h \left|\frac{A_1^2(u,h)}{A_0^2(u,h)}\right| du < \int_0^1 \left|\frac{A_1^2(u,h)}{A_0^2(u,h)}\right| du < (4W_2W_3+1)h; \\ &(iii) \quad 0 < \int_0^1 \left|\frac{A_2(u,h)}{A_0(u,h)}\right| du < \infty. \end{aligned}$$

Lemma 14 *Suppose* $\mathbf{L}^P(\mathbf{u}, h)$ *is defined as in Table S1. Then*

$$\int_{[0,1]^P} |\mathbf{L}^P(\mathbf{u}, h) - 1| d\mathbf{u} = O(h), \quad \text{as} \quad h \to 0.$$

Lemma 15 *Suppose* $\{u_i\}_{i=1}^n$ *are* n *constitute a random sample of* n *observations from the standard uniform distribution. If* $h, b \to 0$ *as* $n \to \infty$*, then*

$$\frac{1}{n}\sum_{i=1}^n [\log A_0(u_i, h) - \log A_0(u_i, b)] = 2(b-h)\int_0^\infty \frac{tK(t)}{\int_{-t}^\infty K(s)ds} dt + o(n^{-\frac{1}{2}}(|b-h|)^{\frac{1}{2}} \log(n)) \ a.s..$$

Lemma 16 *Let* $U_n = \frac{1}{\binom{n}{m}} \sum_c h(X_{i_1}, X_{i_2}, ..., X_{i_m})$ *be a U-statistics and* $\hat{U}_n = \sum_{i=1}^n E(U_n|X_i) - (n-1)E(U_n)$ *be its projection on* $X_1, X_2, ..., X_n$*. Let*

$$U_n = \hat{U}_n + R_n.$$

If $E(U_n - \hat{U}_n)^2 = O(n^{-2})$*, then*

$$R_n = o(n^{-\frac{1}{2}} \log(n)) \ a.s..$$

Lemma 17 *Under Assumptions A.1, A.5 and A.7,*

$$\tilde{T}_1(h) - \theta(h) = o(h^3 + \frac{1}{nh^{P+Q}}) \ a.s.,$$

where

$$\theta(h) = E(\tilde{T}_1(h)) = O(h^4).$$

Furthermore, if Assumption A.8 holds, then

$$\sup_{h \in \mathcal{H}_n} |\tilde{T}_1(h)| = o_P(n^{-\frac{3}{P+Q+3}}).$$

Lemma 18 *Under Assumptions A.1, A.5 and A.7,*

$$T_2(h) = \tilde{\theta}(h) + o(h^3 + \frac{1}{nh^{P+Q}}) \ a.s.,$$

where

$$\tilde{\theta}(h) = E(T_2(h)) = C_2 h^3 + \frac{1}{nh^{(P+Q)}} \Big[\int_{-\infty}^{\infty} K^2(s) ds \Big]^{(P+Q)}$$

with

$$C_2 = \lim_{h \to 0} \frac{1}{h} \Bigg[\int_{[0,1]^{P+Q}} \Bigg[\sum_{p=1}^P \frac{A_1^2(u_p, h)(\frac{\partial c_{\mathbf{UV}}}{\partial u_p}(\mathbf{u}, \mathbf{v}))^2}{c_{\mathbf{UV}}(\mathbf{u}, \mathbf{v}) A_0^2(\mathbf{u}_p, h)} + \sum_{q=1}^Q \frac{A_1^2(v_q, h)(\frac{\partial c_{\mathbf{UV}}}{\partial v_q}(\mathbf{u}, \mathbf{v}))^2}{c_{\mathbf{UV}}(\mathbf{u}, \mathbf{v}) A_0^2(v_q, h)} \Bigg] d\mathbf{u} d\mathbf{v} \Bigg].$$

Furthermore, if Assumption A.8 holds, then

$$\sup_{h \in \mathcal{H}_n} |T_2(h) - \tilde{\theta}(h)| = o_P(n^{-\frac{3}{P+Q+3}}).$$

Lemma 19 *Let Q be a fixed positive number. Write $W_1 = \int_{-1}^{Q} tK(t)dt$, $W_2 = \int_{-\infty}^{\infty} t^2K(t)dt$ and $W_3 = \int_0^{\infty} tK(t)dt$. Let $\mathbf{u}_{-p} = (u_1, u_2, ..., u_{p-1}, u_{p+1}, ..., u_P)$, $\mathbf{u}_{p,c} = (u_1, u_2, ..., u_{p-1}, c, u_{p+1}, ..., u_P)$, $\mathbf{v}_{-q} = (v_1, v_2, ..., v_{q-1}, v_{q+1}, ..., v_Q)$ and $\mathbf{v}_{q,c} = (v_1, v_2, ..., v_{q-1}, c, v_{q+1}, ..., v_Q)$. If*

$$
\begin{aligned}
&M_1 \le c_{\mathbf{UV}}(\mathbf{u},\mathbf{v}) \le M_2 \\
&m_1 = \max_{p,q}\left\{ \int_{[0,1]^{(P+Q-1)}} \left[\frac{\partial c_{\mathbf{UV}}}{\partial u_p}(\mathbf{u}_{p,0},\mathbf{v})\right]^2 d\mathbf{u}_{-p}d\mathbf{v}, \int_{[0,1]^{(P+Q-1)}} \left[\frac{\partial c_{\mathbf{UV}}}{\partial u_p}(\mathbf{u}_{p,1},\mathbf{v})\right]^2 d\mathbf{u}_{-p}d\mathbf{v}, \right. \\
&\qquad\qquad \left. \int_{[0,1]^{(P+Q-1)}} \left[\frac{\partial c_{\mathbf{UV}}}{\partial v_q}(\mathbf{u},\mathbf{v}_{q,0})\right]^2 d\mathbf{u}d\mathbf{v}_{-q}, \int_{[0,1]^{(P+Q-1)}} \left[\frac{\partial c_{\mathbf{UV}}}{\partial v_q}(\mathbf{u},\mathbf{v}_{q,1})\right]^2 d\mathbf{u}d\mathbf{v}_{-q} \right\}, \\
&m_2 = \min_{p,q}\left\{ \sup_{u_p>0}\left[\int_{[0,1]^{(P+Q-1)}} [\frac{\partial c_{\mathbf{UV}}}{\partial \mathbf{u}}(\mathbf{u}_{p,0},\mathbf{v})]^2 d\mathbf{u}_{-p}d\mathbf{v}\right], \sup_{v_q>0}\left[\int_{[0,1]^{P+Q-1}} [\frac{\partial c_{\mathbf{UV}}}{\partial v_q}(\mathbf{u},\mathbf{v}_{q,0})]^2 d\mathbf{u}d\mathbf{v}_{-q}\right] \right\},
\end{aligned}
$$

then the constant C_1 in Theorem 2 satisfies,

$$\frac{m_1 W_1^2}{M_2} - \epsilon < C_1 < \frac{2m_2^2(4W_2W_3+1)}{M_1}.$$

as $h \to 0$. Here ϵ is any positive value.

Lemma 20 *Under Assumptions A.1, A.4 and A.8, for $M > 2$,*

$$\sup_{(\mathbf{u},\mathbf{v})}\left| \min\left[\frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v})}{\mathbf{A}^P_0(\mathbf{u},h)\mathbf{A}^Q_0(\mathbf{v},h)}, M\right] - \min\left[c_{\mathbf{UV}}(\mathbf{u},\mathbf{v}), M\right]\right| = o(1).$$

I. Proofs of Theorems.

Theorem 1 *Under regularity conditions (A.1, A.5 and A.6) on the kernel function K, the copula density $c_{\mathbf{U}}$, $c_{\mathbf{V}}$ and $c_{\mathbf{UV}}$ and bandwidth matrices, we have*

$$\frac{1}{n}\sum_{i=1}^n \log \hat{c}^{\setminus i}_{\mathbf{UV},\mathbf{B_U},\mathbf{B_V}}(\mathbf{u}_i^\star,\mathbf{v}_i^\star) - \frac{1}{n}\sum_{i=1}^n \log c_{\mathbf{UV}}(\mathbf{u},\mathbf{v}) = -C_1\left[\sum_{p=1}^P b_{\mathrm{U}_p} + \sum_{q=1}^Q b_{\mathrm{V}_q}\right] + o_P\left(\sum_{p=1}^P |b_{\mathrm{U}_p}| + \sum_{q=1}^Q |b_{\mathrm{V}_q}|\right),$$

with $C_1 = 2\int_0^\infty \frac{tK(t)}{\int_{-t}^\infty K(s)ds}dt$. Furthermore,

$$\hat{I}_2(\mathbf{H_U},\mathbf{H_V},\mathbf{B_U},\mathbf{B_V}) - \hat{I}_0 = C_1\left[\sum_{p=1}^P (h_{\mathrm{U}_p} - b_{\mathrm{U}_p}) + \sum_{q=1}^Q (h_{\mathrm{V}_q} - b_{\mathrm{V}_q})\right] + o_P\left(\sum_{p=1}^P (|h_{\mathrm{U}_p}| + |b_{\mathrm{U}_p}|) + \sum_{q=1}^Q (|h_{\mathrm{V}_q}| + |b_{\mathrm{V}_q}|)\right).$$

Proof Using the same technique for proving lemmas 13, 14, 17 and 18, we can derive that

$$
\begin{aligned}
|T_3(\mathbf{B_U},\mathbf{B_V},\mathbf{H_U},\mathbf{H_V}) - \tilde{T}_3(\mathbf{B_U},\mathbf{B_V},\mathbf{H_U},\mathbf{H_V})| &= o_P\left(\sum_{p=1}^P |b_{\mathrm{U}_p}| + \sum_{q=1}^Q |b_{\mathrm{V}_q}|\right); \\
|T_4(\mathbf{B_U},\mathbf{B_V},\mathbf{H_U},\mathbf{H_V}) - \tilde{T}_4(\mathbf{B_U},\mathbf{B_V},\mathbf{H_U},\mathbf{H_V})| &= o_P\left(\sum_{p=1}^P |b_{\mathrm{U}_p}| + \sum_{q=1}^Q |b_{\mathrm{V}_q}|\right); \\
|\tilde{T}_3(\mathbf{B_U},\mathbf{B_V},\mathbf{H_U},\mathbf{H_V})| &= o_P\left(\sum_{p=1}^P |b_{\mathrm{U}_p}| + \sum_{q=1}^Q |b_{\mathrm{V}_q}|\right); \\
|\tilde{T}_4(\mathbf{B_U},\mathbf{B_V},\mathbf{H_U},\mathbf{H_V})| &= o_P\left(\sum_{p=1}^P |b_{\mathrm{U}_p}| + \sum_{q=1}^Q |b_{\mathrm{V}_q}|\right).
\end{aligned}
$$

By Taylor expansion and Lemma 8,

$$\frac{1}{n}\sum_{i=1}^n \log \hat{c}^{\setminus i}_{\mathbf{UV},\mathbf{B_U},\mathbf{B_V}}(\mathbf{u}_i^\star,\mathbf{v}_i^\star) - \frac{1}{n}\sum_{i=1}^n \log c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})$$

$$= \frac{1}{n}\sum_{i=1}^{n}\log \hat{c}^{\backslash i}_{\mathbf{UV},\mathbf{B_U},\mathbf{B_V}}(\mathbf{u}_i^\star,\mathbf{v}_i^\star) - \frac{1}{n}\sum_{i=1}^{n}\log c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\prod_{p=1}^{P}A_0(u_i,b_{\mathrm{U}_p})\prod_{q=1}^{Q}A_0(v_i,b_{\mathrm{V}_p})$$

$$+\frac{1}{n}\sum_{i=1}^{n}\sum_{p=1}^{P}\left[\log A_0(u_i,b_{\mathrm{U}_p})\right] + \frac{1}{n}\sum_{i=1}^{n}\sum_{q=1}^{Q}\left[\log A_0(v_i,b_{\mathrm{V}_q})\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{p=1}^{P}\left[\log A_0(u_i,b_{\mathrm{U}_p})\right] + \frac{1}{n}\sum_{i=1}^{n}\sum_{q=1}^{Q}\left[\log A_0(v_i,b_{\mathrm{V}_q})\right]$$

$$+T_3(\mathbf{B_U},\mathbf{B_V},\mathbf{H_U},\mathbf{H_V}) - \left[\frac{1}{2}+o(1)\right]T_4(\mathbf{B_U},\mathbf{B_V},\mathbf{H_U},\mathbf{H_V})$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{p=1}^{P}\left[\log A_0(u_i,b_{\mathrm{U}_p})\right] + \frac{1}{n}\sum_{i=1}^{n}\sum_{q=1}^{Q}\left[\log A_0(v_i,b_{\mathrm{V}_q})\right]$$

$$+\tilde{T}_3(\mathbf{B_U},\mathbf{B_V},\mathbf{H_U},\mathbf{H_V}) - \frac{1}{2}\tilde{T}_4(\mathbf{B_U},\mathbf{B_V},\mathbf{H_U},\mathbf{H_V}) + o_P\left(\sum_{p=1}^{P}|b_{\mathrm{U}_p}| + \sum_{q=1}^{Q}|b_{\mathrm{V}_q}|\right)$$

$$= -2\int_0^\infty \frac{tK(t)}{\int_{-t}^\infty K(s)ds}dt\left[\sum_{p=1}^{P}b_{\mathrm{U}_p} + \sum_{q=1}^{Q}b_{\mathrm{V}_q}\right] + o_P\left(\sum_{p=1}^{P}|b_{\mathrm{U}_p}| + \sum_{q=1}^{Q}|b_{\mathrm{V}_q}|\right).$$

Here the last equality follows from Lemma 15. Similarly,

$$\frac{1}{n}\sum_{i=1}^{n}\log \hat{c}^{\backslash i}_{\mathbf{U},\mathbf{H_U}}(\mathbf{u}_i^\star) - \frac{1}{n}\sum_{i=1}^{n}\log c_{\mathbf{U}}(\mathbf{u}) = -C_1\left[\sum_{p=1}^{P}h_{\mathrm{U}_p}\right] + o_P\left(\sum_{p=1}^{P}|h_{\mathrm{U}_p}|\right),$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\log \hat{c}^{\backslash i}_{\mathbf{V},\mathbf{H_V}}(\mathbf{v}_i^\star) - \frac{1}{n}\sum_{i=1}^{n}\log c_{\mathbf{V}}(\mathbf{v}) = -C_1\left[\sum_{q=1}^{Q}h_{\mathrm{V}_q}\right] + o_P\left(\sum_{q=1}^{Q}|h_{\mathrm{V}_q}|\right).$$

Consequently,

$$\begin{aligned}\hat{I}_2(\mathbf{H_U},\mathbf{H_V},\mathbf{B_U},\mathbf{B_V}) - \hat{I}_0 &= \left[\frac{1}{n}\sum_{i=1}^{n}\log \hat{c}^{\backslash i}_{\mathbf{UV},\mathbf{B_U},\mathbf{B_V}}(\mathbf{u}_i^\star,\mathbf{v}_i^\star) - \frac{1}{n}\sum_{i=1}^{n}\log c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})\right] \\ &\quad - \left[\frac{1}{n}\sum_{i=1}^{n}\log \hat{c}^{\backslash i}_{\mathbf{U},\mathbf{H_U}}(\mathbf{u}_i^\star) - \frac{1}{n}\sum_{i=1}^{n}\log c_{\mathbf{U}}(\mathbf{u})\right] - \left[\frac{1}{n}\sum_{i=1}^{n}\log \hat{c}^{\backslash i}_{\mathbf{V},\mathbf{H_V}}(\mathbf{v}_i^\star) - \frac{1}{n}\sum_{i=1}^{n}\log c_{\mathbf{V}}(\mathbf{v})\right] \\ &= C_1\left[\sum_{p=1}^{P}(h_{\mathrm{U}_p}-b_{\mathrm{U}_p}) + \sum_{q=1}^{Q}(h_{\mathrm{V}_q}-b_{\mathrm{V}_q})\right] + o_P\left(\sum_{p=1}^{P}(|h_{\mathrm{U}_p}|+|b_{\mathrm{U}_p}|) + \sum_{q=1}^{Q}(|h_{\mathrm{V}_q}|+|b_{\mathrm{V}_q}|)\right).\end{aligned}$$

□

Theorem 2 *Under general regularity conditions (A.1, A.5 and A.7) on functions K, $c_{\mathbf{U}}$, $c_{\mathbf{V}}$ and bandwidth,*

(a) *if $MI < \infty$ and $c_{\mathbf{UV}}$ satisfies regularity condition A.2 , then*

$$\hat{I}_3(h) - \hat{I}_0 = -\frac{C_2}{2}h^3 - \frac{C_3}{2nh^{P+Q}} + o_P(h^3 + \frac{1}{nh^{P+Q}}),$$

where C_2 and C_3 are two positive constants given. In particular, $C_2 = 0$ when $\mathbf{X}$ and $\mathbf{Y}$ are independent;

(b) *if $MI = \infty$, and that $\mathbf{Y}$ is a function of $\mathbf{X}$ and satisfies regularity condition A.3 , then*

$$\hat{I}_3(h) = -\min(P,Q)\log h + o_P(\log h).$$

Proof For (a), note that the term $n^{-1/2}h\log n$ is always dominated by $h^3 + 1/(nh^{P+Q})$. By Taylor expansion, Lemma 8 , Lemma 11 and Lemma 12, we have

$$\hat{I}_3(h) - \hat{I}_0$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n}\log\left\{\frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^{P}_{h^2\mathbf{I}_P}(\mathbf{u}_j^\star-\mathbf{u}_i^\star)\mathbf{K}^{Q}_{h^2\mathbf{I}_Q}(\mathbf{v}_j^\star-\mathbf{v}_i^\star)}{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^{P}_{h^2\mathbf{I}_P}(\mathbf{u}_j^\star-\mathbf{u}_i^\star)\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^{Q}_{h^2\mathbf{I}_Q}(\mathbf{v}_j^\star-\mathbf{v}_i^\star)}\right\} - \frac{1}{n}\sum_{i=1}^{n}\log\left\{\frac{c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}_0^P(\mathbf{u}_i,h)\mathbf{A}_0^Q(\mathbf{v}_i,h)}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}_0^P(\mathbf{u}_i,h)c_{\mathbf{V}}(\mathbf{v}_i)\mathbf{A}_0^Q(\mathbf{v}_i,h)}\right\}\\
&= \frac{1}{n}\sum_{i=1}^{n}\left\{\log\left[\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^{P}_{h^2\mathbf{I}_P}(\mathbf{u}_j^\star-\mathbf{u}_i^\star)\mathbf{K}^{Q}_{h^2\mathbf{I}_Q}(\mathbf{v}_j^\star-\mathbf{v}_i^\star)\right]-\log[c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}_0^P(\mathbf{u}_i,h)\mathbf{A}_0^Q(\mathbf{v}_i,h)]\right\}\\
&\quad -\frac{1}{n}\sum_{i=1}^{n}\left\{\log\left[\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^{P}_{h^2\mathbf{I}_P}(\mathbf{u}_j^\star-\mathbf{u}_i^\star)\right]-\log\left[c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}_0^P(\mathbf{u}_i,h)\right]\right\}\\
&\quad -\frac{1}{n}\sum_{i=1}^{n}\left\{\log\left[\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^{Q}_{h^2\mathbf{I}_Q}(\mathbf{v}_j^\star-\mathbf{v}_i^\star)\right]-\log\left[c_{\mathbf{V}}(\mathbf{v}_i)\mathbf{A}_0^Q(\mathbf{v}_i,h)\right]\right\}\\
&= T_1(h)+(-\frac{1}{2}+o(1))T_2(h)\\
&= \tilde{T}_1(h)+(-\frac{1}{2}+o(1))\tilde{T}_2(h).
\end{aligned}
$$

Then, by Lemma 17 and Lemma 18,

$$\hat{I}_3(h)-\hat{I}_0 = -\frac{C_2}{2}h^3-\frac{C_3}{2nh^{P+Q}}+o_P(h^3+\frac{1}{nh^{P+Q}})\ a.s..$$

with

$$C_2=\lim_{h\to 0}\frac{1}{h}\left[\int_{[0,1]^{P+Q}}\left[\sum_{p=1}^{P}\frac{A_1^2(u_p,h)(\frac{\partial c_{\mathbf{UV}}}{\partial u_p}(\mathbf{u},\mathbf{v}))^2}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})A_0^2(\mathbf{u}_p,h)}+\sum_{q=1}^{Q}\frac{A_1^2(v_q,h)(\frac{\partial c_{\mathbf{UV}}}{\partial v_q}(\mathbf{u},\mathbf{v}))^2}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})A_0^2(v_q,h)}\right]d\mathbf{u}d\mathbf{v}\right],$$

and

$$C_3=\left[\int_{-\infty}^{\infty}K^2(s)ds\right]^{(P+Q)}.$$

When $\mathbf{X}$ and $\mathbf{Y}$ are independent, we have $c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})\equiv 1$, and

$$\frac{\partial c_{\mathbf{UV}}}{\partial u_p}(\mathbf{u},\mathbf{v})=\frac{\partial c_{\mathbf{UV}}}{\partial v_q}(\mathbf{u},\mathbf{v})=0 \text{ for } p=1,2,...,P,\ \ q=1,2,...,Q.$$

Consequently, $C_2=0$.

When $\mathbf{X}$ and $\mathbf{Y}$ are not independent, the upper and lower bounds for C_2 are given in Lemma 19.

For (b), we assume $Q\le P$ without loss of generality. According to Lemma 6 and the proof of Lemma 7, under Assumptions A.1, A.5 and A.7,

$$\sup_i\left[\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^{P}_{h^2\mathbf{I}_P}(\mathbf{u}_j^\star-\mathbf{u}_i^\star)\mathbf{K}^{Q}_{h^2\mathbf{I}_Q}(\mathbf{v}_j^\star-\mathbf{v}_i^\star)-\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^{P}_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}^{Q}_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)\right]=o_P(1),$$

and

$$\sup_{u,v}\left[\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^{P}_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u})\mathbf{K}^{Q}_{h^2\mathbf{I}_Q}(\mathbf{\Phi}(\mathbf{u}_j)-\mathbf{\Phi}(\mathbf{u}))-E\Big(\mathbf{K}^{P}_{h^2\mathbf{I}_P}(\mathbf{U}-\mathbf{u})\mathbf{K}^{Q}_{h^2\mathbf{I}_Q}(\mathbf{\Phi}(\mathbf{U})-\mathbf{\Phi}(\mathbf{u}))\Big)\right]=o_P(h^{-Q}).$$

Note

$$
\begin{aligned}
&E\Big[\mathbf{K}^{P}_{h^2\mathbf{I}_P}(\mathbf{U}-\mathbf{u})\mathbf{K}^{Q}_{h^2\mathbf{I}_Q}\Big(\mathbf{\Phi}(\mathbf{U})-\mathbf{\Phi}(\mathbf{u})\Big)\Big]\\
&= \int_{[0,1]^P}\mathbf{K}^{P}_{h^2\mathbf{I}_P}(\mathbf{u}_1-\mathbf{u})\mathbf{K}^{Q}_{h^2\mathbf{I}_Q}\Big(\mathbf{\Phi}(\mathbf{u}_1)-\mathbf{\Phi}(\mathbf{u})\Big)c_{\mathbf{U}}(\mathbf{u}_1)d\mathbf{u}_1\\
&= \frac{1}{h^Q}\int_{(\mathbf{u}+h\mathbf{s})\in[0,1]^P}\mathbf{K}^P(\mathbf{s})\mathbf{K}^Q\Big(\frac{\mathbf{\Phi}(\mathbf{u}+h\mathbf{s})-\mathbf{\Phi}(\mathbf{u})}{h}\Big)c_{\mathbf{U}}(\mathbf{u}+h\mathbf{s})d\mathbf{s}\\
&= \frac{1}{h^Q}\int_{(\mathbf{u}+h\mathbf{s})\in[0,1]^P}\mathbf{K}^P(\mathbf{s})\prod_{q=1}^{Q}K\Big(\frac{\Phi_q(\mathbf{u}+h\mathbf{s})-\Phi_q(\mathbf{u})}{h}\Big)c_{\mathbf{U}}(\mathbf{u}+h\mathbf{s})d\mathbf{s}\\
&= \frac{1}{h^Q}\Big[\int_{\mathcal{R}^P}\mathbf{K}^P(\mathbf{s})\prod_{q=1}^{Q}K\Big(\sum_{p=1}^{P}s_p\frac{\partial\Phi_q}{\partial u_p}(\mathbf{u})\Big)c_{\mathbf{U}}(\mathbf{u})d\mathbf{s}+o(h)\Big]
\end{aligned}
$$

$$= \quad C(\mathbf{u})h^{-Q} + o(h^{-Q}).$$

Here, $C(\mathbf{u})$ is bounded away from 0 and infinity for any $\mathbf{u}$ since $\frac{\partial \Phi_q}{\partial u_p}(\mathbf{u})$ is bounded for any p,q and $\mathbf{u}$. Then,

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n} \log\left\{\frac{1}{n-1}\sum_{j\neq i} \mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}^\star_j - \mathbf{u}^\star_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}^\star_j - \mathbf{v}^\star_i)\right\} \\
= & \frac{1}{n}\sum_{i=1}^{n} \log\left\{\frac{1}{n-1}\sum_{j\neq i} \mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j - \mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j - \mathbf{v}_i)\right\} + o_P(1) \\
= & \frac{1}{n}\sum_{i=1}^{n} \log\left\{\frac{1}{n-1}\sum_{j\neq i} \mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j - \mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}\Big(\mathbf{\Phi}(\mathbf{u}_j) - \mathbf{\Phi}(\mathbf{u}_i)\Big)\right\} + o_P(1) \\
= & \frac{1}{n}\sum_{i=1}^{n} \log\left\{E\bigg(\Big[\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U} - \mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}\Big(\mathbf{\Phi}(\mathbf{U}) - \mathbf{\Phi}(\mathbf{u}_i)\Big)\Big]\Big|\mathbf{u}_i\bigg)\right\} + o_P(1) \\
= & \frac{1}{n}\sum_{i=1}^{n} \log\left\{C(\mathbf{u}_i)h^{-Q} + o(h^{-Q})\right\} + o_P(1) \\
= & -Q\log h + \frac{1}{n}\sum_{i=1}^{n} \log\{C(\mathbf{u}_i)\} + o_P(1) \\
= & -Q\log h + o_P(-\log h).
\end{aligned}$$

Follows the same analysis, we have, under Assumptions A.1, A.5 and A.7,

$$\frac{1}{n}\sum_{i=1}^{n} \log\left\{\frac{1}{n-1}\sum_{j\neq i} \mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}^\star_j - \mathbf{u}^\star_i)\right\} = O_P(1),$$

and

$$\frac{1}{n}\sum_{i=1}^{n} \log\left\{\frac{1}{n-1}\sum_{j\neq i} \mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}^\star_j - \mathbf{v}^\star_i)\right\} = O_P(1).$$

In conclusion,

$$\begin{aligned}
& \hat{I}_3(h) \\
= & \frac{1}{n}\sum_{i=1}^{n}\left\{\log\left[\frac{1}{n-1}\sum_{j\neq i} \mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}^\star_j - \mathbf{u}^\star_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}^\star_j - \mathbf{v}^\star_i)\right]\right\} - \frac{1}{n}\sum_{i=1}^{n}\left\{\log\left[\frac{1}{n-1}\sum_{j\neq i} \mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}^\star_j - \mathbf{u}^\star_i)\right]\right\} \\
& -\frac{1}{n}\sum_{i=1}^{n}\left\{\log\left[\frac{1}{n-1}\sum_{j\neq i} \mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}^\star_j - \mathbf{v}^\star_i)\right]\right\} \\
= & -Q\log h + o_P(-\log h).
\end{aligned}$$

□

Theorem 3 *Under general regularity conditions (A.1, A.5 and A.8) on functions K, $c_\mathbf{U}$, $c_\mathbf{V}$ and bandwidth,*

(a) *if $MI < \infty$ and $c_{\mathbf{UV}}$ satisfies regularity condition A.2 , then*

$$\widehat{JMI}(\mathbf{X}, \mathbf{Y}) - \hat{I}_0 = O_P(n^{-3/(P+Q+3)}).$$

(b) *if $MI = \infty$, and that $\mathbf{Y}$ is a function of $\mathbf{X}$ and satisfies regularity condition A.4 , then*

$$\widehat{JMI}(\mathbf{X}, \mathbf{Y}) \to \infty, \quad a.s..$$

Proof For (a), note that, under Assumptions A.1, A.5 and A.8,

$$\begin{aligned}
\widehat{JMI}(\mathbf{X}, \mathbf{Y}) - \hat{I}_0 &= \sup_{h\in\mathcal{H}_n} \hat{I}_3(h) - \hat{I}_0 \\
&= \sup_{h\in\mathcal{H}_n} [T_1(h) - (\frac{1}{2} + o(1))T_2(h)] \\
&= \sup_{h\in\mathcal{H}_n} [\tilde{T}_1(h) - \frac{1}{2}\tilde{T}_2(h)] + o_P(n^{-\frac{3}{P+Q+3}})
\end{aligned}$$

$$
\begin{aligned}
&= \sup_{h\in\mathcal{H}_n}[-\frac{1}{2}\tilde{T}_2(h)] + o_P(n^{-\frac{3}{P+Q+3}}) \\
&= \sup_{h\in\mathcal{H}_n}[-\frac{1}{2}\tilde{\theta}(h)] + o_P(n^{-\frac{3}{P+Q+3}}) \\
&= O_p(n^{-\frac{3}{P+Q+3}}).
\end{aligned}
$$

Here, the third equality follows from Lemma 11 and Lemma 12, the fourth from Lemma 17 and the fifth from Lemma 18, $\tilde{\theta}(h) = C_2h^3 + \frac{C_3}{nh^{P+Q}}$ is defined in Lemma 18.

For (b), let $Z = \log\frac{c_{\mathbf{UV}}(\mathbf{U},\mathbf{V})}{c_{\mathbf{U}}(\mathbf{U})c_{\mathbf{V}}(\mathbf{V})}$ and $Z_N = \log\frac{\min[c_{\mathbf{UV}}(\mathbf{U},\mathbf{V}),N]}{c_{\mathbf{U}}(\mathbf{U})c_{\mathbf{V}}(\mathbf{V})}$, $N = 1,2,\ldots$. Then $E(Z) = MI$ and

$$Z_1 < Z_2 < \cdots < Z_N \to Z.$$

By the monotone convergence theorem,

$$\lim_{N\to\infty} E(Z_N) = MI.$$

When $MI(X,Y) = \infty$, for $\forall M$, there exists an N_1 such that, for $\forall N > N_1$, $E(Z_N) > M$. And there exist $h_0 \in \mathcal{H}_n$ such that, with probability 1,

$$
\begin{aligned}
\widehat{JMI}(\mathbf{XY}) &\geq \hat{I}_3(h_0) \\
&= \frac{1}{n}\sum_{i=1}^{n}\log\frac{\hat{c}^{\setminus i}_{\mathbf{UV},h^2\mathbf{I}_P,h^2\mathbf{I}_Q}(\mathbf{u}_i^\star,\mathbf{v}_i^\star)}{\hat{c}^{\setminus i}_{\mathbf{U},h^2\mathbf{I}_P}(\mathbf{u}_i^\star)\hat{c}^{\setminus i}_{\mathbf{V},h^2\mathbf{I}_Q}(\mathbf{v}_i^\star)} \\
&= \frac{1}{n}\sum_{i=1}^{n}\log\left\{\frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)} + o(1)\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\log\left\{\frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{c_{\mathbf{U}}(\mathbf{u}_i)c_{\mathbf{V}}(\mathbf{v}_i)\mathbf{A}^P(\mathbf{u}_i,h)\mathbf{A}^Q(\mathbf{v}_i,h)} + o(1)\right\} \\
&\geq \frac{1}{n}\sum_{i=1}^{n}\log\left\{\frac{1}{c_{\mathbf{U}}(\mathbf{u}_i)c_{\mathbf{V}}(\mathbf{v}_i)}\min\left[\frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{\mathbf{A}^P(\mathbf{u}_i,h)\mathbf{A}^Q(\mathbf{v}_i,h)}, N\right] + o(1)\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\log\left\{\frac{\min\left[c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i), N\right]}{c_{\mathbf{U}}(\mathbf{u}_i)c_{\mathbf{V}}(\mathbf{v}_i)} + o(1)\right\} \\
&= E[Z_N] + o(1) > M.
\end{aligned}
$$

Here, the third equality follows from Lemma 6, the fourth from Lemma 8, the sixth from Lemma 20 and the second last from the law of large numbers.

Let $M \to \infty$, we have

$$\widehat{JMI} \to \infty \ \ a.s..$$

□

Theorem A.1 *Let t_α be the critical value for rejecting the null hypothesis at significance level α, i.e. $P(\widehat{JMI} \leq t_\alpha|H_0) \leq \alpha$. Suppose that general regularity conditions A.1, A.5 and A.8 on the kernel, copula and bandwidths hold. Suppose further that the partial derivatives of copula density $c_{UV}(u,v)$ are bounded. Let $\rho_n = Cn^{-3/5}$ for some constant $C > 0$, then, there exists a constant C_α such that, for $0 < \alpha < 1$ and $C > C_\alpha$,*

$$\lim_{n\to\infty} P(\widehat{JMI} > t_\alpha) = 1.$$

Proof When $P = Q = 1$, X and Y are independent, θ and $\tilde{\theta}$ in Lemma 17 and Lemma 18 satisfy

$$\theta = 0 \quad \text{and} \quad \tilde{\theta} = \frac{1}{nh^2}\left[\int_{-\infty}^{\infty} K^2(s)ds\right]^2.$$

Consequently,

$$\hat{I}_3(h) = \tilde{T}_1(h) - (1/2 + o(1))\tilde{T}_2(h) = -\frac{1}{2nh^2}\left[\int_{-\infty}^{\infty} K^2(s)ds\right]^2 + o(h^3 + \frac{1}{nh^2}) \ a.s..$$

As $h \propto n^{-1/(P+Q+1)} = n^{-1/5}$, there exists a C'_α such that $t_\alpha < C'_\alpha n^{-3/5}$. Since $MI(X,Y) = Cn^{-3/5} = o(1)$ and the partial derivatives of $c_{UV}(u,v)$ are bounded,

$$\sup_{u,v}\Big|c_{UV}(u,v) - 1\Big| = o(1).$$

Write $g(u,v) = c_{UV}(u,v) - 1$. Then,

$$\begin{aligned} MI(X,Y) &= \int_0^1\int_0^1 c_{UV}(u,v)\log[c_{UV}(u,v)]dudv \\ &= \int_0^1\int_0^1 (1+g(u,v))\Big[g(u,v) - \frac{1}{2}g^2(u,v) + o(g^2(u,v))\Big]dudv \\ &= \frac{1}{2}\int_0^1\int_0^1 g^2(u,v)dudv + o(\int_0^1\int_0^1 g^2(u,v)dudv). \end{aligned}$$

This implies that

$$\int_0^1\int_0^1 g^2(u,v)dudv = O(n^{-\frac{3}{5}}).$$

Consequently,

$$\begin{aligned} &\int_0^1\int_0^1 c_{UV}(u,v)\log^2[c_{UV}(u,v)]dudv \\ =& \int_0^1\int_0^1 (1+g(u,v))\Big[g(u,v) - \frac{1}{2}g^2(u,v) + o(g^2(u,v))\Big]^2 dudv \\ =& \int_0^1\int_0^1 g^2(u,v)dudv + o(\int_0^1\int_0^1 g^2(u,v)dudv) \\ =& O(n^{-\frac{3}{5}}). \end{aligned}$$

Then,

$$\hat{I}_0 = Cn^{-\frac{3}{5}} + o(n^{-\frac{4}{5}}\log n)\ a.s..$$

Combing this result with Lemma 17 and Lemma 18 we have that

$$\begin{aligned} I_3(h_m) &= I_3(h_m) - \hat{I}_0 + \hat{I}_0 \\ &= -C_1h_m^3 - \frac{1}{nh_m^2}\left(\int_0^1 K^2(t)dt\right)^2 + Cn^{-\frac{3}{5}} + o(n^{-\frac{3}{5}})\ a.s. \\ &= \left[C - C_1c_2^3 - c_2^{-2}\left(\int_0^1 K^2(t)dt\right)^2\right]n^{-\frac{3}{5}} + o(n^{-\frac{3}{5}})\ a.s.. \end{aligned}$$

Here,

$$C_1 \le (2M^2+1)\left(4\int_{-\infty}^{\infty} K^2(t)dt\int_0^{\infty} tK(t)dt + 1\right).$$

Let

$$C_\alpha = c_2^3\left[(2M^2+1)\left(4\int_{-\infty}^{\infty} K^2(t)dt\int_0^{\infty} tK(t)dt + 1\right)\right] + c_2^{-2}\left(\int_0^1 K^2(t)dt\right)^2 + C'_\alpha + 1.$$

When $C > C_\alpha$,

$$I_3(h_m) > C'_\alpha n^{-1/5}\ a.s..$$

As $\widehat{JMI}(X,Y) \ge I_3(h_m)$,

$$\begin{aligned} &\lim_{n\to\infty} P(\widehat{JMI}(X,Y) > t_\alpha) \\ \ge& \lim_{n\to\infty} P(I_3(h_m) > t_\alpha) \\ \ge& \lim_{n\to\infty} P(I_3(h_m) > C'_\alpha n^{-1/5}) \\ =& 1. \end{aligned}$$

□

J. Proofs of technical lemmas in section F. There are several results in Lemmas 4 to Lemma 10, we only prove the first one in each of those Lemmas for simplicity. The remainings can be verified similarly.

Proof of Lemma 1. Let $Z_i, i = 1, 2, ..., n$ be n independent standard normal variables. Then for $|\lambda| \le 1/4$,

$$\begin{aligned} E[e^{\lambda(\chi_n^2 - n)}] &= \prod_{i=1}^{n} E[e^{\lambda(Z_i^2-1)}] \\ &= e^{-n\lambda}(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(\frac{1}{2}-\lambda)z^2}dz)^n \\ &= (\frac{e^{-\lambda}}{\sqrt{1-2\lambda}})^n \\ &\le e^{2n\lambda^2}. \end{aligned}$$

By Chebyshev's inequality (11), for $0 \le t \le 1$,

$$\begin{aligned} P(\frac{1}{n}\chi_n^2 - 1 > t) &= P(\chi_n^2 - n > nt) \\ &\le e^{-\lambda nt}E[e^{\lambda(\chi_n^2-n)}] \\ &= e^{2n\lambda^2 - \lambda nt} \\ &\le e^{-\frac{nt^2}{8}}. \end{aligned}$$

Similarly, for $0 \le t \le 1$,

$$P(\frac{1}{n}\chi_n^2 - 1 < -t) \le e^{-\frac{nt^2}{8}}.$$

Then, for $0 \le t \le 1$

$$P(|\frac{1}{n}\chi_n^2 - 1| > t) \le 2e^{-\frac{nt^2}{8}}.$$

□

Proof of Lemma 2. Without loss of generality, we assume $i < j$. On writing $W_{ij} = u_{(j)} - u_{(i)}$, $W_{ij} \sim Beta(j-i, n-j+i+1) = \frac{X}{X+Y}$ with $X \sim \chi^2_{2(j-i)}$ and $Y \sim \chi^2_{2(n-j+i+1)}$. By Lemma 1,

$$\begin{aligned} & P\left(\left|u_{(j)} - u_{(i)} - \frac{j-i}{n+1}\right| > \frac{\sqrt{j-i}(\log n)}{n}t\right) \\ = \;& P\left(\left|\frac{j-i}{n+1}\frac{\frac{1}{2(j-i)}X}{\frac{1}{2(n+1)}(X+Y)} - \frac{j-i}{n+1}\right| > \frac{t\sqrt{j-i}(\log n)}{n}\right) \\ = \;& P\left(\left|\frac{\frac{1}{2(j-i)}X}{\frac{1}{2(n+1)}(X+Y)} - 1\right| > \frac{t(n+1)\log n}{n\sqrt{j-i}}\right) \\ \le \;& P\left(\frac{1}{\frac{1}{2(n+1)}(X+Y)} > \frac{5(n+1)}{4n}\right) + P\left(\left|\frac{1}{2(j-i)}X - \frac{1}{2(n+1)}(X+Y)\right| > \frac{4t\log n}{5\sqrt{j-i}}\right) \\ \le \;& P\left(\frac{1}{\frac{1}{2(n+1)}(X+Y)} > \frac{5}{4}\right) + P\left(\left|\frac{1}{2(j-i)}X - 1\right| > \frac{2t\log n}{5\sqrt{j-i}}\right) + P\left(\left|\frac{1}{2(n+1)}(X+Y) - 1\right| > \frac{2t\log n}{5\sqrt{j-i}}\right) \\ \le \;& P\left(\left|\frac{1}{2(n+1)}\chi^2_{2(n+1)} - 1\right| > \frac{1}{5}\right) + P\left(\left|\frac{1}{2(j-i)}\chi^2_{2(j-i)} - 1\right| > \frac{2t\log n}{5\sqrt{j-i}}\right) + P\left(\left|\frac{1}{2(n+1)}\chi^2_{2(n+1)} - 1\right| > \frac{2t\log n}{5\sqrt{j-i}}\right) \\ \le \;& 2e^{-\frac{n+1}{100}} + 2e^{-\frac{2(j-i)\frac{4t^2(\log n)^2}{25(j-i)}}{8}} + 2e^{-\frac{2(n+1)\frac{4t^2(\log n)^2}{25(j-i)}}{8}} \\ \le \;& 3e^{-\frac{t^2(\log n)^2}{25}}. \end{aligned}$$

□

Proof of Lemma 3. The first result follows directly from Lemma 2.

For the second result, it can be easily verified that, when $|u^\star_{jp} - u^\star_{ip}| > n^{-9/10}$ and $\left|u_{jp} - u_{ip} - u^\star_{jp} + u^\star_{ip}\right| < \frac{\sqrt{|u^\star_{jp} - u^\star_{ip}|}(\log n)}{\sqrt{n}}$,

$$\frac{5}{6}\sqrt{|u^\star_{jp} - u^\star_{ip}|} < |\sqrt{|u_{jp} - u_{ip}|} < \frac{6}{5}\sqrt{|u^\star_{jp} - u^\star_{ip}|}.$$

Then,

$$
\begin{aligned}
& P\left(\left|u_{jp}-u_{ip}-u_{jp}^{\star}+u_{ip}^{\star}\right|>\frac{t\sqrt{|u_{jp}-u_{ip}|}(\log n)}{\sqrt{n}}\right) \\
\leq\ & P\left(\left|u_{jp}-u_{ip}-u_{jp}^{\star}+u_{ip}^{\star}\right|>\frac{5t\sqrt{|u_{jp}^{\star}-u_{ip}^{\star}|}(\log n)}{6\sqrt{n}}\right) \\
\leq\ & 3e^{-\frac{t^2(\log n)^2}{36}}.
\end{aligned}
$$

□

Proof of Lemma 4. As $c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})$ is bounded, for $\forall h\in\mathcal{H}_n$ and $s,t\in[0,1]$,

$$
\mu(h,\mathbf{s},\mathbf{t})=E\Big[\Big|\frac{1}{h^{P+1}}\frac{\partial\mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{s}}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{t})\Big|\Big]=O(\frac{1}{h})=O(n^{\frac{1}{P+Q+3}}),
$$

and

$$
\sigma^2(h,\mathbf{s},\mathbf{t})=Var\Big[\Big|\frac{1}{h^{P+1}}\frac{\partial\mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{s}}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{t})\Big|\Big]=O(\frac{1}{h^{P+Q+2}})=O(n^{\frac{P+Q+2}{P+Q+3}}).
$$

As $\frac{1}{h^{P+1}}\frac{\partial\mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\frac{\mathbf{v}_j-\mathbf{v}_i}{h})<\frac{C}{h^{P+Q+1}}$, by Bernstein's inequality (12), for $\forall u_i,v_i$ and $h\in\mathcal{H}_n$,

$$
\begin{aligned}
& P\left(\Big|\frac{1}{(n-1)h^{P+1}}\sum_{j\neq i}\Big|\frac{\partial\mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)\Big|-\mu(h,\mathbf{u}_i,\mathbf{v}_i)\Big|>\frac{t\log n}{\sqrt{n}h^{\frac{P+Q+2}{2}}}\Big|\mathbf{u}_i,\mathbf{v}_i\right) \\
\leq\ & 2\exp(-\frac{t^2(\log n)^2}{Ch^{P+Q+2}\sigma^2(h,\mathbf{u}_i,\mathbf{v}_i)+C't(\log n)n^{-1/2}h^{-\frac{P+Q}{2}}}) \\
\leq\ & 2\exp(-C''t^2(\log n)^2).
\end{aligned}
$$

Then,

$$
\begin{aligned}
& P\left(\Big|\frac{1}{(n-1)h^{P+1}}\sum_{j\neq i}\Big|\frac{\partial\mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)\Big|-\mu(h,\mathbf{u}_i,\mathbf{v}_i)\Big|>\frac{t\log n}{\sqrt{n}h^{\frac{P+Q+2}{2}}}\right) \\
=\ & E\left[P\left(\Big|\frac{1}{(n-1)h^{P+1}}\sum_{j\neq i}\Big|\frac{\partial\mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)\Big|-\mu(h,\mathbf{u}_i,\mathbf{v}_i)\Big|>\frac{t\log n}{\sqrt{n}h^{\frac{P+Q+2}{2}}}\Big|\mathbf{u}_i,\mathbf{v}_i\right)\right] \\
\leq\ & 2\exp(-Ct^2(\log n)^2).
\end{aligned}
$$

Then,

$$
\begin{aligned}
& P\left(\Big|\sup_{h\in\mathcal{H}_n}\Big[\frac{1}{(n-1)h^{P+1}}\sum_{j\neq i}\Big|\frac{\partial\mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)\Big|-\mu(h,\mathbf{u}_i,\mathbf{v}_i)\Big|\Big]>\frac{t\log n}{\sqrt{n}h^{\frac{P+Q+2}{2}}}\right) \\
\leq\ & \sum_{s=1}^{m}P\left(\Big|\Big[\frac{1}{(n-1)h_s^{P+1}}\sum_{j\neq i}\Big|\frac{\partial\mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h_s})\mathbf{K}^Q_{h_s^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)\Big|-\mu(h_s,\mathbf{u}_i,\mathbf{v}_i)\Big|\Big]>\frac{t\log n}{\sqrt{n}h_s^{\frac{P+Q+2}{2}}}\right) \\
\leq\ & 2n^{\frac{P+Q}{P+Q+3}-\epsilon}\exp(-Ct^2(\log n)^2),
\end{aligned}
$$

which implies that

$$
\Big|\sup_{h\in\mathcal{H}_n}\Big[\frac{1}{(n-1)h^{P+1}}\sum_{j\neq i}\Big|\frac{\partial\mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)\Big|-\mu(h,\mathbf{u}_i,\mathbf{v}_i)\Big|\Big]=o(n^{-\frac{1}{2(P+Q+3)}}(\log n))\ a.s..
$$

Combining with the result that $\sup_{h\in\mathcal{H}_n}\mu(h,\mathbf{u}_i,\mathbf{v}_i)=O(n^{\frac{1}{P+Q+3}})$, we have

$$
\sup_{h\in\mathcal{H}_n}\Big[\frac{1}{(n-1)h^{P+1}}\sum_{j\neq i}\Big|\frac{\partial\mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)\Big|-\mu(h,\mathbf{u}_i,\mathbf{v}_i)\Big|\Big]=O(n^{\frac{1}{P+Q+3}})+o(n^{-\frac{1}{2(P+Q+3)}}(\log n))\ a.s..
$$

□

Proof of Lemma 5.

$$\begin{aligned}
& \Big|\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\frac{\mathbf{v}_j-\mathbf{v}_i}{h})\Delta u_{ij,p}\Big| \\
= & \Big|\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)\Delta u_{ij,p}\Big|\mathbf{1}(|u^\star_{jp}-u^\star_{ip}|\le n^{-9/10})\mathbf{1}\Big(|\Delta u_{ij,p}|\le \frac{\sqrt{|u^\star_{jp}-u^\star_{ip}|}(\log n)}{\sqrt{n}}\Big) \\
& +\Big|\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)\Delta u_{ij,p}\Big|\mathbf{1}(|u^\star_{jp}-u^\star_{ip}|\le n^{-9/10})\mathbf{1}\Big(|\Delta u_{ij,p}|> \frac{\sqrt{|u^\star_{jp}-u^\star_{ip}|}(\log n)}{\sqrt{n}}\Big) \\
& +\Big|\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)\Delta u_{ij,p}\Big|\mathbf{1}(|u^\star_{jp}-u^\star_{ip}|> n^{-9/10})\mathbf{1}\Big(|\Delta u_{ij,p}|\le \frac{\sqrt{|u_{jp}-u_{ip}|}(\log n)}{\sqrt{n}}\Big) \\
& +\Big|\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)\Delta u_{ij,p}\Big|\mathbf{1}(|u^\star_{jp}-u^\star_{ip}|> n^{-9/10})\mathbf{1}\Big(|\Delta u_{ij,p}|> \frac{\sqrt{|u_{jp}-u_{ip}|}(\log n)}{\sqrt{n}}\Big) \\
= & a(n,h)+b(n,h)+c(n,h)+d(n,h).
\end{aligned}$$

For $b(n,h)$ and $d(n,h)$,

$$\begin{aligned}
P(\sup_{h\in\mathcal{H}_n} b(n,h)>n^{-\frac{P+Q+2}{2(P+Q+3)}}\log(n)) &\le P(|\Delta u_{ij,p}|>\frac{\sqrt{|u^\star_{jp}-u^\star_{ip}|}\log(n)}{\sqrt{n}}) \\
&\le 3e^{-\frac{(\log n)^2}{25}};
\end{aligned}$$

and

$$\begin{aligned}
P(\sup_{h\in\mathcal{H}_n} d(n,h)>n^{-\frac{P+Q+2}{2(P+Q+3)}}\log(n)) &\le P(|\Delta u_{ij,p}|>\frac{\sqrt{|u_{jp}-u_{ip}|}\log(n)}{\sqrt{n}}) \\
&\le 3e^{-\frac{(\log n)^2}{36}}.
\end{aligned}$$

Hence,

$$\sup_{h\in\mathcal{H}_n}\Big[\frac{1}{n-1}\sum_{j\ne i} b(n,h)\Big]=o(n^{-\frac{P+Q+2}{2(P+Q+3)}}\log(n))\ a.s.,$$

and

$$\sup_{h\in\mathcal{H}_n}\Big[\frac{1}{n-1}\sum_{j\ne i} d(n,h)\Big]=o(n^{-\frac{P+Q+2}{2(P+Q+3)}}log(n))\ a.s..$$

For $a(n,h)$ and $c(n,h)$, it follows from Lemma 4 that

$$\begin{aligned}
& \sup_{h\in\mathcal{H}_n}\Big[\frac{1}{n-1}\sum_{j\ne i} a(n,h)\Big] \\
\le\ & n^{-19/20}(\log n)\sup_{h\in\mathcal{H}_n}\Big[\frac{1}{(n-1)h^{P+1}}\sum_{j\ne i}\Big|\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)\Big|\Big] \\
=\ & o(n^{-\frac{P+Q+2}{2(P+Q+3)}}\log(n))\ a.s..
\end{aligned}$$

$$\begin{aligned}
& \sup_{h\in\mathcal{H}_n}\Big[\frac{1}{n-1}\sum_{j\ne i} c(n,h)\Big] \\
\le\ & n^{-1/2}\log(n)\sup_{h\in\mathcal{H}_n}\Big[\frac{1}{(n-1)h^{P+1}}\sum_{j\ne i}\Big|\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)\Big|\sqrt{|u_j-u_i|}\Big] \\
=\ & O(n^{-\frac{P+Q+2}{2(P+Q+3)}}\log(n))\quad a.s..
\end{aligned}$$

In summary,

$$\sup_{h\in\mathcal{H}_n}\Big[\frac{1}{(n-1)h^{P+1}}\sum_{j\ne i}\Big|\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)\Delta u_{ij,p}\Big|\Big]=O(n^{-\frac{P+Q+2}{2(P+Q+3)}}(\log n)^2)+o(n^{-\frac{P+Q+2}{2(P+Q+3)}}\log(n))\ a.s..$$

□

Proof of Lemma 6. With out loss of generality, we assume $Q \le P$, then, according to Lemma 5,

$$
\begin{aligned}
& \sup_{h\in\mathcal{H}_n}\Big|\frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I_P}}(\mathbf{u}^\star_j-\mathbf{u}^\star_i)\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}^\star_j-\mathbf{v}^\star_i)-\frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I_P}}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}_j-\mathbf{v}_i)\Big| \\
\le\ & \sup_{h\in\mathcal{H}_n}\Big[\frac{1}{(n-1)h^{P+1}}\sum_{j\neq i}\sum_{p=1}^{P}\Big|\frac{\partial\mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}_j-\mathbf{v}_i)\widetilde{\Delta u}_{ij,p}\Big|\Big] \\
& +\sup_{h\in\mathcal{H}_n}\Big[\frac{1}{(n-1)h^{Q+1}}\sum_{j\neq i}\sum_{q=1}^{Q}\Big|\mathbf{K}^P_{h^2\mathbf{I_Q}}(\mathbf{u}_j-\mathbf{u}_i)\frac{\partial\mathbf{K}^Q}{\partial v_q}(\frac{\mathbf{v}_j-\mathbf{v}_i}{h})\widetilde{\Delta v}_{ij,q}\Big|\Big] \\
\le\ & \sup_{h\in\mathcal{H}_n}\Big\{\frac{K_0^Q}{(n-1)h^{P+Q+1}}\Big[\sum_{j\neq i}\sum_{p=1}^{P}\Big|\frac{\partial\mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\widetilde{\Delta u}_{ij,p}\Big|+\sum_{j\neq i}\sum_{q=1}^{Q}\Big|\mathbf{K}^P(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\widetilde{\Delta v}_{ij,q}\Big|\Big]\Big\} \\
=\ & O(n^{-\frac{P-Q+2}{2(P+Q+3)}}\log(n))+o(n^{-\frac{P-Q+2}{2(P+Q+3)}}\log(n))\ a.s. \\
=\ & o(1)\ a.s.
\end{aligned}
$$

Here, $\widetilde{\Delta u}_{ij,p}=\tilde\theta_p\Delta u_{ij,p}$, $\widetilde{\Delta v}_{ij,q}=\tilde\delta_q\Delta v_{ij,q}$, $1\le p\le P$, $1\le q\le Q$ with $0<\tilde\theta_p,\tilde\delta_q<1$. The second last equality follows from $|\widetilde{\Delta u}_{ij,p}|<|\Delta u_{ij,p}|$ and $|\widetilde{\Delta v}_{ij,q}|<|\Delta v_{ij,q}|$.

□

Proof of Lemma 7. Under Assumption A.1, A.5 and A.7, for any $\mathbf{u}$,

$$
\begin{aligned}
Var(\frac{1}{n}\sum_{j=1}^{n}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u})) &= \frac{1}{n}Var(\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U}-\mathbf{u})) \\
&\le \frac{1}{n}E(\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U}-\mathbf{u}))^2 \\
&= \frac{1}{nh^{2P}}\int_{[0,1]^P}\Big(\mathbf{K}^P(\frac{\mathbf{u}_1-\mathbf{u}}{h})\Big)^2 c_{\mathbf{U}}(\mathbf{u}_1)d\mathbf{u}_1 \\
&= \frac{c_{\mathbf{U}}(\mathbf{u})}{nh^P}\Big(\int_{-\infty}^{\infty}K^2(t)dt\Big)^P+o(\frac{1}{nh^P}).
\end{aligned}
$$

As $\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u})=O(\frac{1}{h^P})$, by Bernstein's inequality (12) ,

$$
\begin{aligned}
& P\Big(\Big|\frac{1}{n}\sum_{j=1}^{n}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u})-E(\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U}-\mathbf{u}))\Big|>(nh^P)^{-\frac{1}{2}}(\log n)t\Big) \\
\le\ & 2\exp\Big\{-\frac{t^2(\log n)^2nh^{-P}}{Cnh^{-P}+C'n^{\frac{1}{2}}h^{-\frac{3P}{2}}t}\Big\}.
\end{aligned}
$$

Consider the set $\mathcal{S}=\Big\{(\frac{i_1}{n+1},\frac{i_2}{n+1},...,\frac{i_P}{n+1})|,i_1,i_2,...,i_P=1,2,...,n\Big\}$.

First,

$$
\begin{aligned}
& P\Big(\sup_{u\in\mathcal{S}}\Big|\frac{1}{n}\sum_{j=1}^{n}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u})-E(\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U}-\mathbf{u}))\Big|>(nh^P)^{-\frac{1}{2}}(\log n)t\Big) \\
\le\ & \sum_{u\in\mathcal{S}}P\Big(\Big|\frac{1}{n}\sum_{j=1}^{n}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u})-E(\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U}-\mathbf{u}))\Big|>(nh^P)^{-\frac{1}{2}}(\log n)t\Big) \\
\le\ & n^P\exp\Big\{-\frac{t^2(\log n)^2nh^{-P}}{Cnh^{-P}+C'n^{\frac{1}{2}}h^{-\frac{3P}{2}}t}\Big\} \\
=\ & o(\frac{1}{n^2}).
\end{aligned}
$$

Which implies,

$$
\sup_{u\in\mathcal{S}}\Big|\frac{1}{n}\sum_{j=1}^{n}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u})-E(\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U}-\mathbf{u}))\Big|=o(1)\ a.s..
$$

Second, for any $\mathbf{u} \in [0,1]^P$, there exists a $\mathbf{u_0} \in \mathcal{S}$ such that $d(\mathbf{u}, \mathbf{u}_0) \leq \frac{\sqrt{P}}{2n}$. As $||\frac{dK}{dt}||_{\sup} < K_0$, we have, for any $\mathbf{u} \in [0,1]^P$, there exists a $\mathbf{u_0} \in \mathcal{S}$ such that

$$\Big|\frac{1}{n}\sum_{j=1}^{n}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j - \mathbf{u}) - \frac{1}{n}\sum_{j=1}^{n}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j - \mathbf{u}_0)\Big| = O(\frac{1}{nh^P}) = o(1),$$

and

$$\Big|E(\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U} - \mathbf{u})) - E(\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U} - \mathbf{u}_0))\Big| = O(\frac{1}{nh^P}) = o(1).$$

Then,

$$\begin{aligned}
& \sup_{u\in[0,1]^P}\Big|\frac{1}{n}\sum_{j=1}^{n}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j - \mathbf{u}) - E(\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U} - \mathbf{u}))\Big| \\
= & \sup_{u\in\mathcal{S}}\Big|\frac{1}{n}\sum_{j=1}^{n}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j - \mathbf{u}) - E(\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U} - \mathbf{u}))\Big| + o(1) \\
= & o(1) \ a.s..
\end{aligned}$$

□

Proof of Lemma 8. Note that

$$\begin{aligned}
& \Big|\hat{c}^{\setminus i}_{\mathbf{U},\mathbf{H}_\mathrm{U}}(\mathbf{u}) - c_\mathbf{U}(\mathbf{u})\prod_{p=1}^{P}A_0(u_p, h_{\mathrm{U_p}})\Big| \\
\leq & \Big|\frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^P_{\mathbf{H}_\mathrm{U}}(\mathbf{u}^\star_j - \mathbf{u}) - \frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^P_{\mathbf{H}_\mathrm{U}}(\mathbf{u}_j - \mathbf{u})\Big| \\
& + \Big|\frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^P_{\mathbf{H}_\mathrm{U}}(\mathbf{u}_j - \mathbf{u}) - E[\mathbf{K}^P_{\mathbf{H}_\mathrm{U}}(\mathbf{U} - \mathbf{u})]\Big| \\
& + \Big|E[\mathbf{K}^P_{\mathbf{H}_\mathrm{U}}(\mathbf{U} - \mathbf{u})] - c_\mathbf{U}(\mathbf{u})\prod_{p=1}^{P}A_0(u_p, h_{\mathrm{U_p}})\Big| \\
= & I(\mathbf{u}) + II(\mathbf{u}) + III(\mathbf{u}).
\end{aligned}$$

First, by Lemma 7,

$$\sup_{\mathbf{u}}|II(\mathbf{u}, h)| = o(1), \quad a.s..$$

Second, for $\forall \mathbf{u}$,

$$\begin{aligned}
E[\mathbf{K}^P_{\mathbf{H}_\mathrm{U}}(\mathbf{U} - \mathbf{u})] &= \int_{[0,1]^P}\mathbf{K}^P_{\mathbf{H}_\mathrm{U}}(\mathbf{u}_1 - \mathbf{u})c_\mathbf{U}(\mathbf{u}_1)d\mathbf{u}_1 \\
&= \int_{(\mathbf{u}+\mathbf{H}_\mathrm{U}^{1/2}\mathbf{s})\in[0,1]^P}\mathbf{K}^P(\mathbf{s})c_\mathbf{U}(\mathbf{u} + \mathbf{H}_\mathrm{U}^{1/2}\mathbf{s})d\mathbf{s} \\
&= \int_{(\mathbf{u}+\mathbf{H}_\mathrm{U}^{1/2}\mathbf{s})\in[0,1]^P}\mathbf{K}^P(\mathbf{s})(c_\mathbf{U}(\mathbf{u}) + o(1))d\mathbf{s} \\
&= c_\mathbf{U}(\mathbf{u})\prod_{p=1}^{P}A_0(u_p, h_{\mathrm{U}_p}) + o(1),
\end{aligned}$$

i.e. $\sup_\mathbf{u}|III(\mathbf{u})| = o(1)$.

It remains to prove that $\sup_\mathbf{u} I(\mathbf{u}) = o(1), \quad a.s..$ Note that, for $\forall \mathbf{u}$,

$$\begin{aligned}
& \Big|\hat{c}^{\setminus i}_{\mathbf{U},\mathbf{B}_\mathrm{U}}(\mathbf{u}) - \frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{\mathbf{H_U}}(\mathbf{u}_j - \mathbf{u})\Big| \\
= & \Big|\frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I_P}}(\mathbf{u}^\star_j - \mathbf{u}) - \frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I_P}}(\mathbf{u}_j - \mathbf{u})\Big| \\
= & \Big|\frac{1 + o(1)}{(n-1)h^{P+1}}\sum_{j\neq i}\sum_{p=1}^{P}\frac{\partial\mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j - \mathbf{u}}{h})|u^\star_{jp} - u_{jp}|\Big|
\end{aligned}$$

$$\leq \frac{C}{(n-1)h^{P+1}} \sum_{j\neq i} \sum_{p=1}^{P} \Big| \frac{\partial \mathbf{K}^P}{\partial u_p} (\frac{\mathbf{u}_j - \mathbf{u}}{h}) |u^\star_{jp} - u_{jp} \Big|.$$

Using the same argument for proving Lemma 5 and Lemma 7 we can derive that, for $\forall p$,

$$\sup_{\mathbf{u}} \Big\{ \frac{1}{(n-1)h^{P+1}} \sum_{j\neq i} \Big| \frac{\partial \mathbf{K}^P}{\partial u_p} (\frac{\mathbf{u}_j - \mathbf{u}}{h}) \Big| \Big\} = O(\frac{1}{h}), \quad a.s..$$

Moreover, by Dvoretzky-Kiefer-Wolfowitz (DKW) inequality (13),

$$\begin{aligned} \sup_{j,p} |u^\star_{jp} - u_{jp}| &= \sup_{j,p} |F_{X_p,n}(x_{jp}) - F_{X_p}(x_{jp})| \\ &= o(\frac{\log n}{\sqrt{n}}), \quad a.s.. \end{aligned}$$

Consequently, $\sup_{\mathbf{u}} |I(\mathbf{u})| = o(1)$, which completes the proof. □

Proof of Lemma 9.

$$\begin{aligned} & \Big| \frac{\partial \mathbf{K}^P}{\partial u_p} (\frac{\mathbf{u}_j - \mathbf{u}_i}{h}) \Delta u_{ij,p} \Big| \\ = & \Big| \frac{\partial \mathbf{K}^P}{\partial u_p} (\frac{\mathbf{u}_j - \mathbf{u}_i}{h}) \Delta u_{ij,p} \Big| \mathbf{1}(|u^\star_{jp} - u^\star_{ip}| \leq n^{-9/10}) \mathbf{1}\Big(|\Delta u_{ij,p}| \leq \frac{\sqrt{|u^\star_{jp} - u^\star_{ip}|} \log(n)}{\sqrt{n}} \Big) \\ & + \Big| \frac{\partial \mathbf{K}^P}{\partial u_p} (\frac{\mathbf{u}_j - \mathbf{u}_i}{h}) \Delta u_{ij,p} \Big| \mathbf{1}(|u^\star_{jp} - u^\star_{ip}| \leq n^{-9/10}) \mathbf{1}\Big(|\Delta u_{ij,p}| > \frac{\sqrt{|u^\star_{jp} - u^\star_{ip}|} \log(n)}{\sqrt{n}} \Big) \\ & + \Big| \frac{\partial \mathbf{K}^P}{\partial u_p} (\frac{\mathbf{u}_j - \mathbf{u}_i}{h}) \Delta u_{ij,p} \Big| \mathbf{1}(|u^\star_{jp} - u^\star_{ip}| > n^{-9/10}) \mathbf{1}\Big(|\Delta u_{ij,p}| \leq \frac{\sqrt{|u_{jp} - u_{ip}|} \log(n)}{\sqrt{n}} \Big) \\ & + \Big| \frac{\partial \mathbf{K}^P}{\partial u_p} (\frac{\mathbf{u}_j - \mathbf{u}_i}{h}) \Delta u_{ij,p} \Big| \mathbf{1}(|u^\star_{jp} - u^\star_{ip}| > n^{-9/10}) \mathbf{1}\Big(|\Delta u_{ij,p}| > \frac{\sqrt{|u_{jp} - u_{ip}|} \log(n)}{\sqrt{n}} \Big) \\ = & a'(n,h) + b'(n,h) + c'(n,h) + d'(n,h). \end{aligned}$$

Note that

$$\begin{aligned} E(a'(n,h)) &= O(n^{-19/20} \log(n)); \\ E(b'(n,h)) &\leq P(|\Delta u_{ij,p}| > \frac{\sqrt{|u^\star_{jp} - u^\star_{ip}|} \log(n)}{\sqrt{n}}) \leq 3e^{-\frac{(\log n)^2}{25}}; \\ E(c'(n,h)) &\leq O(n^{-1/2} h^{P+1/2} (\log n)); \\ E(d'(n,h)) &\leq P(|\Delta u_{ij,p}| > \frac{\sqrt{|u_{jp} - u_{ip}|}(\log n)}{\sqrt{n}}) \leq 3e^{-\frac{(\log n)^2}{36}}. \end{aligned}$$

In summary,

$$E \Big| \frac{\partial \mathbf{K}^P}{\partial u_p} (\frac{\mathbf{u}_j - \mathbf{u}_i}{h}) \Delta u_{ij,p} \Big| = O(n^{-1/2} h^{P+1/2} \log(n)).$$

□

Proof of Lemma 10. According to Lemma 9, for $i = 1, 2$,

$$E \Big| \frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_i}} (\frac{\mathbf{u}_j - \mathbf{u}_i}{h}) \Delta u_{ij,p_i}}{c_{\mathbf{U}}(\mathbf{u}_i) \mathbf{A}^P_0(\mathbf{u}_i, h)} w(\mathbf{u}_j) \Big| = O(n^{-1/2} h^{P+5/2} (\log n));$$

and

$$E \Big| \frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_1}} (\frac{\mathbf{u}_j - \mathbf{u}_i}{h}) \Delta u_{ij,p_1}}{c_{\mathbf{U}}(\mathbf{u}_i) \mathbf{A}^P_0(\mathbf{u}_i, h)} w(\mathbf{u}_j) \frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_1}} (\frac{\mathbf{u}_j - \mathbf{u}_i}{h}) \Delta u_{ij,p_2}}{c_{\mathbf{U}}(\mathbf{u}_i) \mathbf{A}^P_0(\mathbf{u}_i, h)} w(\mathbf{u}_j) \Big| = O(n^{-1} h^{P+5} (\log n)^2).$$

Then,

$$\begin{aligned} & Cov \Big[\frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_1}} (\frac{\mathbf{u}_j - \mathbf{u}_i}{h}) \Delta u_{ij,p_1}}{c_{\mathbf{U}}(\mathbf{u}_i) \mathbf{A}^P_0(\mathbf{u}_i, h)} w(\mathbf{u}_j), \frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_1}} (\frac{\mathbf{u}_j - \mathbf{u}_i}{h}) \Delta u_{ij,p_2}}{c_{\mathbf{U}}(\mathbf{u}_i) \mathbf{A}^P_0(\mathbf{u}_i, h)} w(\mathbf{u}_j) \Big] \\ = & O(n^{-1} h^{P+5} (\log n)^2). \end{aligned}$$

$\square$

Proof of Lemma 11.

$$\begin{aligned}
& T_1(h) - \tilde{T}_1(h) \\
= \quad & \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I_P}}(\mathbf{u}^\star_j-\mathbf{u}^\star_i)\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}^\star_j-\mathbf{v}^\star_i) - \frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I_P}}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}_j-\mathbf{v}_i)}{c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)}\right] \\
& -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I_P}}(\mathbf{u}^\star_j-\mathbf{u}^\star_i) - \frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I_P}}(\mathbf{u}_j-\mathbf{u}_i)}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)}\right] \\
& -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}^\star_j-\mathbf{v}^\star_i) - \frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}_j-\mathbf{v}_i)}{c_{\mathbf{V}}(\mathbf{v}_i)\mathbf{A}^Q_0(\mathbf{v}_i,h)}\right] \\
= \quad & \frac{1}{n(n-1)h^{P+1}}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{p=1}^{P}\left[\frac{\frac{\partial\mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}_j-\mathbf{v}_i)\Delta u_{ij,p}}{c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)} - \frac{\frac{\partial\mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)}\right] \\
& +\frac{1}{n(n-1)h^{Q+1}}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{q=1}^{Q}\left[\frac{\mathbf{K}^P_{h^2\mathbf{I_P}}(\mathbf{u}_j-\mathbf{u}_i)\frac{\partial\mathbf{K}^Q}{\partial v_q}(\frac{\mathbf{v}_j-\mathbf{v}_i}{h})\Delta v_{ij,q}}{c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)} - \frac{\frac{\partial\mathbf{K}^Q}{\partial v_q}(\frac{\mathbf{v}_j-\mathbf{v}_i}{h})\Delta v_{ij,q}}{c_{\mathbf{V}}(\mathbf{v}_i)\mathbf{A}^Q_0(\mathbf{v}_i,h)}\right] \\
& +\frac{1}{2n(n-1)h^{P+2}}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{p_1=1}^{P}\sum_{p_2=1}^{P}\left[\frac{\frac{\partial^2\mathbf{K}^P}{\partial u_{P_1}\partial u_{P_2}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}_j-\mathbf{v}_i)\Delta u'_{ij,p_1}\Delta u'_{ij,p_2}}{c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)}\right] \\
& +\frac{1}{2n(n-1)h^{Q+2}}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{q_1=1}^{Q}\sum_{q_2=1}^{Q}\left[\frac{\mathbf{K}^P_{h^2\mathbf{I_P}}(\mathbf{u}_j-\mathbf{u}_i)\frac{\partial^2\mathbf{K}^Q}{\partial v_{q_1}\partial v_{q2}}(\frac{\mathbf{v}_j-\mathbf{v}_i}{h})\Delta v'_{ij,q_1}\Delta v'_{ij,q2}}{c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)}\right] \\
& +\frac{1}{n(n-1)h^{P+Q+2}}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{p=1}^{P}\sum_{q=1}^{Q}\left[\frac{\frac{\partial\mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\frac{\partial\mathbf{K}^Q}{\partial v_q}(\frac{\mathbf{v}_j-\mathbf{v}_i}{h})\Delta u'_{ij,p}\Delta v'_{ij,q}}{c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)}\right] \\
& -\frac{1}{2n(n-1)h^{P+2}}\left[\frac{\frac{\partial^2\mathbf{K}^P}{\partial u_{P_1}\partial u_{P_2}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u''_{ij,p_1}\Delta u''_{ij,p_2}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)}\right] \\
& -\frac{1}{2n(n-1)h^{Q+2}}\left[\frac{\frac{\partial^2\mathbf{K}^Q}{\partial v_{q_1}\partial v_{q2}}(\frac{\mathbf{v}_j-\mathbf{v}_i}{h})\Delta v''_{ij,q_1}\Delta v''_{ij,q_2}}{c_{\mathbf{V}}(\mathbf{v}_i)\mathbf{A}^Q_0(\mathbf{v}_i,h)}\right] \\
= \quad & A(h)+B(h)+C(h)+D(h)+E(h)+F(h).
\end{aligned}$$

Here $\Delta u'_{ij,p} = \theta'_p\Delta u_{ij,p}$, $\Delta u''_{ij,p} = \theta''_p\Delta u_{ij,p}$, $\Delta v'_{ij,q} = \delta'_q\Delta v_{ij,q}$ and $\Delta v''_{ij,q} = \delta''_q\Delta v_{ij,q}$ with $1 \le p \le P$, $1 \le q \le Q$ and $0 < \theta'_p, \theta''_p, \delta'_q, \delta''_q < 1$. According to Lemma 5,

$$\sup_{h\in\mathcal{H}_n}|C(h)+D(h)+E(h)+F(h)| = o(n^{-\frac{3}{P+Q+3}})\ a.s..$$

As $A(h)$ and $B(h)$ are similar, it remains to prove that

$$\sup_{h\in\mathcal{H}_n}|A(h)| = o_P(n^{-\frac{3}{P+Q+3}}).$$

Write

$$G_1(\mathbf{u}_1,\mathbf{u}_2,...,\mathbf{u}_n,\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_n) = \frac{1}{n(n-1)h^{P+1}}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{p=1}^{P}\frac{\frac{\partial\mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}_j-\mathbf{v}_i)\Delta u_{ij,p}}{c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)},$$

and

$$G_2(\mathbf{u}_1,\mathbf{u}_2,...,\mathbf{u}_n) = \frac{1}{n(n-1)h^{P+1}}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{p=1}^{P}\frac{\frac{\partial\mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)}.$$

Let $f(\mathbf{v}|\mathbf{U}=\mathbf{u})$ be the conditional density of $\mathbf{V}$ given $\mathbf{U}=\mathbf{u}$. Then, $f(\mathbf{v}|\mathbf{u}) = \frac{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})}{c_{\mathbf{U}}(\mathbf{u})}$.

$$E(G_1 - G_2|\mathbf{u_1},\mathbf{u_2},...,\mathbf{u_n})$$

$$
\begin{aligned}
&= \frac{1}{n(n-1)h^{P+1}} \sum_{i=1}^{n}\sum_{j\neq i}\sum_{p=1}^{P} \frac{\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p}}{\mathbf{A}_0^P(\mathbf{u}_i,h)c_{\mathbf{U}}(\mathbf{u}_i)c_{\mathbf{U}}(\mathbf{u}_j)} \int_{[0,1]^{2Q}} \frac{\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}_j-\mathbf{v}_i)c_{\mathbf{UV}}(\mathbf{u}_j,\mathbf{v_j})}{\mathbf{A}_0^Q(\mathbf{v}_i,h)} d\mathbf{v}_i d\mathbf{v}_j \\
&\quad - \frac{1}{n(n-1)h^{P+1}} \sum_{i=1}^{n}\sum_{j\neq i}\sum_{p=1}^{P} \frac{\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}_0^P(\mathbf{u}_i,h)} \\
&= \frac{1}{n(n-1)h^{P+1}} \sum_{i=1}^{n}\sum_{j\neq i}\sum_{p=1}^{P} \frac{\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}_0^P(\mathbf{u}_i,h)} \left[\sum_{q=1}^{Q}\int_{[0,1]^Q} h\frac{A_1(v_q,h)}{A_0(v_q,h)}\frac{\partial c_{\mathbf{UV}}}{\partial v_q}(\mathbf{u}_j,\mathbf{v})d\mathbf{v} + O(h^2)\right] \\
&= \frac{1}{n(n-1)h^{P+1}} \sum_{i=1}^{n}\sum_{j\neq i}\sum_{p=1}^{P} \frac{\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}_0^P(\mathbf{u}_i,h)} w(\mathbf{u}_j).
\end{aligned}
$$

Here, $\forall j,\ |w(\mathbf{u}_j)| = \left[\sum_{q=1}^{Q}\int_{[0,1]^Q} h\frac{A_1(v_q,h)}{A_0(v_q,h)}\frac{\partial c_{\mathbf{UV}}}{\partial v_q}(\mathbf{u}_j,\mathbf{v})d\mathbf{v} + O(h^2)\right] \le Ch^2$. Then,

$$
\begin{aligned}
& E(G_1-G_2) \\
\le\ & E\Big[\frac{1}{n(n-1)h^{P+1}} \sum_{i=1}^{n}\sum_{j\neq i}\sum_{p=1}^{P} \frac{\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}_0^P(\mathbf{u}_i,h)} w(\mathbf{u}_j)\Big] \\
\le\ & E\Big[\frac{C}{n(n-1)h^{P-1}} \sum_{i=1}^{n}\sum_{j\neq i}\sum_{p=1}^{P} \Big|\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Big||\Delta u_{ij,p}|\Big] \\
=\ & O(n^{-1/2}h^{3/2}(\log n)).
\end{aligned}
$$

Note that,

$$
\begin{aligned}
& Var(E(G_1-G_2|u_1,u_2,...,u_n)) \\
=\ & \frac{1}{n^2(n-1)^2h^{2(P+1)}} \sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j\neq i}\sum_{l\neq k}\sum_{p_1=1}^{P}\sum_{p_2=1}^{P} Cov\Big[\frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_1}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p_1}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}_0^P(\mathbf{u}_i,h)} w(\mathbf{u}_j), \frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_2}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p_2}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}_0^P(\mathbf{u}_i,h)} w(\mathbf{u}_j)\Big] \\
=\ & \frac{1}{n^2(n-1)^2h^{2(P+1)}} \sum_{\substack{1\le i,j,k,l\le n\\ i\neq j\neq k\neq l}}\sum_{p_1=1}^{P}\sum_{p_2=1}^{P} Cov\Big[\frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_1}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p_1}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}_0^P(\mathbf{u}_i,h)} w(\mathbf{u}_j), \frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_2}}(\frac{\mathbf{u}_l-\mathbf{u}_k}{h})\Delta u_{kl,p_2}}{c_{\mathbf{U}}(\mathbf{u}_k)\mathbf{A}_0^P(\mathbf{u}_k,h)} w(\mathbf{u}_l)\Big] \\
& + \frac{1}{n^2(n-1)^2h^{2(P+1)}} \sum_{\substack{1\le i,j,k\le n\\ i\neq j\neq k}}\sum_{p_1=1}^{P}\sum_{p_2=1}^{P} Cov\Big[\frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_1}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p_1}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}_0^P(\mathbf{u}_i,h)} w(\mathbf{u}_j), \frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_1}}(\frac{\mathbf{u}_i-\mathbf{u}_k}{h})\Delta u_{ki,p_2}}{c_{\mathbf{U}}(\mathbf{u}_k)\mathbf{A}_0^P(\mathbf{u}_k,h)} w(\mathbf{u}_i)\Big] \\
& + \frac{1}{n^2(n-1)^2h^{2(P+1)}} \sum_{\substack{1\le i,j,k\le n\\ i\neq j\neq k}}\sum_{p_1=1}^{P}\sum_{p_2=1}^{P} Cov\Big[\frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_1}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p_1}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}_0^P(\mathbf{u}_i,h)} w(\mathbf{u}_j), \frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_2}}(\frac{\mathbf{u}_k-\mathbf{u}_i}{h})\Delta u_{ik,p_2}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}_0^P(\mathbf{u}_i,h)} w(\mathbf{u}_k)\Big] \\
& + \frac{1}{n^2(n-1)^2h^{2(P+1)}} \sum_{\substack{1\le i,j,k\le n\\ i\neq j\neq k}}\sum_{p_1=1}^{P}\sum_{p_2=1}^{P} Cov\Big[\frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_1}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p_1}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}_0^P(\mathbf{u}_i,h)} w(\mathbf{u}_j), \frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_2}}(\frac{\mathbf{u}_j-\mathbf{u}_k}{h})\Delta u_{kj,p_2}}{c_{\mathbf{U}}(\mathbf{u}_k)\mathbf{A}_0^P(\mathbf{u}_k,h)} w(\mathbf{u}_j)\Big] \\
& + \frac{1}{n^2(n-1)^2h^{2(P+1)}} \sum_{\substack{1\le i,j,k\le n\\ i\neq j\neq k}}\sum_{p_1=1}^{P}\sum_{p_2=1}^{P} Cov\Big[\frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_1}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p_1}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}_0^P(\mathbf{u}_i,h)} w(\mathbf{u}_j), \frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_2}}(\frac{\mathbf{u}_k-\mathbf{u}_j}{h})\Delta u_{jk,p_2}}{c_{\mathbf{U}}(\mathbf{u}_j)\mathbf{A}_0^P(\mathbf{u}_j,h)} w(\mathbf{u}_k)\Big] \\
& + \frac{1}{n^2(n-1)^2h^{2(P+1)}} \sum_{i=1}^{n}\sum_{j\neq i}\sum_{p_1=1}^{P}\sum_{p_2=1}^{P} Cov\Big[\frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_1}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p_1}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}_0^P(\mathbf{u}_i,h)} w(\mathbf{u}_j), \frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_2}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p_2}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}_0^P(\mathbf{u}_i,h)} w(\mathbf{u}_j)\Big] \\
& + \frac{1}{n^2(n-1)^2h^{2(P+1)}} \sum_{i=1}^{n}\sum_{j\neq i}\sum_{p_1=1}^{P}\sum_{p_2=1}^{P} Cov\Big[\frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_1}}(\frac{\mathbf{u}_j-\mathbf{u}_i}{h})\Delta u_{ij,p_1}}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}_0^P(\mathbf{u}_i,h)} w(\mathbf{u}_j), \frac{\frac{\partial \mathbf{K}^P}{\partial u_{p_2}}(\frac{\mathbf{u}_i-\mathbf{u}_j}{h})\Delta u_{ji,p_2}}{c_{\mathbf{U}}(\mathbf{u}_j)\mathbf{A}_0^P(\mathbf{u}_j,h)} w(\mathbf{u}_i)\Big] \\
=\ & O\left(\frac{h^3(\log n)^2}{n} + \frac{(\log n)^2}{n^3h^{P-3}}\right).
\end{aligned}
$$

Here the last equality follows from Lemma 10. Follow a similar tedious calucation as above we can derive that

$$E(Var(G_1 - G_2|u_1, u_2, ..., u_n)) = O\left(\frac{(\log n)^2}{n^2 h} + \frac{(\log n)^2}{n^3 h^{P+Q+1}}\right).$$

In summary,

$$\begin{aligned} Var(G_1 - G_2) &= E(Var(G_1 - G_2|u_1, u_2, ..., u_n)) + Var(E(G_1 - G_2|u_1, u_2, ...u_n)) \\ &= O\left(\frac{h^3(\log n)^2}{n} + \frac{(\log n)^2}{n^2 h} + \frac{(\log n)^2}{n^3 h^{P+Q+1}}\right). \end{aligned}$$

When $h \propto n^{-1/(P+Q+3)}$, $E(A(h)) = O(n^{-1/2}h^{3/2}(\log n)) = o(n^{-3/P+Q+3})$. By Chebyshev's inequality (11),

$$P(|A(h)| > n^{-3/P+Q+3}t) = O(n^{-\frac{P+Q}{P+Q+3}}(\log n)^2)).$$

Then,

$$P(\sup_{h\in\mathcal{H}_n} |A(h)| > n^{-3/P+Q+3}t) \leq n^{\frac{P+Q}{P+Q+3}-\epsilon} O(n^{-\frac{P+Q}{P+Q+3}}(\log n)^2)) \to 0.$$

In other words, $\sup_{h\in\mathcal{H}_n} |A(h)| = o_P(n^{-3/(P+Q+3)})$. □

Proof of Lemma 12. According to Lemma 5,

$$\begin{aligned} &\sup_{h\in\mathcal{H}_n}\left|\frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I_P}}(\mathbf{u}^\star_j - \mathbf{u}^\star_i)\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}^\star_j - \mathbf{v}^\star_i) - \frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I_P}}(\mathbf{u}_j - \mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}_j - \mathbf{v}_i)\right| \\ \leq\ &\sup_{h\in\mathcal{H}_n}\left[\frac{1}{(n-1)h^{P+1}}\sum_{j\neq i}\sum_{p=1}^{P}\left|\frac{\partial \mathbf{K}^P}{\partial u_p}(\frac{\mathbf{u}_j - \mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}_j - \mathbf{v}_i)\Delta u_{ij,p}\right|\right] \\ &+\sup_{h\in\mathcal{H}_n}\left[\frac{1}{(n-1)h^{Q+1}}\sum_{j\neq i}\sum_{q=1}^{Q}\left|\mathbf{K}^P_{h^2\mathbf{I_Q}}(\mathbf{u}_j - \mathbf{u}_i)\frac{\partial \mathbf{K}^Q}{\partial v_q}(\frac{\mathbf{v}_j - \mathbf{v}_i}{h})\Delta v_{ij,q}\right|\right] \\ &+\sup_{h\in\mathcal{H}_n}\left[\frac{1}{2(n-1)h^{P+2}}\sum_{j\neq i}\sum_{p_1=1}^{P}\sum_{p_2=1}^{P}\left|\frac{\partial^2 \mathbf{K}^P}{\partial u_{p_1}\partial u_{p_2}}(\frac{\mathbf{u}_j - \mathbf{u}_i}{h})\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}_j - \mathbf{v}_i)\Delta u'_{ij,p_1}\Delta u'_{ij,p_2}\right|\right] \\ &+\sup_{h\in\mathcal{H}_n}\left[\frac{1}{2(n-1)h^{Q+2}}\sum_{j\neq i}\sum_{q_1=1}^{Q}\sum_{q_2=1}^{Q}\left|\mathbf{K}^P_{h^2\mathbf{I_P}}(\mathbf{u}_j - \mathbf{u}_i)\frac{\partial^2 \mathbf{K}^Q}{\partial v_{q_1}\partial v_{q_2}}(\frac{\mathbf{v}_j - \mathbf{v}_i}{h})\Delta v'_{ij,q_1}\Delta v'_{ij,q_2}\right|\right] \\ &+\sup_{h\in\mathcal{H}_n}\left[\frac{1}{(n-1)h^{P+Q+2}}\sum_{j\neq i}\sum_{p=1}^{P}\sum_{q=1}^{Q}\left|\frac{\partial \mathbf{K}^P}{\partial u_p}(\mathbf{u}_j - \mathbf{u}_i)\frac{\partial \mathbf{K}^Q}{\partial v_q}(\frac{\mathbf{v}_j - \mathbf{v}_i}{h})\Delta u'_{ij,p}\Delta v'_{ij,q}\right|\right], \\ =\ &O\left(n^{-\frac{P+Q+2}{2(P+Q+3)}}(\log n)\right) + o\left(n^{-\frac{P+Q+2}{2(P+Q+3)}}(\log n)\right)\quad a.s.. \end{aligned}$$

Here $\Delta u'_{ij,p} = \theta'_p \Delta u_{ij,p}$and $\Delta v'_{ij,q} = \delta'_q \Delta v_{ij,q}$ with $1 \leq p \leq P$, $1 \leq q \leq Q$ and $0 < \theta'_p, \delta'_q < 1$.

It can be shown that

$$\sup_{h\in\mathcal{H}_n}\left\{\frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I_P}}(\mathbf{u}_j - \mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}_j - \mathbf{v}_i) - c_{\mathbf{UV}}(\mathbf{u}_i, \mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i, h)\mathbf{A}^Q_0(\mathbf{v}_i, h)\right|\right\} = o(n^{-\frac{3}{2(P+Q+3)}}(\log n))\quad a.s..$$

Consequently,

$$\begin{aligned} &\sup_{h\in\mathcal{H}_n} |T_2(h) - \tilde{T}_2(h)| \\ =\ &\sup_{h\in\mathcal{H}_n}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I_P}}(\mathbf{u}^\star_j - \mathbf{u}^\star_i)\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}^\star_j - \mathbf{v}^\star_i) - \frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I_P}}(\mathbf{u}_j - \mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}_j - \mathbf{v}_i)}{c_{\mathbf{UV}}(\mathbf{u}_i, \mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i, h)\mathbf{A}^Q_0(\mathbf{v}_i, h)}\right]^2 \\ &+\sup_{h\in\mathcal{H}_n}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I_P}}(\mathbf{u}^\star_j - \mathbf{u}^\star_i)\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}^\star_j - \mathbf{v}^\star_i) - \frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I_P}}(\mathbf{u}_j - \mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}_j - \mathbf{v}_i)}{c_{\mathbf{UV}}(\mathbf{u}_i, \mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i, h)\mathbf{A}^Q_0(\mathbf{v}_i, h)}\right] \\ &\times\left[\frac{\frac{1}{(n-1)}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I_P}}(\mathbf{u}_j - \mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I_Q}}(\mathbf{v}_j - \mathbf{v}_i)}{c_{\mathbf{UV}}(\mathbf{u}_i, \mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i, h)\mathbf{A}^Q_0(\mathbf{v}_i, h)} - 1\right] \end{aligned}$$

$$= o\left(n^{-\frac{P+Q+5}{2(P+Q+3)}}(\log n)\right) \quad a.s.$$
$$= o_P\left(n^{-\frac{3}{P+Q+3}}\right).$$

□

Proof of Lemma 13. Consider (i). For given positive value Q, $\frac{1-h}{h} > Q$ when $h = o(1)$.
On the one hand,

$$\int_0^1 \left|\frac{A_1(u,h)}{A_0(u,h)}\right| du > \int_0^h A_1(u,h)du > h\int_{-1}^{\frac{1-h}{h}} tK(t)dt > W_1 h.$$

On the other hand,

$$\begin{aligned}
&\int_0^1 \left|\frac{A_1(u,h)}{A_0(u,h)}\right| du \\
< \; & 2\int_0^1 |A_1(u,h)|du \\
= \; & 4\int_0^{\frac{1}{2}} A_1(u,h)dx \\
= \; & 4[\int_0^{\frac{1}{2h}}\int_0^{\frac{1}{2}} tK(t)dudt + \int_{\frac{1}{2h}}^{\frac{1}{h}}\int_0^{1-th} tK(t)dudt + \int_{-\frac{1}{2h}}^{0}\int_{-th}^{\frac{1}{2}} tK(t)dudt \\
= \; & 2\int_0^{\frac{1}{2h}} tK(t)dt + \int_{\frac{1}{2h}}^{\frac{1}{h}} 4(1-th)tK(t)dt + \int_{-\frac{1}{2h}}^{0} 4(\frac{1}{2}+th)tK(t)dt \\
= \; & 1 + o(h) + o(h) - 1 + o(h) + 4h\int_{-\infty}^{0} t^2K(t)dt + o(h) \\
< \; & (2W_2+1)h.
\end{aligned}$$

(ii) can be verified using the same argument as that for (i).
(iii) holds since

$$\frac{1}{2} \le |A_0(u,h)| = \left|\int_{-\frac{u}{h}}^{\frac{1-u}{h}} K(t)\right| \le 1,$$

and

$$0 < \int_0^{\infty} t^2K(t) \le |A_2(u,h)| \le 2\int_{-\infty}^{\infty} t^2K(t) < \infty.$$

□

Proof of Lemma 14. Note that $\mathbf{L}^P(\mathbf{u}) - 1$ can be expressed as a polynomial function of $L(u_p,h), p = 1,2,...,P$. We only need to prove

$$\int_0^1 |L^P(u_p) - 1|du_p = O(h), \quad \text{for} \quad p = 1,2,...,P.$$

Now,

$$\begin{aligned}
\int_0^1 |L(u_p) - 1|du_p &= \int_0^1 \left|\int_0^1 \frac{K_h(s-u_p)}{A_0(s,h)}ds - 1\right| du_p \\
&= \int_0^1 \left|\int_0^1 \frac{K_h(s-u_p)}{A_0(s,h)}ds - \int_0^1 \frac{K_h(s-u_p)}{A_0(u_p,h)}ds\right| du_p \\
&= \int_0^1 \left|\int_0^1 K_h(s-u_p)\left[\frac{1}{A_0(s,h)} - \frac{1}{A_0(u_p,h)}\right]ds\right| du_p \\
&\le 4\int_0^1\int_0^1 K_h(s-u_p)\left|A_0(s,h) - A_0(u_p,h)\right| ds du_p
\end{aligned}$$

$$
\begin{aligned}
&= 4\int_0^1 \int_{-\frac{u_p}{h}}^{\frac{1-u_p}{h}} K(t)|A_0(u_p+th,h)-A_0(u_p,h)|dtdu_p \\
&= 4\int_0^1 \int_{-\frac{u_p}{h}}^{\frac{1-u_p}{h}} K(t)\left|tK(-\frac{u_p}{h})-tK(\frac{1-u_p}{h})\right|dtdu_p+O(h) \\
&\le 4\int_0^1 \int_{-\frac{u_p}{h}}^{\frac{1-u_p}{h}} |tK(t)|dt\left|K(-\frac{u_p}{h})-K(\frac{1-u_p}{h})\right|du_p+O(h) \\
&\le 4C\int_0^1 \left|K(-\frac{u_p}{h})-K(\frac{1-u_p}{h})\right|du_p+O(h) \\
&= 8C\int_0^{\frac{1}{2}} \left[K(\frac{u_p}{h})-K(\frac{1-u_p}{h})\right]du_p+O(h) \\
&= 8Ch\int_0^{\frac{1}{2h}} \left[K(t)-K(\frac{1}{h}-t)\right]dt+O(h) \\
&= O(h).
\end{aligned}
$$

□

Proof of Lemma 15. Without loss of generality, we assume $b \ge h$. Note that

$$
\begin{aligned}
\lim_{h\to 0}\frac{E\log A_0(U,h)}{h} &= \lim_{h\to 0}\frac{\int_0^1 \log A_0(u,h)du}{h} \\
&= \lim_{h\to 0}\int_0^1 \frac{-\frac{1-u}{h^2}K(\frac{1-u}{h})-\frac{u}{h^2}K(-\frac{u}{h})}{A_0(u,h)}du \\
&= \lim_{h\to 0} 2\int_0^1 \frac{-\frac{u}{h^2}K(\frac{u}{h})}{A_0(u,h)}du \\
&= -2\int_0^\infty \frac{tK(t)}{\int_{-t}^\infty K(s)ds}dt.
\end{aligned}
$$

Then,

$$
E[\log A_0(U,h)-\log A_0(U,b)] = 2(b-h)\int_0^\infty \frac{tK(t)}{\int_{-t}^\infty K(s)ds}dt.
$$

Similarly, it can be verified that

$$
E[\log A_0(U,b)-\log A_0(U,h)]^2 = O(|b-h|).
$$

Because $1/2 < A_0(u,h) < 2$ for any $0 < u < 1$ and $h > 0$, $|\log A_0(U,b)-\log A_0(U,h) \le 2\log 2$. Then, by Bernstein's inequality (12),

$$
\frac{1}{n}\sum_{i=1}^n [A_0(u_i,h)-A_0(u_i,b)]du = 2(b-h)\int_0^\infty \frac{tK(t)}{\int_{-t}^\infty K(s)ds}dt + o(n^{-\frac{1}{2}}(|b-h|)^{\frac{1}{2}}\log n)\ a.s..
$$

□

Proof of Lemma 16. The proof can be found in 5.3.3 of (14).

□

Proof of Lemma 17. Note that

$$
\begin{aligned}
&\tilde{T}_1(h) \\
= \ & \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j\ne i}\left[\frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)} - \frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)} - \frac{\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{c_{\mathbf{V}}(\mathbf{v}_i)\mathbf{A}^Q_0(\mathbf{v}_i,h)}+1\right] \\
= \ & \frac{2}{n(n-1)}\sum_{i<j}\left\{\left[\frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{2c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)} + \frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{2c_{\mathbf{UV}}(\mathbf{u}_j,\mathbf{v}_j)\mathbf{A}^P_0(\mathbf{u}_j,h)\mathbf{A}^Q_0(\mathbf{v}_j,h)}\right]\right.
\end{aligned}
$$

$$
-\left[\frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)}{2c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)}+\frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)}{2c_{\mathbf{U}}(\mathbf{u}_j)\mathbf{A}^P_0(\mathbf{u}_j,h)}\right]-\left[\frac{\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{2c_{\mathbf{V}}(\mathbf{v}_i)\mathbf{A}^Q_0(\mathbf{v}_i,h)}+\frac{\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{2c_{\mathbf{V}}(\mathbf{v}_j)\mathbf{A}^Q_0(\mathbf{v}_j,h)}\right]+1\Bigg\}
$$

$$
= \frac{2}{n(n-1)}\sum_{i<j} g(\mathbf{u}_i,\mathbf{v}_i,\mathbf{u}_j,\mathbf{v}_j,h).
$$

Let

$$
\tilde{g}_1(\mathbf{u},\mathbf{v}) = E[g(\mathbf{u},\mathbf{v},\mathbf{U},\mathbf{V},h)],
$$

and

$$
\theta(h)=E[g(\mathbf{U}_1,\mathbf{V}_1,\mathbf{U}_2,\mathbf{V}_2,h)],\quad \zeta_1(h)=Var[\tilde{g}_1(\mathbf{U},\mathbf{V},h)],\quad \zeta_2(h)=Var[g(\mathbf{U}_1,\mathbf{V}_1,\mathbf{U}_2,\mathbf{V}_2,h)].
$$

We now analyze the above three terms. It can be easily verified that

$$
\zeta_2(h)=O(\frac{1}{h^{P+Q}}).
$$

Write $g=Z_1-Z_2-Z_3+1$ with

$$
Z_1(\mathbf{u}_i,\mathbf{v}_i,\mathbf{u}_j,\mathbf{v}_j,h) = \frac{1}{2}\left[\frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)}+\frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{c_{\mathbf{UV}}(\mathbf{u}_j,\mathbf{v}_j)\mathbf{A}^P_0(\mathbf{u}_j,h)\mathbf{A}^Q_0(\mathbf{v}_j,h)}\right],
$$

$$
Z_2(\mathbf{u}_i,\mathbf{u}_j,h) = \frac{1}{2}\left[\frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)}{c_{\mathbf{U}}(\mathbf{u}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)}+\frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)}{c_{\mathbf{U}}(\mathbf{u}_j)\mathbf{A}^P_0(\mathbf{u}_j,h)}\right],
$$

and

$$
Z_3(\mathbf{v}_i,\mathbf{v}_j,h) = \frac{1}{2}\left[\frac{\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{c_{\mathbf{V}}(\mathbf{v}_i)\mathbf{A}^Q_0(\mathbf{v}_i,h)}+\frac{\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{c_{\mathbf{V}}(\mathbf{v}_j)\mathbf{A}^Q_0(\mathbf{v}_j,h)}\right].
$$

By Taylor expansion,

$$
\begin{aligned}
& E[Z_1(\mathbf{u},\mathbf{v},\mathbf{U},\mathbf{V},h)] \\
= & \int_{[0,1]^{P+Q}} \frac{1}{2}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{x}-\mathbf{u})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{y}-\mathbf{v})\left[\frac{c_{\mathbf{UV}}(\mathbf{x},\mathbf{y})}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})\mathbf{A}^P_0(\mathbf{u},h)\mathbf{A}^Q_0(\mathbf{v},h)}+\frac{1}{\mathbf{A}^P_0(\mathbf{x},h)\mathbf{A}^Q_0(\mathbf{y},h)}\right]d\mathbf{x}d\mathbf{y} \\
= & \frac{1}{2}\mathbf{L}^P(\mathbf{u})\mathbf{L}^Q(\mathbf{v})+\frac{1}{2}\int_{[0,1]^{P+Q}}\frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{x}-\mathbf{u})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{y}-\mathbf{v})c_{\mathbf{UV}}(\mathbf{x},\mathbf{y})}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})\mathbf{A}^P_0(\mathbf{u},h)\mathbf{A}^Q_0(\mathbf{v},h)}d\mathbf{x}d\mathbf{y} \\
= & \frac{1}{2}\mathbf{L}^P(\mathbf{u})\mathbf{L}^Q(\mathbf{v})+\frac{1}{2}\int_{(\mathbf{u}+h\mathbf{s},\mathbf{v}+h\mathbf{t})\in[0,1]^{P+Q}}\frac{\mathbf{K}^P(\mathbf{s})\mathbf{K}^Q(\mathbf{t})c_{\mathbf{UV}}(\mathbf{u}+h\mathbf{s},\mathbf{v}+h\mathbf{t})}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})\mathbf{A}^P_0(\mathbf{u},h)\mathbf{A}^Q_0(\mathbf{v},h)}d\mathbf{s}d\mathbf{t} \\
= & \frac{1}{2}\mathbf{L}^P(\mathbf{u})\mathbf{L}^Q(\mathbf{v})+\frac{1}{2}+\frac{h}{2}\left[\sum_{p=1}^{P}\frac{\frac{\partial c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})}{\partial u_p}A_1(u_p,h)}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})A_0(u_p,h)}+\sum_{q=1}^{Q}\frac{\frac{\partial c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})}{\partial v_q}A_1(v_q,h)}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})A_0(v_q,h)}\right]+O(h^2).
\end{aligned}
$$

Similarly,

$$
E[Z_2(\mathbf{u},\mathbf{U},h)] = \frac{1}{2}+\frac{1}{2}\mathbf{L}^P(\mathbf{u})+\frac{h}{2}\left[\sum_{p=1}^{P}\frac{\frac{\partial c_{\mathbf{U}}(\mathbf{u})}{\partial u_p}A_1(u_p,h)}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})A_0(u_p,h)}\right]+O(h^2),
$$

$$
E[Z_3(\mathbf{v},\mathbf{V},h)] = \frac{1}{2}+\frac{1}{2}\mathbf{L}^Q(\mathbf{v})+\frac{h}{2}\left[\sum_{q=1}^{Q}\frac{\frac{\partial c_{\mathbf{V}}(\mathbf{v})}{\partial v_q}A_1(v_q,h)}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})A_0(v_q,h)}\right]+O(h^2).
$$

As a result,

$$
\tilde{g}_1(\mathbf{u},\mathbf{v},h) = E[g(\mathbf{u},\mathbf{v},\mathbf{U},\mathbf{V},h)] = \frac{1}{2}[\mathbf{L}^P(\mathbf{u})-1][\mathbf{L}^Q(\mathbf{v})-1]+h\xi(\mathbf{u},\mathbf{v},h)+O(h^2),
$$

with $\xi(\mathbf{u},\mathbf{v},h)=\frac{1}{2}\left[\sum_{p=1}^{P}\frac{\frac{\partial c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})}{\partial u_p}A_1(u_p,h)}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})A_0(u_p,h)}+\sum_{q=1}^{Q}\frac{\frac{\partial c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})}{\partial v_q}A_1(v_q,h)}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})A_0(v_q,h)}-\sum_{p=1}^{P}\frac{\frac{\partial c_{\mathbf{U}}(\mathbf{u})}{\partial u_p}A_1(u_p,h)}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})A_0(u_p,h)}-\sum_{q=1}^{Q}\frac{\frac{\partial c_{\mathbf{V}}(\mathbf{v})}{\partial v_q}A_1(v_q,h)}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})A_0(v_q,h)}\right]$.

Then,

$$
\begin{aligned}
\zeta_1(h) = & E[\tilde{g}^2_1(\mathbf{U},\mathbf{V},h)]-\{E[\tilde{g}^2_1(\mathbf{U},\mathbf{V},h)]\}^2 \\
& +\int_{[0,1]^{P+Q}}\frac{1}{4}[\mathbf{L}^P(\mathbf{u})-1]^2[\mathbf{L}^Q(\mathbf{v})-1]^2c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})d\mathbf{u}d\mathbf{v}
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{[0,1]^{P+Q}} h[\mathbf{L}^P(\mathbf{u}) - 1][\mathbf{L}^Q(\mathbf{v}) - 1]\xi(\mathbf{u},\mathbf{v}) d\mathbf{u} d\mathbf{v} + O(h^2) \\
&\le M_2 \int_0^1 \int_0^1 |\mathbf{L}^P(\mathbf{u}) - 1||\mathbf{L}^Q(\mathbf{v}) - 1| d\mathbf{u} d\mathbf{v} + O(h^2) \\
&= M_2 \int_{[0,1]^{P+Q}} |\mathbf{L}^P(\mathbf{u}) - 1| d\mathbf{u} \int_0^1 |\mathbf{L}^Q(\mathbf{v}) - 1| d\mathbf{v} + O(h^2) \\
&= O(h^2). \qquad [4]
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
& E[Z_1(\mathbf{U}_1, \mathbf{V}_1, \mathbf{U}_2, \mathbf{V}_2, h)] \\
= & \int_{[0,1]^{2(P+Q)}} \frac{1}{2} \mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_2 - \mathbf{u}_1) \mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_2 - \mathbf{v}_1) \left[\frac{c_{\mathbf{UV}}(\mathbf{u_2}, \mathbf{v_2})}{\mathbf{A}_0^P(\mathbf{u}_1, h)\mathbf{A}_0^Q(\mathbf{v}_1, h)} + \frac{c_{\mathbf{UV}}(\mathbf{u_1}, \mathbf{v_1})}{\mathbf{A}_0^P(\mathbf{u}_2, h)\mathbf{A}_0^Q(\mathbf{v}_2, h)} \right] d\mathbf{u}_1 d\mathbf{v}_1 d\mathbf{u}_2 d\mathbf{v}_2 \\
= & \int_{[0,1]^{2(P+Q)}} \frac{1}{2} \mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_2 - \mathbf{u}_1) \mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_2 - \mathbf{v}_1) \frac{c_{\mathbf{UV}}(\mathbf{u_2}, \mathbf{v_2})}{\mathbf{A}_0^P(\mathbf{u}_1, h)\mathbf{A}_0^Q(\mathbf{v}_1, h)} d\mathbf{u}_1 d\mathbf{v}_1 d\mathbf{u}_2 d\mathbf{v}_2 \\
= & \int_{[0,1]^{P+Q}} \frac{1}{\mathbf{A}_0^P(\mathbf{u}_1, h)\mathbf{A}_0^Q(\mathbf{v}_1, h)} \left[\int_{[0,1]^{P+Q}} \mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_2 - \mathbf{u}_1) \mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_2 - \mathbf{v}_1) c_{\mathbf{UV}}(\mathbf{u}_2, \mathbf{v}_2) d\mathbf{u}_2 d\mathbf{v}_2 \right] d\mathbf{u}_1 d\mathbf{v}_1 \\
= & \int_{[0,1]^{P+Q}} \frac{1}{\mathbf{A}_0^P(\mathbf{u}_1, h)\mathbf{A}_0^Q(\mathbf{v}_1, h)} \left[\int_{(\mathbf{u}_1 + h\mathbf{s}, \mathbf{v}_1 + h\mathbf{t}) \in [0,1]^{P+Q}} \mathbf{K}^P(\mathbf{s}) \mathbf{K}^Q(\mathbf{t}) c_{\mathbf{UV}}(\mathbf{u}_1 + h\mathbf{s}, \mathbf{v}_1 + h\mathbf{t}) d\mathbf{s} d\mathbf{t} \right] d\mathbf{u}_1 d\mathbf{v}_1 \\
= & 1 + h \sum_{p=1}^P \int_{[0,1]^{P+Q}} \frac{A_1(u_{1p}, h)}{A_0(u_{1p}, h)} \frac{\partial c_{\mathbf{UV}}}{\partial u_p}(\mathbf{u}_1, \mathbf{v}_1) d\mathbf{u}_1 d\mathbf{v}_1 + h \sum_{q=1}^Q \int_{[0,1]^{P+Q}} \frac{A_1(v_{1q}, h)}{A_0(v_{1q}, h)} \frac{\partial c_{\mathbf{UV}}}{\partial v_q}(\mathbf{u}_1, \mathbf{v}_1) d\mathbf{u}_1 d\mathbf{v}_1 \\
& + \frac{h^2}{2} \sum_{p=1}^P \int_{[0,1]^{P+Q}} \frac{A_2(u_{1p}, h)}{A_0(u_{1p}, h)} \frac{\partial^2 c_{\mathbf{UV}}}{\partial u_p^2}(\mathbf{u}_1, \mathbf{v}_1) d\mathbf{u}_1 d\mathbf{v}_1 + \frac{h^2}{2} \sum_{q=1}^Q \int_{[0,1]^{P+Q}} \frac{A_2(v_{1q}, h)}{A_0(v_{1q}, h)} \frac{\partial^2 c_{\mathbf{UV}}}{\partial v_q^2}(\mathbf{u}_1, \mathbf{v}_1) d\mathbf{u}_1 d\mathbf{v}_1 \\
& + \frac{h^2}{2} \sum_{p_1=1}^P \sum_{p_2 \ne p_1} \int_{[0,1]^{P+Q}} \frac{A_1(u_{1p_1}, h) A_1(u_{1p_2}, h)}{A_0(u_{1p_1}, h) A_0(u_{1p_2}, h)} \frac{\partial^2 c_{\mathbf{UV}}}{\partial u_{p_1} \partial u_{p_2}}(\mathbf{u}_1, \mathbf{v}_1) d\mathbf{u}_1 d\mathbf{v}_1 \\
& + \frac{h^2}{2} \sum_{q_1=1}^Q \sum_{q_2 \ne q_1} \int_{[0,1]^{P+Q}} \frac{A_1(v_{1q_1}, h) A_1(v_{1q_2}, h)}{A_0(v_{1q_1}, h) A_0(v_{1q_2}, h)} \frac{\partial^2 c_{\mathbf{UV}}}{\partial v_{q_1} \partial v_{q_2}}(\mathbf{u}_1, \mathbf{v}_1) d\mathbf{u}_1 d\mathbf{v}_1 \\
& + h^2 \sum_{p=1}^P \sum_{q=1}^Q \int_{[0,1]^{P+Q}} \frac{A_1(u_{1p}, h) A_1(v_{1q_2}, h)}{A_0(u_{1p}, h) A_0(v_{1q}, h)} \frac{\partial^2 c_{\mathbf{UV}}}{\partial u_p \partial v_q}(\mathbf{u}_1, \mathbf{v}_1) d\mathbf{u}_1 d\mathbf{v}_1 \\
& + \frac{h^3}{6} \sum_{p=1}^P \int_{[0,1]^{P+Q}} \frac{A_3(u_{1p}, h)}{A_0(u_{1p}, h)} \frac{\partial^3 c_{\mathbf{UV}}}{\partial u_p^3}(\mathbf{u}_1, \mathbf{v}_1) d\mathbf{u}_1 d\mathbf{v}_1 + \frac{h^3}{6} \sum_{q=1}^Q \int_{[0,1]^{P+Q}} \frac{A_3(v_{1q}, h)}{A_0(v_{1q}, h)} \frac{\partial^3 c_{\mathbf{UV}}}{\partial v_q^3}(\mathbf{u}_1, \mathbf{v}_1) d\mathbf{u}_1 d\mathbf{v}_1 \\
& + \frac{h^2}{2} \sum_{p_1=1}^P \sum_{p_2 \ne p_1} \int_{[0,1]^{P+Q}} \frac{A_2(u_{1p_1}, h) A_1(u_{1p_2}, h)}{A_0(u_{1p_1}, h) A_0(u_{1p_2}, h)} \frac{\partial^3 c_{\mathbf{UV}}}{\partial u_{p_1}^2 \partial u_{p_2}}(\mathbf{u}_1, \mathbf{v}_1) d\mathbf{u}_1 d\mathbf{v}_1 \\
& + \frac{h^2}{2} \sum_{q_1=1}^Q \sum_{q_2 \ne q_1} \int_{[0,1]^{P+Q}} \frac{A_2(v_{1q_1}, h) A_1(v_{1q_2}, h)}{A_0(v_{1q_1}, h) A_0(v_{1q_2}, h)} \frac{\partial^3 c_{\mathbf{UV}}}{\partial v_{q_1}^2 \partial v_{q_2}}(\mathbf{u}_1, \mathbf{v}_1) d\mathbf{u}_1 d\mathbf{v}_1 \\
& + \frac{h^2}{6} \sum_{\substack{1 \le p_1, p_2, p_3 \le P \\ p_1 \ne p_2 \ne p_3}} \int_{[0,1]^{P+Q}} \frac{A_1(u_{1p_1}, h) A_1(u_{1p_2}, h) A_1(u_{1p_3}, h)}{A_0(u_{1p_1}, h) A_0(u_{1p_2}, h) A_0(u_{1p_3}, h)} \frac{\partial^3 c_{\mathbf{UV}}}{\partial u_{p_1} \partial u_{p_2} \partial u_{p_3}}(\mathbf{u}_1, \mathbf{v}_1) d\mathbf{u}_1 d\mathbf{v}_1 \\
& + \frac{h^2}{6} \sum_{\substack{1 \le q_1, q_2, q_3 \le Q \\ q_1 \ne q_2 \ne q_3}} \int_{[0,1]^{P+Q}} \frac{A_1(v_{1q_1}, h) A_1(v_{1q_2}, h) A_1(v_{1q_3}, h)}{A_0(v_{1q_1}, h) A_0(v_{1q_2}, h) A_0(v_{1q_3}, h)} \frac{\partial^3 c_{\mathbf{UV}}}{\partial v_{q_1} \partial v_{q_2} \partial v_{q_3}}(\mathbf{u}_1, \mathbf{v}_1) d\mathbf{u}_1 d\mathbf{v}_1 \\
& + \frac{h^2}{2} \sum_{p=1}^P \sum_{q=1}^Q \int_{[0,1]^{P+Q}} \frac{A_2(u_{1p}, h) A_1(v_{1q}, h)}{A_0(u_{1p}, h) A_0(v_{1q}, h)} \frac{\partial^3 c_{\mathbf{UV}}}{\partial u_p^2 \partial v_q}(\mathbf{u}_1, \mathbf{v}_1) d\mathbf{u}_1 d\mathbf{v}_1 \\
& + \frac{h^2}{2} \sum_{p=1}^P \sum_{q=1}^Q \int_{[0,1]^{P+Q}} \frac{A_1(u_{1p}, h) A_2(v_{1q}, h)}{A_0(u_{1p}, h) A_0(v_{1q}, h)} \frac{\partial^3 c_{\mathbf{UV}}}{\partial u_p \partial v_q^2}(\mathbf{u}_1, \mathbf{v}_1) d\mathbf{u}_1 d\mathbf{v}_1
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{h^2}{2}\sum_{p_1=1}^{P}\sum_{p_2\neq p_1}\sum_{q=1}^{Q}\int_{[0,1]^{P+Q}}\frac{A_1(u_{1p_1},h)A_1(u_{1p_2},h)A_1(v_{1q},h)}{A_0(u_{1p_1},h)A_0(u_{1p_2},h)A_0(v_{1q},h)}\frac{\partial^3 c_{\mathbf{UV}}}{\partial u_{p_1}\partial u_{p_2}\partial v_q}(\mathbf{u}_1,\mathbf{v}_1)d\mathbf{u}_1 d\mathbf{v}_1\\
&+\frac{h^2}{2}\sum_{p=1}^{P}\sum_{q_1=1}^{Q}\sum_{q_2\neq q_1}\int_{[0,1]^{P+Q}}\frac{A_1(u_{1p},h)A_1(v_{1q_1},h)A_1(v_{1q_2},h)}{A_0(u_{1p},h)A_0(v_{1q_1},h)A_0(v_{1q_2},h)}\frac{\partial^3 c_{\mathbf{UV}}}{\partial u_p\partial v_{q_1}\partial v_{q_2}}(\mathbf{u}_1,\mathbf{v}_1)d\mathbf{u}_1 d\mathbf{v}_1+O(h^4)\\
=&\quad 1+O(h^4).
\end{aligned}
$$

Let x, y and z denote any three different entries of $\mathbf{u}$ or $\mathbf{v}$. Let X, Y and Z be the corresponding entries in $\mathbf{U}$ or $\mathbf{V}$. Let $f_{\mathbf{X}}$ be the density function of $\mathbf{X}$ with $\mathbf{X}$ being either a random vector or a random variable. Then $f_X = f_Y = f_Z = 1$ as $U_p, p = 1, ..., P$ and $V_q, q = 1, ..., Q$ are uniformly distributed. The last equality above follows from the following facts.

$$
\begin{aligned}
(I)\quad & \int_{[0,1]^{P+Q}}\frac{A_1(x,h)}{A_0(x,h)}\frac{\partial c_{\mathbf{UV}}}{\partial x}(\mathbf{u},\mathbf{v})d\mathbf{u}d\mathbf{v}=\int_0^1\frac{A_1(x,h)}{A_0(x,h)}\frac{df_X}{dx}(x)dx=0;\\
(II)\quad & \int_{[0,1]^{P+Q}}\frac{A_2(x,h)}{A_0(x,h)}\frac{\partial c_{\mathbf{UV}}}{\partial x}(\mathbf{u},\mathbf{v})d\mathbf{u}d\mathbf{v}=\int_0^1\frac{A_2(x,h)}{A_0(x,h)}\frac{d^2f_X}{dx^2}(x)dx=0;\\
(III)\quad & \int_{[0,1]^{P+Q}}\frac{A_3(x,h)}{A_0(x,h)}\frac{\partial c_{\mathbf{UV}}}{\partial x}(\mathbf{u},\mathbf{v})d\mathbf{u}d\mathbf{v}=\int_0^1\frac{A_3(x,h)}{A_0(x,h)}\frac{d^3f_X}{dx^3}(x)dx=0;\\
(IV)\quad & \int_{[0,1]^{P+Q}}\frac{A_1(x,h)A_1(y,h)}{A_0(x,h)A_0(y,h)}\frac{\partial^2 c_{\mathbf{UV}}}{\partial x\partial y}(\mathbf{u}_1,\mathbf{v}_1)d\mathbf{u}_1 d\mathbf{v}_1=\int_0^1\int_0^1\frac{A_1(x,h)A_1(y,h)}{A_0(x,h)A_0(y,h)}\frac{\partial f_{XY}}{\partial x\partial y}(x,y)dxdy=O(h^2);\\
(V)\quad & \int_{[0,1]^{P+Q}}\frac{A_2(x,h)A_1(y,h)}{A_0(x,h)A_0(y,h)}\frac{\partial^3 c_{\mathbf{UV}}}{\partial x^2\partial y}(\mathbf{u}_1,\mathbf{v}_1)d\mathbf{u}_1 d\mathbf{v}_1=\int_0^1\int_0^1\frac{A_2(x,h)A_1(y,h)}{A_0(x,h)A_0(y,h)}\frac{\partial f^3_{XY}}{\partial x^2\partial y}(x,y)dxdy=O(h);\\
(VI)\quad & \int_{[0,1]^{P+Q}}\frac{A_1(x,h)A_1(y,h)A_1(z,h)}{A_0(x,h)A_0(y,h)A_0(z,h)}\frac{\partial^3 c_{\mathbf{UV}}}{\partial x\partial y\partial z}(\mathbf{u}_1,\mathbf{v}_1)d\mathbf{u}_1 d\mathbf{v}_1\\
=\;& \int_0^1\int_0^1\int_0^1\frac{A_1(x,h)A_1(y,h)A_1(z,h)}{A_0(x,h)A_0(y,h)A_0(z,h)}\frac{\partial f^3_{XY}}{\partial x\partial y\partial z}(x,y,z)dxdydz=O(h^3).
\end{aligned}
$$

By similar arguments,

$$
E[Z_2(\mathbf{U}_1,\mathbf{U}_2,h)]=1+O(h^4);
$$
$$
E[Z_3(\mathbf{V}_1,\mathbf{V}_2,h)]=1+O(h^4).
$$

Then,

$$
\theta(h)=O(h^4).
$$

Let $U_n=\frac{2}{n(n-1)}\sum_{i<j}g(\mathbf{u}_i,\mathbf{v}_i,\mathbf{u}_j,\mathbf{v}_j,h)$, $\hat{U}_n=\frac{2}{n}\sum_{i=1}^n\tilde{g}_1(\mathbf{u_i},\mathbf{v_i},h)-(n-1)E(U_n)$ be the projection of $E(U_n)$ and $U_n=\hat{U}_n+R_n$. As $\tilde{g}_1(\mathbf{U}_i,\mathbf{V}_i,h)<M$ and $Var(\tilde{g}_1(\mathbf{U},\mathbf{V},h))=\zeta_1(h)=O(h^2)$, by Bernstein's inequality (12),

$$
P[|\hat{U}_n-\theta(h))|>n^{-\frac{3}{P+Q+3}}t]\le 2\exp\left\{-\frac{t^2n^{\frac{P+Q-3}{P+Q+3}}}{Ch^2+C'n^{-\frac{3}{P+Q+3}}t}\right\}.
$$

and

$$
\hat{U}_n-\theta(h)=o(n^{-\frac{1}{2}}h\log n)\ a.s..
$$

Note that

$$
\begin{aligned}
E[U_n-\hat{U}_n]^2 &= \frac{4(n-2)}{n(n-1)}\zeta_1+\frac{2}{n(n-1)}\zeta_2-\frac{4}{n}\zeta_1\\
&= O(\frac{1}{n^2h^{P+Q}}).
\end{aligned}
$$

Then, by Chebyshev's inequality (11),

$$
P[|R_n|>n^{-3/(P+Q+3)}t]\le t^{-2}ER_n^2n^{6/(P+Q+3)}=C_3t^{-2}n^{-\frac{2P+2Q}{P+Q+3}}h^{-(P+Q)}.
$$

Clearly, $h^{(P+Q)/2}U_n$ is also a U-statistics and $h^{(P+Q)/2}\hat{U}_n$ is its projection. Then,

$$
E[h^{\frac{P+Q}{2}}U_n-h^{\frac{P+Q}{2}}\hat{U}_n]^2 \quad=\quad O(n^{-2}).
$$

According to Lemma 16,

$$
R_n=o(n^{-1}h^{-\frac{P+Q}{2}}\log n)\ a.s..
$$

As a result,

$$\begin{aligned}T_1(h)-\theta(h)=\hat{U}_n-\theta(h)+R_n &= o(n^{-\frac{1}{2}}h\log n+n^{-1}h^{-\frac{P+Q}{2}}\log n)\ a.s.\\ &= o(h^3+\frac{1}{nh^{P+Q}})\ a.s..\end{aligned}$$

Furthermore,

$$\begin{aligned}&P[\sup_{h\in\mathcal{H}_n}|T_1(h)-\theta(h)|>n^{-\frac{3}{P+Q+3}}t]\\ \leq\ &\sum_{i=1}^m P[|T_1(h_i)-\theta(h_i)|>n^{-\frac{3}{P+Q+3}}t]\\ \leq\ &\sum_{i=1}^m P[|\hat{U}_n(h_i)-\theta(h_i)|>\frac{1}{2}n^{-\frac{3}{P+Q+3}}t]+\sum_{i=1}^m P[|R_n(h_i)|>\frac{1}{2}n^{-\frac{3}{P+Q+3}}t]\\ \leq\ &\sum_{i=1}^m\Big[2\exp\Big\{-\frac{t^2n^{\frac{P+Q-3}{P+Q+3}}}{Ch^2+C'n^{-\frac{3}{P+Q+3}}t}\Big\}+C''t^{-2}n^{-\frac{2P+2Q}{P+Q+3}}h^{-(P+Q)}\Big]\\ =\ &o(1).\end{aligned}$$

Combining this with the result that $\sup_{h\in\mathcal{H}_n}|\theta(h)|=O(\sup_i h_i^4)=O(n^{-\frac{4}{P+Q+3}})$, we have,

$$\sup_{h\in\mathcal{H}_n}|T_1(h)|=o_P(n^{-\frac{3}{P+Q+3}}).$$

□

Proof of Lemma 18. We begin with analyzing the term $\frac{1}{n}\sum_{i=1}^n\Big[\frac{\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)}-1\Big]^2$, which can be expressed in terms of two U-statistics, as we shall show.

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n\Big[\frac{\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)}-1\Big]^2\\
=\ &\frac{1}{n(n-1)^2}\sum_{i=1}^n\sum_{j\neq i}\sum_{l\neq i}\Big[\frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)}-1\Big]\Big[\frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_l-\mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_l-\mathbf{v}_i)}{c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)}-1\Big]\\
=\ &\frac{1}{n(n-1)^2}\sum_{i<j}\Big\{\Big[\frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)}-1\Big]^2+\Big[\frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_i-\mathbf{u}_j)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_i-\mathbf{v}_j)}{c_{\mathbf{UV}}(\mathbf{u}_j,\mathbf{v}_j)\mathbf{A}^P_0(\mathbf{u}_j,h)\mathbf{A}^Q_0(\mathbf{v}_j,h)}-1\Big]^2\Big\}\\
&+\frac{2}{n(n-1)^2}\sum_{i<j<l}\Big\{\Big[\frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)}-1\Big]\Big[\frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_l-\mathbf{u}_i)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_l-\mathbf{v}_i)}{c_{\mathbf{UV}}(\mathbf{u}_i,\mathbf{v}_i)\mathbf{A}^P_0(\mathbf{u}_i,h)\mathbf{A}^Q_0(\mathbf{v}_i,h)}-1\Big]\\
&\quad+\Big[\frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_i-\mathbf{u}_j)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_i-\mathbf{v}_j)}{c_{\mathbf{UV}}(\mathbf{u}_j,\mathbf{v}_j)\mathbf{A}^P_0(\mathbf{u}_j,h)\mathbf{A}^Q_0(\mathbf{v}_j,h)}-1\Big]\Big[\frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_l-\mathbf{u}_j)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_l-\mathbf{v}_j)}{c_{\mathbf{UV}}(\mathbf{u}_j,\mathbf{v}_j)\mathbf{A}^P_0(\mathbf{u}_j,h)\mathbf{A}^Q_0(\mathbf{v}_j,h)}-1\Big]\\
&\quad+\Big[\frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_i-\mathbf{u}_l)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_i-\mathbf{v}_l)}{c_{\mathbf{UV}}(\mathbf{u}_l,\mathbf{v}_l)\mathbf{A}^P_0(\mathbf{u}_l,h)\mathbf{A}^Q_0(\mathbf{v}_l,h)}-1\Big]\Big[\frac{\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u}_l)\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_l)}{c_{\mathbf{UV}}(\mathbf{u}_l,\mathbf{v}_l)\mathbf{A}^P_0(\mathbf{u}_l,h)A^Q_0(\mathbf{v}_l,h)}-1\Big]\Big\}\\
=\ &\frac{1}{n(n-1)^2}\sum_{i<j}g_1(\mathbf{u}_i,\mathbf{v}_i,\mathbf{u}_j,\mathbf{v}_j,h)+\frac{1}{n(n-1)^2}\sum_{i<j<l}g_2(\mathbf{u}_i,\mathbf{v}_i,\mathbf{u}_j,\mathbf{v}_j,\mathbf{u}_l,\mathbf{v}_l,h)\\
=\ &\frac{1}{2(n-1)}U_{n,1}+\frac{n-2}{3(n-1)}U_{n,2}.\end{aligned}$$

Now, we use the same technique as in Lemma 17 to analyse these two U-statistics. Let

$$\begin{aligned}\tilde{g}_{1,1}(\mathbf{u},\mathbf{v},h) &= E[g_1(\mathbf{u},\mathbf{v},\mathbf{U},\mathbf{V},h)];\quad \theta_1(h)=E[g_1(\mathbf{U}_1,\mathbf{V}_1,\mathbf{U}_2,\mathbf{V}_2,h)];\\ \zeta_{1,1}(h) &= Var[\tilde{g}_{1,1}(\mathbf{U},\mathbf{V},h)];\quad \zeta_{1,2}(h)=Var[g_1(\mathbf{U}_1,\mathbf{V}_1,\mathbf{U}_2,\mathbf{V}_2,h)];\\ \tilde{g}_{2,1}(\mathbf{u},\mathbf{v},h) &= E[g_2(\mathbf{u},\mathbf{v},\mathbf{U}_1,\mathbf{V}_1,\mathbf{U}_2,\mathbf{V}_2,h)];\quad \tilde{g}_{2,2}(\mathbf{u}_1,\mathbf{v}_1,\mathbf{u}_2,\mathbf{v}_2,h)=E[g_2(\mathbf{u}_1,\mathbf{v}_1,\mathbf{u}_2,\mathbf{v}_2,\mathbf{U},\mathbf{V}),h];\\ \theta_2(h) &= E[g_2(\mathbf{U}_1,\mathbf{V}_1,\mathbf{U}_2,\mathbf{V}_2,\mathbf{U}_3,\mathbf{V}_3,h)];\quad \zeta_{2,1}(h)=Var[\tilde{g}_{2,1}(\mathbf{U},\mathbf{V},h)];\end{aligned}$$

$$\zeta_{2,2}(h) = Var[\tilde{g}_{2,2}(\mathbf{U}_1,\mathbf{V}_1,\mathbf{U}_2,\mathbf{V}_2,h)]; \quad \zeta_{2,3}(h) = Var[g_2(\mathbf{U}_1,\mathbf{V}_1,\mathbf{U}_2,\mathbf{V}_2,\mathbf{U}_3,\mathbf{V}_3,h)].$$

After some tedious calculations, we have

$$\theta_1(h) = \frac{2}{h^{P+Q}}\left[\int_{-\infty}^{\infty} K^2(t)dt\right]^{P+Q} + O(\frac{1}{h^{P+Q-1}});$$
$$\theta_2(h) = 3h^2\int_{[0,1]^{P+Q}}\left[\sum_{p=1}^{P}\frac{A_1^2(u_p,h)(\frac{\partial c_{\mathbf{UV}}}{\partial u_p}(\mathbf{u},\mathbf{v}))^2}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})A_0^2(\mathbf{u}_p,h)} + \sum_{q=1}^{Q}\frac{A_1^2(v_q,h)(\frac{\partial c_{\mathbf{UV}}}{\partial v_q}(\mathbf{u},\mathbf{v}))^2}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})A_0^2(v_q,h)}\right]d\mathbf{u}d\mathbf{v} + O(h^4).$$

Moreover,

$$\zeta_{1,1}(h) = O(\frac{1}{h^{2(P+Q)}});$$
$$\zeta_{1,2}(h) = O(\frac{1}{h^{3(P+Q)}});$$
$$\zeta_{2,1}(h) = O(h^2);$$
$$\zeta_{2,2}(h) = O(\frac{1}{h^{P+Q}});$$
$$\zeta_{2,3}(h) = O(\frac{1}{h^{2(P+Q)}}).$$

Write $U_{n,1} = \hat{U}_{n,1} + R_{n,1}$ and $U_{n,2} = \hat{U}_{n,2} + R_{n,2}$ with $\hat{U}_{n,1} = \frac{2}{n}\sum\tilde{g}_{1,1}(\mathbf{u}_i,\mathbf{v}_i,h)$ and $\hat{U}_{n,2} = \frac{3}{n}\sum\tilde{g}_{2,1}(\mathbf{u}_i,\mathbf{v}_i,h)$. By Bernstein's inequality (12) and Lemma 16,

$$\hat{U}_{n,1} = \frac{2}{h^{P+Q}}\left[\int_{-\infty}^{\infty} K^2(t)dt\right]^{P+Q} + O(\frac{1}{h^{(P+Q-1)}}) + o(n^{-\frac{1}{2}}h^{-(P+Q)}\log n)\ a.s.;$$
$$R_{n,1} = o(n^{-1}h^{-\frac{3(P+Q)}{2}}\log n)\ a.s.;$$
$$\hat{U}_{n,2} = 3C_1h^3 + O(h^4) + o(n^{-\frac{1}{2}}h\log n)\ a.s.;$$
$$R_{n,2} = o(n^{-1}h^{-\frac{P+Q}{2}}\log n)\ a.s..$$

Consequently,

$$\frac{1}{2(n-1)}U_{n,1} + \frac{n-2}{3(n-1)}U_{n,2} = C_1h^3 + \frac{1}{nh^{P+Q}}\left[\int_{-\infty}^{\infty} K^2(t)dt\right]^{P+Q} + o(h^3 + \frac{1}{nh^{P+Q}})\ a.s..$$

When $h \propto n^{-1/(P+Q+3)}$, by Bernstein's inequality (12) for U-statistics ((15)),

$$P\Big(\frac{1}{n}|U_{n,1} - \theta_1(h)| > n^{-3/(P+Q+3)}t\Big) \le 4\exp\left\{-\frac{t^2 n^{\frac{3(P+Q+1)}{P+Q+3}}}{Ch^{-4} + C'n^{\frac{P+Q}{P+Q+3}}h^{-4}}\right\} = o(n^{-\frac{P+Q}{P+Q+3}}),$$

since $\zeta_{1,1} = O(\frac{1}{h^4})$ and $||U_{n,1}||_\infty \le C/h^4$.

Moreover, when $h \propto n^{-1/(P+Q+3)}$, by Bernstein's inequality (12) and Chebyshev's inequality (11),

$$P(|U_{n,2} - \theta_2(h)| > n^{-3/(P+Q+3)}t) \le 2\exp\left\{-\frac{t^2 n^{\frac{P+Q-3}{P+Q+3}}}{C_1h^2 + C_2n^{-\frac{3}{P+Q+3}}t}\right\} + C_3t^{-2}n^{-\frac{2(P+Q)}{P+Q+3}}h^{-(P+Q)} = O(n^{-\frac{P+Q}{P+Q+3}}).$$

Repeating the above procedure we have

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\sum_{j\neq i}\mathbf{K}^P_{h\mathbf{I}_P}(\mathbf{u}_j - \mathbf{u}_i)}{\mathbf{A}_0^P(\mathbf{u}_i,h)} - 1\right]^2 = o(h^3 + \frac{1}{nh^{P+Q}})\ a.s.;$$
$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\sum_{j\neq i}\mathbf{K}^Q_{h\mathbf{I}_Q}(\mathbf{v}_j - \mathbf{v}_i)}{\mathbf{A}_0^Q(\mathbf{v}_i,h)} - 1\right]^2 = o(h^3 + \frac{1}{nh^{P+Q}})\ a.s..$$

When $h \propto n^{-1/(P+Q+3)}$,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\sum_{j\neq i}\mathbf{K}^P_{h\mathbf{I}_P}(\mathbf{u}_j - \mathbf{u}_i)}{\mathbf{A}_0^P(\mathbf{u}_i,h)} - 1\right]^2\right| > n^{-\frac{3}{P+Q+3}}t\right) = O(n^{-\frac{P+Q}{P+Q+3}}),$$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\sum_{j\neq i}\mathbf{K}^{Q}_{h\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}_i)}{\mathbf{A}^{Q}_0(\mathbf{v}_i,h)}-1\right]^2\right|>n^{-\frac{3}{P+Q+3}}t\right)=O(n^{-\frac{P+Q}{P+Q+3}}),$$

which implies that

$$\begin{aligned}\tilde{\theta}(h) &= \frac{1}{2(n-1)}\theta_1(h)+\frac{n-2}{3(n-1)}\theta_2(h)+o(h^3+\frac{1}{nh^{P+Q}})\\ &= C_2h^3+\frac{1}{nh^{(P+Q)}}\left[\int_{-\infty}^{\infty}K^2(s)ds\right]^{(P+Q)}+o(h^3+\frac{1}{nh^{P+Q}});\end{aligned}$$

$$T_2(h)=\tilde{\theta}(h)+o(h^3+\frac{1}{nh^{P+Q}})\ a.s.,$$

and

$$\sup_{h\in\mathcal{H}_n}|T_2(h)-\tilde{\theta}(h)|=o_P(n^{-\frac{3}{P+Q+3}}).$$

□

Proof of Lemma 19. The result follows directly from Lemma 13. Without loss of generality, we assume

$$m_1=\int_{[0,1]^{(P+Q-1)}}\left[\frac{\partial c_{\mathbf{UV}}}{\partial u_p}(\mathbf{u}_{p,0},\mathbf{v})\right]^2 d\mathbf{u}_{-p}d\mathbf{v}.$$

Then,

$$\begin{aligned}C_1 &> \frac{1}{h}\left[\int_{[0,1]^{P+Q}}\left[\frac{A_1^2(u_p,h)(\frac{\partial c_{\mathbf{UV}}}{\partial u_p}(\mathbf{u},\mathbf{v}))^2}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})A_0^2(u_p,h)}\right]d\mathbf{u}d\mathbf{v}\right]\\ &> \frac{1}{h}\left[\int_{[0,1]^{P+Q-1}}\int_0^h\frac{A_1^2(u_p,h)(\frac{\partial c_{\mathbf{UV}}}{\partial u_p}(\mathbf{u},\mathbf{v}))^2}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})A_0^2(u_p,h)}du_pd\mathbf{u}_{-p}d\mathbf{v}\right]\\ &= \frac{1}{h}\left[\int_{[0,1]^{P+Q-1}}\int_0^h\frac{A_1^2(u_p,h)(\frac{\partial c_{\mathbf{UV}}}{\partial u_p}(\mathbf{u}_{p,0},\mathbf{v})+o(1))^2}{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})A_0^2(u_p,h)}du_pd\mathbf{u}_{-p}d\mathbf{v}\right]\\ &= \frac{(m_1+o(1))}{hM_2}\left[\int_0^h\frac{A_1^2(u_p,h)}{A_0^2(u_p,h)}du_p\right]\\ &> \frac{m_1W_1^2}{M_2}-\epsilon.\end{aligned}$$

Additionally,

$$\begin{aligned}C_1 &< \frac{2m_2^2}{hM_1}\left[\int_0^1\frac{A_1^2(u_p,h)}{A_0^2(u_p,h)}du_p\right]\\ &< \frac{2m_2^2(4W_2W_3+1)}{M_1}.\end{aligned}$$

□

Proof of Lemma 20. Let $A_M=\{(\mathbf{u},\mathbf{v})|c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})>3M\}$ with $M>2$. Since $c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})$ is absolutely continuous on A_M^c, by same argument for proving Lemma 8,

$$\sup_{(\mathbf{u},\mathbf{v})\in A_M^c}|\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v})-c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})\mathbf{A}^P_0(\mathbf{u},h)\mathbf{A}^Q_0(\mathbf{v},h)|\to 0.\ \ a.s.$$

To prove the result on A_M, we just need to prove that

$$\lim_{n\to\infty}\inf_{(\mathbf{u},\mathbf{v})\in A_M}\frac{\frac{1}{n-1}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U}-\mathbf{u})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{V}-\mathbf{v})}{\mathbf{A}^P_0(\mathbf{u},h)\mathbf{A}^Q_0(\mathbf{v},h)}>M.\ \ a.s.$$

Write $c^M_{\mathbf{UV}}(\mathbf{u},\mathbf{v})=\min[c_{\mathbf{UV}}(\mathbf{u},\mathbf{v}),M]$. By the absolute continuity of $c^{3M}_{\mathbf{UV}}$, for $\forall(\mathbf{u}_1,\mathbf{v}_1)\in A_M^c$ and $\forall\epsilon>0$, there exist a $\delta>0$ such that

$$\sup_{d((\mathbf{u},\mathbf{v}),(\mathbf{u}_1,\mathbf{v}_1))<\delta}|c^{3M}_{\mathbf{UV}}(\mathbf{u},\mathbf{v})-c^{3M}_{\mathbf{UV}}(\mathbf{u}_1,\mathbf{v}_1)|<\epsilon$$

with d the Euclidean distance. This implies that

$$\inf_{d((\mathbf{u},\mathbf{v}),A_M)<\delta)} c^{3M}_{\mathbf{UV}}(\mathbf{u},\mathbf{v}) > \frac{11M}{4},$$

where $d((\mathbf{u},\mathbf{v}),A_M) = \inf_{(\mathbf{u}_1,\mathbf{v}_1)\in A_M} d((\mathbf{u},\mathbf{v}),(\mathbf{u}_1,\mathbf{v}_1))$. On noting that $K(t)$ is integrable on $\mathcal{R}$, for $\forall\epsilon>0$, there exists a positive number N_0 such that

$$\int_{-N_0}^{N_0} K(t)dt > 1-\epsilon.$$

Choose N_0 such that

$$\int_{\max(-N_0,\frac{-u}{h})}^{\min(N_0,\frac{1-u}{h})} K(t)dt > (\frac{3}{4})^{\frac{1}{P+Q}} \int_{\frac{-u}{h}}^{\frac{1-u}{h}} K(t)dt.$$

Then, for ϵ small enough,

$$\begin{aligned}
&\inf_{(\mathbf{u},\mathbf{v})\in A_M} \frac{E\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U}-\mathbf{u})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{V}-\mathbf{v})}{\mathbf{A}^P_0(\mathbf{u},h)\mathbf{A}^Q_0(\mathbf{v},h)} \\
= &\inf_{(\mathbf{u},\mathbf{v})\in A_M} \int_{(\mathbf{u}+h\mathbf{s},\mathbf{v}+h\mathbf{t})\in[0,1]^{P+Q}} \frac{c_{\mathbf{UV}}(\mathbf{u}+h\mathbf{s},\mathbf{v}+h\mathbf{t})}{\mathbf{A}^P_0(\mathbf{u},h)\mathbf{A}^Q_0(\mathbf{v},h)}\mathbf{K}^P(\mathbf{s})\mathbf{K}^Q(\mathbf{t})d\mathbf{s}d\mathbf{t} \\
\geq &\inf_{(\mathbf{u},\mathbf{v})\in A_M} \int_{(\mathbf{u}+h\mathbf{s},\mathbf{v}+h\mathbf{t})\in[0,1]^{P+Q}} \frac{c^{3M}_{\mathbf{UV}}(\mathbf{u}+h\mathbf{s},\mathbf{v}+h\mathbf{t})}{\mathbf{A}^P_0(\mathbf{u},h)\mathbf{A}^Q_0(\mathbf{v},h)}\mathbf{K}^P(\mathbf{s})\mathbf{K}^Q(\mathbf{t})d\mathbf{s}d\mathbf{t} \\
= &\inf_{(\mathbf{u},\mathbf{v})\in A_M} \int_{\substack{(\mathbf{u}+h\mathbf{s},\mathbf{v}+h\mathbf{t})\in[0,1]^{P+Q} \\ (\mathbf{s},\mathbf{t}\in[-N_0,N_0]^{P+Q})}} \frac{c^{3M}_{\mathbf{UV}}(\mathbf{u}+h\mathbf{s},\mathbf{v}+h\mathbf{t})}{\mathbf{A}^P_0(\mathbf{u},h)\mathbf{A}^Q_0(\mathbf{v},h)}\mathbf{K}^P(\mathbf{s})\mathbf{K}^Q(\mathbf{t})d\mathbf{s}d\mathbf{t} + C\epsilon \\
> &\frac{11M}{4\mathbf{A}^P_0(\mathbf{u},h)\mathbf{A}^Q_0(\mathbf{v},h)}\big[(\frac{3}{4})^{\frac{P}{P+Q}}\int_{\mathbf{u}+h\mathbf{s}\in[0,1]^P} K(\mathbf{s})d\mathbf{s}\big]\big[(\frac{3}{4})^{\frac{Q}{P+Q}}\int_{\mathbf{v}+h\mathbf{t}\in[0,1]^Q} K(\mathbf{t})d\mathbf{t}\big] + C\epsilon \\
> &2M.
\end{aligned}$$

Notice that, for any $(\mathbf{u},\mathbf{v})\in A_M$,

$$\begin{aligned}
Var\Big(\frac{1}{n}\sum_{j\neq i}\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v})\Big) &= \frac{1}{n}Var(\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U}-\mathbf{u})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{V}-\mathbf{v})) \\
&\leq \frac{1}{n}E(\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U}-\mathbf{u})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{V}-\mathbf{v}))^2 \\
&= \frac{1}{nh^{2P+2Q}}\int_{[0,1]^P}\Big(\mathbf{K}^P(\frac{\mathbf{u}_1-\mathbf{u}}{h})\mathbf{K}^Q(\frac{\mathbf{v}_1-\mathbf{v}}{h})\Big)^2 c_{\mathbf{UV}}(\mathbf{u}_1,\mathbf{v}_1)d\mathbf{u}_1d\mathbf{v}_1 \\
&\leq \frac{K_0^{P+Q}}{nh^{P+Q}}E(\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U}-\mathbf{u})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{V}-\mathbf{v})).
\end{aligned}$$

Using the same argument in proof of lemma 7, we have

$$\sup_{(\mathbf{u},\mathbf{v})\in A_M}\left|\frac{\frac{1}{n}\sum_{j=1}^n \mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}) - E\big(\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U}-\mathbf{u})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{V}-\mathbf{v})\big)}{\sqrt{E(\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U}-\mathbf{u})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{V}-\mathbf{v}))}}\right| = o(1)\ a.s..$$

And thus, for all $(\mathbf{u},\mathbf{v})\in A_M$ with $M>2$,

$$\frac{1}{n}\sum_{j=1}^n \mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v}) \geq \frac{1}{2}E\big(\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U}-\mathbf{u})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{V}-\mathbf{v})\big)\ \ a.s..$$

Consequently,

$$\begin{aligned}
&\inf_{(\mathbf{u},\mathbf{v})\in A_M}\frac{\frac{1}{n}\sum_{j=1}^n \mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{u}_j-\mathbf{u})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{v}_j-\mathbf{v})}{\mathbf{A}^P_0(\mathbf{u},h)\mathbf{A}^Q_0(\mathbf{v},h)} \\
\geq &\inf_{(\mathbf{u},\mathbf{v})\in A_M}\frac{\frac{1}{2}E\big(\mathbf{K}^P_{h^2\mathbf{I}_P}(\mathbf{U}-\mathbf{u})\mathbf{K}^Q_{h^2\mathbf{I}_Q}(\mathbf{V}-\mathbf{v})\big)}{\mathbf{A}^P_0(\mathbf{u},h)\mathbf{A}^Q_0(\mathbf{v},h)}\ \ a.s. \\
> &M\ \ a.s.
\end{aligned}$$

□

References

1. Krishnamurthy A, Kandasamy K, Poczos B, Wasserman L (2014) Nonparametric estimation of renyi divergence and friends in *Proceedings of the 31st International Conference on Machine Learning (ICML 2014)*. (Beijing, China), pp. 919–927.
2. Moon KR, Sricharan K, Hero AO (2017) Ensemble estimation of mutual information in *2017 IEEE International Symposium on Information Theory (ISIT)*. (IEEE, Aachen, Germany), pp. 3030–3034.
3. Silverman BW (1978) Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *The Annals of Statistics* 6(1):177–184.
4. Giné E, Guillou A (2002) Rates of strong uniform consistency for multivariate kernel density estimators. *Annales de l'Institut Henri Poincare (B) Probability and Statistics* 38(6):907–921.
5. Duong T, Hazelton ML (2005) Cross-validation bandwidth matrices for multivariate kernel density estimation. *Scandinavian Journal of Statistics,* 32(3):485–506.
6. Scott DW (2015) *Multivariate density estimation: theory, practice, and visualization.* (John Wiley & Sons., New York, NY).
7. Gao W, Kannan S, Oh S, Viswanath P (2017) Estimating mutual information for discrete-continuous mixtures in *Advances in Neural Information Processing Systems 30*, eds. Guyon I, et al. (Curran Associates, Inc., Red Hook, NY), pp. 5986–5997.
8. Heller R, Heller Y, Gorfine M (2012) A consistent multivariate test of association based on ranks of distances. *Biometrika*.
9. Ding AA, Li Y (2013) Copula correlation: An equitable dependence measure and extension of pearson's correlation. *arXiv preprint arXiv:1312.7214*.
10. Kinney JB, Atwal GS (2014) Equitability, mutual information, and the maximal information coefficient. *Proceedings of the National Academy of Sciences* 111(9):3354–3359.
11. Tchebichef P (1867) "des valeurs moyennes". *Journal de mathématiques pures et appliquées* 2(12):177–184.
12. Bernstein S (1924) On a modification of chebyshev's inequality and of the error formula of laplace. *Ann. Sci. Inst. Sav. Ukraine, Sect. Math* 1(4):38–49.
13. Dvoretzky A, Kiefer J, Wolfowitz J (1956) Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics* pp. 642–669.
14. Serfling RJ (2009) *Approximation theorems of mathematical statistics.* (John Wiley and Sons, New York, NY) Vol. 162.
15. Arcones MA (1995) A bernstein-type inequality for u-statistics and u-processes. *Statistics* & *probability letters* 22(3):239–247.

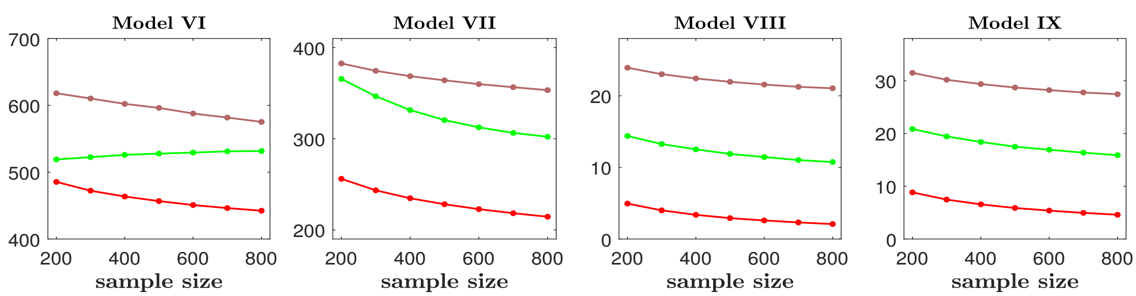

Fig. S1. MSEs against sample sizes for different estimation methods of MI based on different models. JMI: Red, mixed KSG: Green, copula-based KSG: Brown.

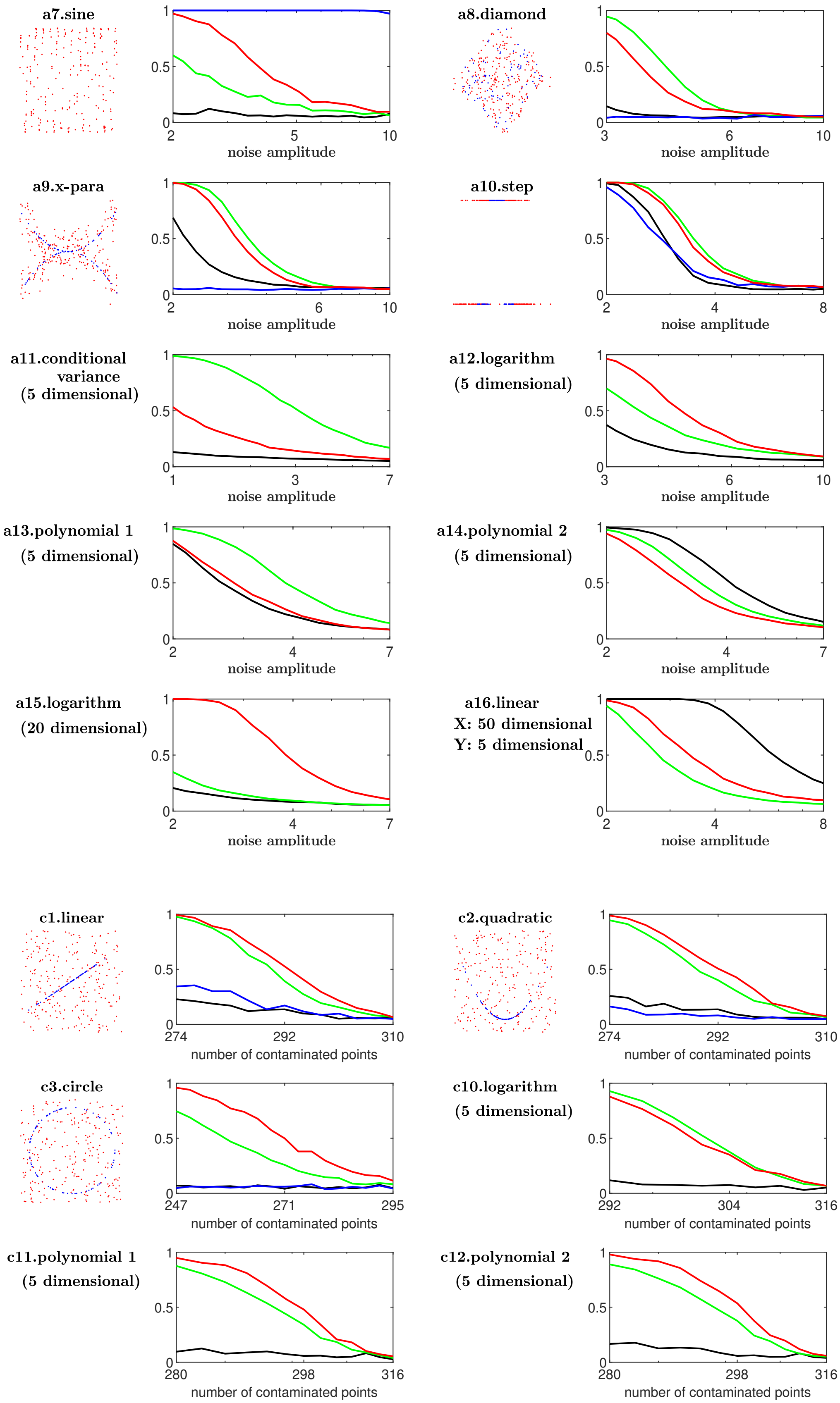

Fig. S2. In each pair of panels, the left is the model and data, and the right is the powers of tests with significance level $\alpha = 0.05$ for models (a7)-(a16) with additive noises, and models (c1)-c(3),(c10)-(c12) with contaminated noises. For the curves of powers, correspondence between colors and different methods are as follows, red for JMI; green for HHG; blue for MIC; black for dCor.