Supplementary Materials for "A semiparametric kernel independence test with application to mutational signatures"

DongHyuk Lee and Bin Zhu*

Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

*email: bin.zhu@nih.gov

S.1 Proof of Theorems

Our works are based on the standard conditions in the kernel density estimation (Hall, 1984; Rosenblatt and Wahlen, 1992) with conditions on zero proportions.

(A1) The pdfs $g(x), g(y), f(x), f(y)$ are supported on $\mathbb{R}$, $f(x,y)$ is supported on $\mathbb{R}^2$ and they are uniformly continuous on respective space.

(A2) The kernel function K is bounded and non-negative on $\mathbb{R}$ and satisfy $\int K(z)dz = 1$, $\int zK(z)dz = 0$ and $\int z^2K(z)dz < \infty$.

(A3) $0 < \beta_{0\cdot}, \beta_{\cdot 0} < 1$, $0 < \beta_{00}, \beta_{10}, \beta_{01}, \beta_{11} < 1$ and $\beta_{11} = 1 - \beta_{00} - \beta_{01} - \beta_{10}$.

(A4) $h \to 0$ and $nh^2 \to \infty$ as $n \to \infty$.

S.1.1 Proof of Theorem 1 (i)

Let $\boldsymbol{Z}_i = [I(X_i = 0, Y_i = 0), I(X_i = 0), I(Y_i = 0)]^T$. Then $E(\boldsymbol{Z}_1) = \boldsymbol{\beta}$ and $Cov(\boldsymbol{Z}_1) = \Sigma_0$ under H_0 which is positive definite since $|\Sigma_0^{(1)}| = \beta_{00}(1-\beta_{00}) > 0$, $|\Sigma_0^{(2)}| = \beta_{00}\beta_{0\cdot}(1-\beta_{0\cdot})(1-\beta_{\cdot 0}) > 0$, and $|\Sigma_0| = \beta_{00}^2(1-\beta_{0\cdot})^2(1-\beta_{\cdot 0})^2 > 0$ under (A3), where $\Sigma_0^{(k)}$ is the order k leading principal submatrix and $|A|$ is the determinant of a matrix A. Then by the multivariate central limit theorem,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Z}_i - \boldsymbol{\beta}\right) \to \mathrm{N}(\boldsymbol{0}, \Sigma_0)$$

in distribution. However, $J(\boldsymbol{\beta}) = \beta_{00} - \beta_{0\cdot}\beta_{\cdot 0} = 0$ under H_0 and the delta method will lead to $\sqrt{n}(\widehat{\beta}_{00} - \widehat{\beta}_{0\cdot}\widehat{\beta}_{\cdot 0}) \to \mathrm{N}(0, \nabla^T J(\boldsymbol{\beta})\Sigma_0 \nabla J(\boldsymbol{\beta}))$ in distribution and the result follows from $\widehat{T}_1 = (\widehat{\beta}_{00} - \widehat{\beta}_{0\cdot}\widehat{\beta}_{\cdot 0})^2$.

S.1.2 Proof of Theorem 1 (ii)

First we note that

$$
\begin{aligned}
\widehat{T}_2 &= \int \{\widehat{\beta}_{10}\widehat{f}(x) - (1-\widehat{\beta}_{0\cdot})\widehat{\beta}_{\cdot 0}\widehat{g}(x)\}^2 dx \\
&= \int \Big\{ \frac{1}{nh}\sum_{i=1}^n I(X_i \neq 0, Y_i = 0) K\Big(\frac{x - X_i}{h}\Big) - \frac{1}{nh}\sum_{i=1}^n \widehat{\beta}_{\cdot 0} I(X_i \neq 0) K\Big(\frac{x - X_i}{h}\Big)\Big\}^2 dx \\
&= \int \Big[\frac{1}{nh}\sum_{i=1}^n \{I(Y_i = 0) - \widehat{\beta}_{\cdot 0}\} I(X_i \neq 0) K\Big(\frac{x - X_i}{h}\Big)\Big]^2 dx \\
&= \frac{1}{n^2h^2}\sum_{i=1}^n \int \{I(Y_i = 0) - \widehat{\beta}_{\cdot 0}\}^2 I(X_i \neq 0) K^2\Big(\frac{x - X_i}{h}\Big) dx \\
&+ \frac{2}{n^2h^2}\sum_{1\le i<j\le n} \int \Big[\{I(Y_i = 0) - \widehat{\beta}_{\cdot 0}\} I(X_i \neq 0) K\Big(\frac{x - X_i}{h}\Big)\Big]\Big[\{I(Y_j = 0) - \widehat{\beta}_{\cdot 0}\} I(X_j \neq 0) K\Big(\frac{x - X_j}{h}\Big)\Big] dx \\
&\equiv \widehat{T}_2^{(1)} + \widehat{T}_2^{(2)}.
\end{aligned}
$$

Now let $\widehat{T}_2^{(1)} = (1/n^2h^2)\sum_{i=1}^n Z_{2,i}$ and $\widehat{T}_2^{(2)} = (2/n^2h^2)\sum_{1\le i<j\le n} H_n(X_i, Y_i, X_j, Y_j)$. Then $H_n(X_i, Y_i, X_j, Y_j)$ is symmetric by definition and $E[H_n(X_i, Y_i, X_j, Y_j)|X_i, Y_i] = 0$ under H_0. Then we have the following result.

Lemma 1. $E[\{I(Y_1 = 0) - \widehat{\beta}_{\cdot 0}\}]^4 = \beta_{\cdot 0}(1-\beta_{\cdot 0})(1 - 3\beta_{\cdot 0} + 3\beta_{\cdot 0}^2) + O(1/n)$.

Proof. From $E[\{I(Y_1 = 0) - \widehat{\beta}_{\cdot 0}\}]^4 = E\{I(Y_1 = 0)\} - 4E\{I(Y_1 = 0)\widehat{\beta}_{\cdot 0}\} + 6E\{I(Y_1 = 0)\widehat{\beta}_{\cdot 0}^2\} - 4E\{I(Y_1 = 0)\widehat{\beta}_{\cdot 0}^3\} + E(\widehat{\beta}_{\cdot 0}^4)$,

$$
\begin{aligned}
E(\widehat{\beta}_{\cdot 0}^4) &= \frac{1}{n^4}E\Big[\sum_{i=1}^n I^4(Y_i = 0) + \sum_{1\le i<j\le n}\{4I^3(Y_i = 0)I(Y_j = 0) + 6I^2(Y_i = 0)I^2(Y_j = 0) + 4I(Y_i = 0)I^3(Y_j = 0)\} \\
&\quad + 12\sum_{1\le i<j<k\le n}\{I^2(Y_i = 0)I(Y_j = 0)I(Y_k = 0) + I(Y_i = 0)I^2(Y_j = 0)I(Y_k = 0) + I(Y_i = 0)I(Y_j = 0)I^2(Y_k = 0)\} \\
&\quad + 24\sum_{1\le i<j<k<\ell\le n} I(Y_i = 0)I(Y_j = 0)I(Y_k = 0)I(Y_\ell = 0)\Big] \\
&= \frac{1}{n^3}\{\beta_{\cdot 0} + 7(n-1)\beta_{\cdot 0}^2 + 6(n-1)(n-2)\beta_{\cdot 0}^3 + (n-1)(n-2)(n-3)\beta_{\cdot 0}^4\} = \beta_{\cdot 0}^4 + O(1/n),
\end{aligned}
$$

$$
\begin{aligned}
E\{I(Y_1=0)\widehat{\beta}_{\cdot 0}^3\} = \frac{1}{n^3}E\Big[& I^4(Y_1=0)+\sum_{i=2}^n I(Y_1=0)I^3(Y_i=0)+3\sum_{i=2}^n I^3(Y_1=0)I(Y_i=0) \\
&+3\sum_{2\le i<j\le n} I(Y_1=0)I^2(Y_i=0)I(Y_j=0)+3\sum_{i=2}^n I^2(Y_1=0)I^2(Y_i=0) \\
&+3\sum_{2\le i<j\le n} I(Y_1=0)I(Y_i=0)I^2(Y_j=0)+6\sum_{2\le i<j\le n} I^2(Y_1=0)I(Y_i=0)I(Y_j=0) \\
&+6\sum_{2\le i<j<k\le n} I(Y_1=0)I(Y_i=0)I(Y_j=0)I(Y_k=0)\Big] \\
&= \frac{1}{n^3}\{\beta_{\cdot 0}+7(n-1)\beta_{\cdot 0}^2+6(n-1)(n-2)\beta_{\cdot 0}^3+(n-1)(n-2)(n-3)\beta_{\cdot 0}^4\} = \beta_{\cdot 0}^4+O(1/n),
\end{aligned}
$$

$$
\begin{aligned}
E\{I(Y_1=0)\widehat{\beta}_{\cdot 0}^2\} = \frac{1}{n^2}E\Big[& I^3(Y_1=0)+\sum_{i=2}^n I(Y_1=0)I^2(Y_i=0)+2\sum_{i=2}^n I^2(Y_1=0)I(Y_i=0) \\
&+2\sum_{2\le i<j\le n} I(Y_1=0)I(Y_i=0)I(Y_j=0)\Big] \\
&= \frac{1}{n^2}\{\beta_{\cdot 0}+3(n-1)\beta_{\cdot 0}^2+(n-1)(n-2)\beta_{\cdot 0}^3\} = \beta_{\cdot 0}^3+O(1/n),
\end{aligned}
$$

$E\{I(Y_1=0)\widehat{\beta}_{\cdot 0}\} = (1/n)E\{I(Y_1=0)+\sum_{i=2}^n I(Y_1=0)I(Y_i=0)\} = \beta_{\cdot 0}^2+O(1/n)$ and $E\{I(Y_1=0)\} = \beta_{\cdot 0}$. □

The next Lemmas show that the condition in which $\widehat{T}_2^{(2)}$ is asymptotically normally distributed for Theorem 1 in Hall (1984). All the expectations in these Lemmas are taken under H_0.

Lemma 2. *$E\{H_n^2(X_i,Y_i,X_j,Y_j)\} = h^3C_1+o(h^3)$, where $C_1 = \{\beta_{\cdot 0}(1-\beta_{\cdot 0})(1-\beta_{0\cdot})\}^2\{\int g(x)^2dx\}\iint \{\int K(z)K(z+z')\}^2dz'$.*

Proof.

$$
\begin{aligned}
&E\{H_n^2(X_i,Y_i,X_j,Y_j)\} \\
&= E\Big(\int\Big[\big\{I(Y_i=0)-\widehat{\beta}_{\cdot 0}\big\}I(X_i\ne 0)K\Big(\frac{x-X_i}{h}\Big)\Big]\Big[\big\{I(Y_j=0)-\widehat{\beta}_{\cdot 0}\big\}I(X_j\ne 0)K\Big(\frac{x-X_j}{h}\Big)\Big]dx\Big)^2 \\
&= \iint E\Big(\Big[\big\{I(Y_i=0)-\widehat{\beta}_{\cdot 0}\big\}I(X_i\ne 0)K\Big(\frac{x-X_i}{h}\Big)\Big]\Big[\big\{I(Y_i=0)-\widehat{\beta}_{\cdot 0}\big\}I(X_i\ne 0)K\Big(\frac{y-X_i}{h}\Big)\Big]\Big) \\
&\quad\times E\Big(\Big[\big\{I(Y_j=0)-\widehat{\beta}_{\cdot 0}\big\}I(X_j\ne 0)K\Big(\frac{x-X_j}{h}\Big)\Big]\Big[\big\{I(Y_j=0)-\widehat{\beta}_{\cdot 0}\big\}I(X_j\ne 0)K\Big(\frac{y-X_j}{h}\Big)\Big]\Big)dxdy
\end{aligned}
$$

$$= \iint \Big\{ E\Big(\Big[\{ I(Y_i = 0) - \widehat{\beta}_{\cdot 0} \} I(X_i \neq 0) K\Big(\frac{x - X_i}{h}\Big) \Big] \Big[\{ I(Y_i = 0) - \widehat{\beta}_{\cdot 0} \} I(X_i \neq 0) K\Big(\frac{y - X_i}{h}\Big) \Big] \Big) \Big\}^2 dxdy$$
$$= \iint \Big(E\Big[\{ I(Y_i = 0) - \widehat{\beta}_{\cdot 0} \}^2 I(X_i \neq 0) K\Big(\frac{x - X_i}{h}\Big) K\Big(\frac{y - X_i}{h}\Big) \Big] \Big)^2 dxdy$$
$$= \Big[E\{ I(Y_i = 0) - \widehat{\beta}_{\cdot 0} \}^2 \Big]^2 \iint \Big[E\{ I(X_i \neq 0) K\Big(\frac{x - X_i}{h}\Big) K\Big(\frac{y - X_i}{h}\Big) \} \Big]^2 dxdy,$$

where $E\{I(Y_i = 0) - \widehat{\beta}_{\cdot 0}\}^2 = E\{I(Y_1 = 0)\} - 2E\{I(Y_1 = 0)\widehat{\beta}_{\cdot 0}\} + E\{\widehat{\beta}^2_{\cdot 0}\} = \beta_{\cdot 0} - 2\beta^2_{\cdot 0} - 2\beta_{\cdot 0}(1 - \beta_{\cdot 0})/n + \beta^2_{\cdot 0} + \beta_{\cdot 0}(1 - \beta_{\cdot 0})/n = (1 - 1/n)\beta_{\cdot 0}(1 - \beta_{\cdot 0})$ and

$$\iint \Big[E\{ I(X_i \neq 0) K\Big(\frac{x - X_i}{h}\Big) K\Big(\frac{y - X_i}{h}\Big) \} \Big]^2 dxdy$$
$$= \iint \Big\{ \int (1 - \beta_{0\cdot}) K\Big(\frac{x - u}{h}\Big) K\Big(\frac{y - u}{h}\Big) g(u) du \Big\}^2 dxdy$$
$$= (1 - \beta_{0\cdot})^2 h^3 \Big\{ \int g(x) dx \Big\} \int \Big\{ \int K(z) K(z + z') dz \Big\}^2 dxdz' + o(h^3).$$

$\square$

Lemma 3. *$E\{H_n^4(X_i, Y_i, X_j, Y_j)\} = h^5 C_2 + o(h^5)$, where $C_2 = \beta_{\cdot 0}(1 - \beta_{\cdot 0})(1 - 3\beta_{\cdot 0} + 3\beta^2_{\cdot 0})(1 - \beta_{0\cdot})^2 \{\int g^2(x) dx\} [\iint K(s_1) K(s_2) \{\int K(s_1 + s_2) K(s_1' + s_2) ds_2\}^3 ds_1 ds_1']$.*

Proof.

$$E\{H_n^4(X_i, Y_i, X_j, Y_j)\}$$
$$= E\Big(\int \Big[\{ I(Y_i = 0) - \widehat{\beta}_{\cdot 0} \} I(X_i \neq 0) K\Big(\frac{x - X_i}{h}\Big) \Big] \Big[\{ I(Y_j = 0) - \widehat{\beta}_{\cdot 0} \} I(X_j \neq 0) K\Big(\frac{x - X_j}{h}\Big) \Big] dx \Big)^4$$
$$= \iiiint E\Big(\Big[\{ I(Y_i = 0) - \widehat{\beta}_{\cdot 0} \} I(X_i \neq 0) K\Big(\frac{x - X_i}{h}\Big) \Big] \Big[\{ I(Y_i = 0) - \widehat{\beta}_{\cdot 0} \} I(X_i \neq 0) K\Big(\frac{y - X_i}{h}\Big) \Big]$$
$$\times \Big[\{ I(Y_i = 0) - \widehat{\beta}_{\cdot 0} \} I(X_i \neq 0) K\Big(\frac{z - X_i}{h}\Big) \Big] \Big[\{ I(Y_i = 0) - \widehat{\beta}_{\cdot 0} \} I(X_i \neq 0) K\Big(\frac{w - X_i}{h}\Big) \Big] \Big)$$
$$\times E\Big(\Big[\{ I(Y_j = 0) - \widehat{\beta}_{\cdot 0} \} I(X_j \neq 0) K\Big(\frac{x - X_j}{h}\Big) \Big] \Big[\{ I(Y_j = 0) - \widehat{\beta}_{\cdot 0} \} I(X_j \neq 0) K\Big(\frac{y - X_j}{h}\Big) \Big]$$
$$\times \Big[\{ I(Y_j = 0) - \widehat{\beta}_{\cdot 0} \} I(X_j \neq 0) K\Big(\frac{x - X_j}{h}\Big) \Big] \Big[\{ I(Y_j = 0) - \widehat{\beta}_{\cdot 0} \} I(X_j \neq 0) K\Big(\frac{y - X_j}{h}\Big) \Big] \Big) dxdydzdw$$
$$= [E\{ I(Y_1 = 0) - \widehat{\beta}_{\cdot 0} \}^4]^2 \iiiint E\Big\{ I^4(X_i \neq 0) K\Big(\frac{x - X_i}{h}\Big) K\Big(\frac{y - X_i}{h}\Big) K\Big(\frac{z - X_i}{h}\Big) K\Big(\frac{w - X_i}{h}\Big) \Big\}^2 dxdydzdw,$$

where $[E\{I(Y_1 = 0) - \widehat{\beta}_{\cdot 0}\}^4]^2$ is given in Lemma 1 and

$$\iiiint E\Big\{I^4(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-X_i}{h}\Big)K\Big(\frac{z-X_i}{h}\Big)K\Big(\frac{w-X_i}{h}\Big)\Big\}^2 dxdydzdw$$
$$=\iiiint \Big\{(1-\beta_{0\cdot})\int K\Big(\frac{x-u}{h}\Big)K\Big(\frac{y-u}{h}\Big)K\Big(\frac{z-u}{h}\Big)K\Big(\frac{w-u}{h}\Big)g(u)du\Big\}^2 dxdydzdw$$
$$=(1-\beta_{0\cdot})^2h^5\Big\{\int g^2(x)dx\Big\}\Big[\iiint K(s_1)K(s_2)\Big\{\int K(s_1+s_2)K(s_1'+s_2)ds_2\Big\}^3 ds_1ds_1'\Big]+o(h^5).$$

□

Lemma 4. *$E\{G_n^2(X_i,Y_i,X_j,Y_j)\}=h^7C_3+o(h^7)$, where $C_3=\{\beta_{\cdot 0}(1-\beta_{\cdot 0})(1-\beta_{0\cdot})\}^4\{\int g^4(x)dx\}\iiiint$ $\{\int K(z)K(z+s_1)dz\}\{\int K(z)K(z+s_2)dz\}\{\int K(z)K(z+s_3)dz\}\{\int K(z)K(z+s_1+s_2-s_3)dz\}ds_1ds_2ds_3$ and $G_n(x_1,y_1,x_2,y_2)=E\{H_n(X_i,Y_i,x_1,y_1)H_n(X_i,Y_i,x_2,y_2)\}$.*

Proof. First we compute

$$\begin{aligned}
&G_n(x_1,y_1,x_2,y_2)\\
&=\iint E\Big(\Big[\{I(Y_i=0)-\widehat{\beta}_{\cdot 0}\}I(X_i\neq 0)K\Big(\frac{a-X_i}{h}\Big)\Big]\Big[\{I(y_1=0)-\widehat{\beta}_{\cdot 0}\}I(x_1\neq 0)K\Big(\frac{a-x_1}{h}\Big)\Big]\\
&\qquad\Big[\{I(Y_i=0)-\widehat{\beta}_{\cdot 0}\}I(X_i\neq 0)K\Big(\frac{b-X_i}{h}\Big)\Big]\Big[\{I(y_2=0)-\widehat{\beta}_{\cdot 0}\}I(x_2\neq 0)K\Big(\frac{b-x_2}{h}\Big)\Big]\Big)dadb\\
&=\big[E\{I(Y_i=0)-\widehat{\beta}_{\cdot 0}\}^2\big]\{I(y_1=0)-\widehat{\beta}_{\cdot 0}\}\{I(y_2=0)-\widehat{\beta}_{\cdot 0}\}\\
&\quad\iint E\Big[\Big\{I(X_i\neq 0)K\Big(\frac{a-X_i}{h}\Big)K\Big(\frac{b-X_i}{h}\Big)\Big\}\Big]\Big\{I(x_1\neq 0)K\Big(\frac{a-x_1}{h}\Big)\Big\}\Big\{I(x_2\neq 0)K\Big(\frac{b-x_2}{h}\Big)\Big\}dadb.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&E\{G_n^2(X_i,Y_i,X_j,Y_j)\}=\big[E\{I(Y_i=0)-\widehat{\beta}_{\cdot 0}\}^2\big]^2\big[E\{I(Y_i=0)-\widehat{\beta}_{\cdot 0}\}^2\{I(Y_j=0)-\widehat{\beta}_{\cdot 0}\}^2\big]\\
&\quad E\Big(\iint E\Big[\Big\{I(X_i\neq 0)K\Big(\frac{a-X_i}{h}\Big)K\Big(\frac{b-X_i}{h}\Big)\Big\}\Big]\Big\{I(X_i\neq 0)K\Big(\frac{a-X_i}{h}\Big)\Big\}\Big\{I(X_j\neq 0)K\Big(\frac{b-X_j}{h}\Big)\Big\}dadb\Big)^2\\
&=\big[E\{I(Y_i=0)-\widehat{\beta}_{\cdot 0}\}^2\big]^4\iiiint E\Big\{I(X_i\neq 0)K\Big(\frac{a-X_i}{h}\Big)K\Big(\frac{b-X_i}{h}\Big)\Big\}E\Big\{I(X_i\neq 0)K\Big(\frac{a'-X_i}{h}\Big)K\Big(\frac{b'-X_i}{h}\Big)\Big\}\\
&\qquad E\Big\{I(X_i\neq 0)K\Big(\frac{a-X_i}{h}\Big)K\Big(\frac{a'-X_i}{h}\Big)\Big\}E\Big\{I(X_j\neq 0)K\Big(\frac{b-X_j}{h}\Big)K\Big(\frac{b'-X_j}{h}\Big)\Big\}dadbda'db'\\
&=\Big\{\beta_{\cdot 0}(1-\beta_{\cdot 0})\Big(1-\frac{1}{n}\Big)\Big\}^4(1-\beta_{0\cdot})^4h^4\iiiint\Big\{\int K(z)K\Big(\frac{b-a}{h}+z\Big)g(a-zh)dz\Big\}\\
&\qquad\Big\{\int K(z)K\Big(\frac{b'-a'}{h}+z\Big)g(a'-zh)dz\Big\}\Big\{\int K(z)K\Big(\frac{a-a'}{h}+z\Big)g(a-zh)dz\Big\}\\
&\qquad\Big\{\int K(z)K\Big(\frac{b-b'}{h}+z\Big)g(a'-zh)dz\Big\}dadbda'db'
\end{aligned}$$

$$= \big\{\beta_{\cdot 0}(1-\beta_{\cdot 0})(1-\beta_{0\cdot})\big\}^4 h^7 \Big\{\int g^4(x)dx\Big\} \iiiint \Big\{\int K(z)K(z+s_1)dz\Big\}\Big\{\int K(z)K(z+s_2)dz\Big\}$$
$$\Big\{\int K(z)K(z+s_3)dz\Big\}\Big\{\int K(z)K(z+s_1+s_2-s_3)dz\Big\} ds_1 ds_2 ds_3 + o(h^7).$$

□

Now we have

$$\frac{E\{G_n^2(X_i,Y_i,X_j,Y_j)\}}{[E\{H_n^2(X_i,Y_i,X_j,Y_j)\}]^2} = \frac{h^7 C_3 + o(h^7)}{h^6 C_1 + o(h^6)} = h\frac{C3+o(1)}{C_1+o(1)} \to 0$$

and

$$\frac{E\{H_n^4(X_i,Y_i,X_j,Y_j)\}}{n[E\{H_n^2(X_i,Y_i,X_j,Y_j)\}]^2} = \frac{h^5 C_2 + o(h^5)}{nh^6 C_1 + o(nh^6)} = \frac{1}{nh}\frac{C3+o(1)}{C_1+o(1)} \to 0$$

as $n \to \infty$ since $h \to 0$ and $nh \to \infty$. Therefore, $\widehat{T}_2^{(2)}$ is asymptotically normal $\mathrm{N}(0, 2C_1/n^2h)$ due to Theorem 1 in Hall (1984).

Proof of Theorem 1 (ii).
For $\widehat{T}_2^{(1)} = (1/n^2h^2)\sum_{i=1}^n Z_{2,i}$,

$$E(Z_{2,i}) = E\Big[\int \{I(Y_i=0)-\widehat{\beta}_{\cdot 0}\}^2 I(X_i \neq 0)K^2\Big(\frac{x-X_i}{h}\Big)dx\Big]$$
$$= E\Big[\big\{I(Y_i=0)-\widehat{\beta}_{\cdot 0}\big\}^2\Big]\int E\Big\{I(X_i\neq 0)K^2\Big(\frac{x-X_i}{h}\Big)\Big\}dx,$$

where

$$E[\{I(Y_i=0)-\widehat{\beta}_{\cdot 0}\}^2] = E\{I(Y_i=0)\} - 2E\{I(Y_i=0)\widehat{\beta}_{\cdot 0}\} + E(\widehat{\beta}_{\cdot 0}^2) = \frac{(n-1)}{n}\beta_{\cdot 0}(1-\beta_{\cdot 0})$$

and

$$\int E\{I(X_i\neq 0)K^2\Big(\frac{x-X_i}{h}\Big)dx = \iint (1-\beta_{0\cdot})K^2\Big(\frac{x-u}{h}\Big)g(u)dudx = (1-\beta_{0\cdot})h\int K^2(z)dz.$$

Therefore, $E(Z_{2,i}) = (1-1/n)\beta_{\cdot 0}(1-\beta_{\cdot 0})(1-\beta_{0\cdot})h\int K^2(z)dz$. Next,

$$E(Z_{2,i}^2) = E\Big[\int \{I(Y_i=0)-\widehat{\beta}_{\cdot 0}\}^2 I(X_i\neq 0)K^2\Big(\frac{x-X_i}{h}\Big)dx\Big]^2$$
$$= E\iint \Big[\big\{I(Y_i=0)-\widehat{\beta}_{\cdot 0}\big\}^2 I(X_i\neq 0)K^2\Big(\frac{x-X_i}{h}\Big)\Big]\Big[\big\{I(Y_i=0)-\widehat{\beta}_{\cdot 0}\big\}^2 I(X_i\neq 0)K^2\Big(\frac{y-X_i}{h}\Big)\Big]dxdy$$
$$= E\Big[\big\{I(Y_i=0)-\widehat{\beta}_{\cdot 0}\big\}^4\Big]\iint E\Big[I(X_i\neq 0)K^2\Big(\frac{x-X_i}{h}\Big)K^2\Big(\frac{y-X_i}{h}\Big)\Big]dxdy,$$

where $E[\{I(Y_i=0)-\widehat{\beta}_{\cdot 0}\}^4] = \beta_{\cdot 0}(1-\beta_{\cdot 0})(1-3\beta_{\cdot 0}+3\beta_{\cdot 0}^2)+O(1/n)$ by Lemma 1 and

$$
\begin{aligned}
&\iint E\Big[I(X_i\neq 0)K^2\Big(\frac{x-X_i}{h}\Big)K^2\Big(\frac{y-X_i}{h}\Big)\Big]dxdy \\
&\qquad = (1-\beta_{0\cdot})\iiint K^2\Big(\frac{x-u}{h}\Big)K^2\Big(\frac{y-u}{h}\Big)g(u)dudxdy \\
&\qquad = (1-\beta_{0\cdot})h^2\iiint K^2(z)K^2(z+z')g(x-zh)dxdzdz' \\
&\qquad = (1-\beta_{0\cdot})h^2\iint K^2(z)K^2(z+z')dzdz'.
\end{aligned}
$$

However, since

$$
\begin{aligned}
Var(\widehat{T}_2^{(1)}) &= \frac{1}{n^3h^4}Var(Z_{2,i}^2) \\
&\leq \frac{1}{n^3h^4}E(Z_{2,i}^2) \\
&= \frac{1}{n^3h^2}\beta_{\cdot 0}(1-\beta_{\cdot 0})(1-3\beta_{\cdot 0}+3\beta_{\cdot 0}^2)(1-\beta_{0\cdot})\iint K^2(z)K^2(z+z')dzdz' + O\Big(\frac{1}{n^4h^2}\Big) \\
&= O\Big(\frac{1}{n^3h^2}\Big),
\end{aligned}
$$

we have

$$
\widehat{T}_2 = \frac{1}{nh}\beta_{\cdot 0}(1-\beta_{\cdot 0})(1-\beta_{0\cdot})\int K^2(z)dz + O\Big(\frac{1}{n}\Big) + O_p\Big(\frac{1}{n^{3/2}h}\Big) + \frac{\sqrt{2}}{n\sqrt{h}}\sqrt{C_1}Z + o_p\Big(\frac{1}{n\sqrt{h}}\Big),
$$

where Z is a standard normal random variable and the result follows from $O(1/n) = o_p(1/n\sqrt{h})$ and $O(1/n^{3/2}h) = o_p(1/n\sqrt{h})$. □

Note that the proof of Theorem 1 (iii) is completely analogous to the proof of Theorem 1 (ii).

S.1.3 Proof of Theorem 1 (iv)

Under H_0 and the formulations $\widehat{f}(x,y)$, $\widehat{g}(x)$ and $\widehat{g}(y)$ in Section 3.2,

$$
\begin{aligned}
E\{\widehat{\beta}_{11}\widehat{f}(x,y)\} &= E\Big\{\frac{1}{nh^2}\sum_{i=1}^n I(X_i\neq 0, Y_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big)\Big\} \\
&= \frac{1}{h^2}E\Big\{I(X_1\neq 0)K\Big(\frac{x-X_1}{h}\Big)\Big\}E\Big\{I(Y_1\neq 0)K\Big(\frac{y-Y_1}{h}\Big)\Big\} \\
&= E\Big\{\frac{1}{nh}\sum_{i=1}^n I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)\}E\Big\{\frac{1}{nh}\sum_{i=1}^n I(Y_i\neq 0)K\Big(\frac{y-Y_i}{h}\Big)\} \\
&= E\{(1-\beta_{0\cdot})\widehat{g}(x)\}E\{(1-\beta_{\cdot 0})\widehat{g}(y)\}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\widehat{T}_4 &= \iint \left[\widehat{\beta}_{11}\widehat{f}(x,y) - E\{\widehat{\beta}_{11}\widehat{f}(x,y)\} + E\{(1-\beta_{0\cdot})\widehat{g}(x)\}E\{(1-\beta_{\cdot 0})\widehat{g}(y)\} - (1-\widehat{\beta}_{0\cdot})(1-\widehat{\beta}_{\cdot 0})\widehat{g}(x)\widehat{g}(y)\right]^2 dxdy \\
&= \iint \left[\widehat{\beta}_{11}\widehat{f}(x,y) - E\{\widehat{\beta}_{11}\widehat{f}(x,y)\}\right]^2 dxdy \\
&+ \iint \left[(1-\widehat{\beta}_{0\cdot})(1-\widehat{\beta}_{\cdot 0})\widehat{g}(x)\widehat{g}(y) - E\{(1-\beta_{0\cdot})\widehat{g}(x)\}E\{(1-\beta_{\cdot 0})\widehat{g}(y)\}\right]^2 dxdy \\
&- 2\iint \left[\widehat{\beta}_{11}\widehat{f}(x,y) - E\{\widehat{\beta}_{11}\widehat{f}(x,y)\}\right]\left[(1-\widehat{\beta}_{0\cdot})(1-\widehat{\beta}_{\cdot 0})\widehat{g}(x)\widehat{g}(y) - E\{(1-\beta_{0\cdot})\widehat{g}(x)\}E\{(1-\beta_{\cdot 0})\widehat{g}(y)\}\right] dxdy \\
&\equiv \widehat{T}_4^{(1)} + \widehat{T}_4^{(2)} - 2\widehat{T}_4^{(3)}.
\end{aligned}
$$

The following Lemmas are modifications of Lemmas 1, 2 and 3 in Rosenblatt and Wahlen (1992) mainly due to zero proportions and in that sense $\widehat{T}_4$ is a generalization of their statistic I_n. A detailed proof for I_n can be found in Wahlen (1992).

Lemma 5. *$\widehat{T}_4^{(1)} = (1/nh^2)(1-\beta_{0\cdot})(1-\beta_{\cdot 0})\{\int K^2(z)dz\}^2 + O(1/n) + O_p(1/n^{3/2}h^2) + (\sqrt{2}/nh)\sqrt{C_4}Z$, where $C_4 = (1-\beta_{0\cdot})^2(1-\beta_{\cdot 0})^2\{\int g^2(x)dx\}\{\int g^2(y)dy\}[\int\{\int K(z)K(z+z')dz\}^2 dz']^2$.*

Proof.

$$
\begin{aligned}
&\widehat{T}_4^{(1)} \\
&= \iint \left(\frac{1}{nh^2}\sum_{i=1}^n \left[I(X_i\neq 0, Y_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big) - E\Big\{I(X_i\neq 0, Y_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big)\Big\}\right]\right)^2 dxdy \\
&= \frac{1}{(nh^2)^2}\sum_{i=1}^n \iint \left[I(X_i\neq 0, Y_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big) - E\Big\{I(X_i\neq 0, Y_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big)\Big\}\right]^2 dxdy \\
&+ \frac{2}{(nh^2)^2}\sum_{1\le i<j\le n} \iint \left[I(X_i\neq 0, Y_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big) - E\Big\{I(X_i\neq 0, Y_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big)\Big\}\right] \\
&\qquad \left[I(X_j\neq 0, Y_j\neq 0)K\Big(\frac{x-X_j}{h}\Big)K\Big(\frac{y-Y_j}{h}\Big) - E\Big\{I(X_j\neq 0, Y_j\neq 0)K\Big(\frac{x-X_j}{h}\Big)K\Big(\frac{y-Y_j}{h}\Big)\Big\}\right] dxdy \\
&\equiv \widehat{T}_{41}^{(1)} + \widehat{T}_{42}^{(1)}.
\end{aligned}
$$

Similarly, let $\widehat{T}_{41}^{(1)} = (1/n^2h^4)\sum_{i=1}^n Z_{41,i}$ and $\widehat{T}_{41}^{(2)} = (2/n^2h^4)\sum_{1\le i<j\le n} H_{4n}(X_i, Y_i, X_j, Y_j)$. Then,

$$
\begin{aligned}
E(Z_{41,i}) = &\iint E\Big\{I(X_i\neq 0, Y_i\neq 0)K^2\Big(\frac{x-X_i}{h}\Big)K^2\Big(\frac{y-Y_i}{h}\Big)\Big\} dxdy \\
&- \iint \left[E\Big\{I(X_i\neq 0, Y_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big)\Big\}\right]^2 dxdy,
\end{aligned}
$$

where

$$\begin{aligned}
&\iint E\Big\{I(X_i \neq 0, Y_i \neq 0)K^2\Big(\frac{x-X_i}{h}\Big)K^2\Big(\frac{y-Y_i}{h}\Big)\Big\}dxdy \\
&= \int E\Big\{I(X_i \neq 0)K^2\Big(\frac{x-X_i}{h}\Big)\Big\}dx \int E\Big\{I(Y_i \neq 0)K^2\Big(\frac{y-Y_i}{h}\Big)\Big\}dy \\
&= \iint (1-\beta_{0\cdot})K^2\Big(\frac{x-u}{h}\Big)g(u)du \iint (1-\beta_{\cdot 0})K^2\Big(\frac{y-u}{h}\Big)g(u)du \\
&= (1-\beta_{0\cdot})(1-\beta_{\cdot 0})h^2\Big\{\int K^2(z)dz\Big\}^2
\end{aligned}$$

and

$$\begin{aligned}
&\iint \Big[E\Big\{I(X_i \neq 0, Y_i \neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big)\Big\}\Big]^2 dxdy \\
&= \iint \Big\{(1-\beta_{0\cdot})(1-\beta_{\cdot 0})\iint K\Big(\frac{x-u}{h}\Big)K\Big(\frac{y-v}{h}\Big)g(u)g(v)\Big\}^2 dxdy \\
&= \{(1-\beta_{0\cdot})(1-\beta_{\cdot 0})\}^2 h^4\Big\{\iint K(z)K(z+z')dzdz'\Big\}^2\Big\{\int g^2(x)dx\Big\}\Big\{\int g^2(y)dy\Big\} + o(h^4).
\end{aligned}$$

Next, from Lemma 2 of Hall (1984) and Lemma 1 of Rosenblatt and Wahlen (1992)

$$E(Z_{41,i}^2) = (1-\beta_{0\cdot})(1-\beta_{\cdot 0})h^4\Big\{\iint K^2(z)K^2(z+z')dzdz'\Big\} + o(h^4).$$

Now, $H_{4n}(X_i, Y_i, X_j, Y_j)$ is symmetric and degenerate. In addition,

$$\begin{aligned}
&E\{H_{4n}^2(X_i, Y_i, X_j, Y_j)\} \\
&= E\Big(\iint \Big[I(X_i \neq 0, Y_i \neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big) - E\Big\{I(X_i \neq 0, Y_i \neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big)\Big\}\Big] \\
&\qquad \Big[I(X_j \neq 0, Y_j \neq 0)K\Big(\frac{x-X_j}{h}\Big)K\Big(\frac{y-Y_j}{h}\Big) - E\Big\{I(X_j \neq 0, Y_j \neq 0)K\Big(\frac{x-X_j}{h}\Big)K\Big(\frac{y-Y_j}{h}\Big)\Big\}\Big]dxdy\Big)^2 \\
&= \iiiint \Big[E\Big\{I(X_i \neq 0, Y_i \neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big)K\Big(\frac{x'-X_i}{h}\Big)K\Big(\frac{y'-Y_i}{h}\Big)\Big\}\Big]^2 dxdydx'dy' \\
&+ \iiiint \Big[E\Big\{I(X_i \neq 0, Y_i \neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big)\Big\}E\Big\{I(X_i \neq 0, Y_i \neq 0)K\Big(\frac{x'-X_i}{h}\Big)K\Big(\frac{y'-Y_i}{h}\Big)\Big\}\Big]^2 dxdydx'dy' \\
&- 2\iiiint E\Big\{I(X_i \neq 0, Y_i \neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big)K\Big(\frac{x'-X_i}{h}\Big)K\Big(\frac{y'-Y_i}{h}\Big)\Big\} \\
&\qquad E\Big\{I(X_i \neq 0, Y_i \neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big)\Big\}E\Big\{I(X_i \neq 0, Y_i \neq 0)K\Big(\frac{x'-X_i}{h}\Big)K\Big(\frac{y'-Y_i}{h}\Big)\Big\}dxdydx'dy' \\
&= (1-\beta_{0\cdot})^2(1-\beta_{\cdot 0})^2 h^6\Big\{\int g^2(x)dx\Big\}\Big\{\int g^2(y)dy\Big\}\Big[\int\Big\{\int K(z)K(z+z')dz\Big\}^2 dz'\Big]^2 \\
&\qquad + (1-\beta_{0\cdot})^4(1-\beta_{\cdot 0})^4 h^8\Big\{\int g^2(x)dx\Big\}^2\Big\{\int g^2(y)dy\Big\}^2 \\
&\qquad - 2(1-\beta_{0\cdot})^3(1-\beta_{\cdot 0})^3 h^8\Big\{\int g^2(x)dx\Big\}\Big\{\int g^2(y)dy\Big\} + o(h^8) = h^6 C_4 + o(h^6).
\end{aligned}$$

The result follows from Theorem 1 of Hall (1984) where $\widehat{T}_{42}^{(1)}$ is asymptotically normal $\mathrm{N}(0, 2C_4/n^2h^2)$ and $Var(\widehat{T}_{41}^{(1)}) \leq O(1/n^3h^4)$. $\square$

Lemma 6. $E(\widehat{T}_4^{(2)}) = (1/nh)(1-\beta_{0\cdot})(1-\beta_{\cdot 0})\{\int K^2(z)dz\}\{(1-\beta_{\cdot 0})\int g^2(x)dx + (1-\beta_{0\cdot})\int g^2(y)dy\} + o(1/nh)$.

Proof. Similar to Lemma 2 of Rosenblatt and Wahlen (1992), we first note that

$$\begin{aligned}
&(1-\widehat{\beta}_{0\cdot})(1-\widehat{\beta}_{\cdot 0})\widehat{g}(x)\widehat{g}(y) - E\{(1-\beta_{0\cdot})\widehat{g}(x)\}E\{(1-\beta_{\cdot 0})\widehat{g}(y)\}\\
&= \big[(1-\widehat{\beta}_{0\cdot})\widehat{g}(x) - E\{(1-\widehat{\beta}_{0\cdot})\widehat{g}(x)\}\big]\big[(1-\widehat{\beta}_{\cdot 0})\widehat{g}(y) - E\{(1-\widehat{\beta}_{\cdot 0})\widehat{g}(y)\}\big]\\
&+ \big[(1-\widehat{\beta}_{0\cdot})\widehat{g}(x) - E\{(1-\widehat{\beta}_{0\cdot})\widehat{g}(x)\}\big]E\{(1-\widehat{\beta}_{\cdot 0})\widehat{g}(y)\} + \big[(1-\widehat{\beta}_{\cdot 0})\widehat{g}(y) - E\{(1-\widehat{\beta}_{\cdot 0})\widehat{g}(y)\}\big]E\{(1-\widehat{\beta}_{0\cdot})\widehat{g}(x)\}\\
&\equiv S_1(x,y) + S_2(x,y) + S_3(x,y).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\widehat{T}_4^{(2)} &= \iint S_1^2(x,y)dxdy + \iint S_2^2(x,y)dxdy + \iint S^3(x,y)dxdy\\
&+ \iint S_1(x,y)S_2(x,y)dxdy + \iint S_1(x,y)S_3(x,y)dxdy + \iint S_2(x,y)S_3(x,y)dxdy.
\end{aligned}$$

Now,

$$\begin{aligned}
E\Big\{\iint S_1^2(x,y)dxdy\Big\} &= \int E\big[(1-\widehat{\beta}_{0\cdot})\widehat{g}(x) - E\{(1-\widehat{\beta}_{0\cdot})\widehat{g}(x)\}\big]^2dx \int E\big[(1-\widehat{\beta}_{\cdot 0})\widehat{g}(y) - E\{(1-\widehat{\beta}_{\cdot 0})\widehat{g}(y)\}\big]^2dy\\
&= \frac{1}{n^2h^2}(1-\beta_{0\cdot})(1-\beta_{\cdot 0})\Big\{\int K^2(z)dz\Big\}^2 + O\Big(\frac{1}{n^2h}\Big),
\end{aligned}$$

since

$$\begin{aligned}
&\int E\big[(1-\widehat{\beta}_{0\cdot})\widehat{g}(x) - E\{(1-\widehat{\beta}_{0\cdot})\widehat{g}(x)\}\big]^2dx = \int Var\Big\{\frac{1}{nh}\sum_{i=1}^n I(X_i \neq 0)K\Big(\frac{x-X_i}{h}\Big)\Big\}dx\\
&= \frac{1}{nh^2}\bigg(\int E\Big\{I(X_i\neq 0)K^2\Big(\frac{x-X_i}{h}\Big)\Big\}dx - \int\Big[E\Big\{I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)\Big\}\Big]^2dx\bigg)\\
&= \frac{1}{nh^2}\Big\{\iint(1-\beta_{0\cdot})K^2\Big(\frac{x-u}{h}\Big)g(u)dudx - \iiint(1-\beta_{0\cdot})^2K\Big(\frac{x-u}{h}\Big)K\Big(\frac{x-u'}{h}\Big)g(u)g(u')dudu'dx\Big\}\\
&= \frac{1}{nh}(1-\beta_{0\cdot})\int K^2(z)dz - \frac{1}{n}(1-\beta_{0\cdot})^2\int g^2(x)dx = \frac{1}{nh}(1-\beta_{0\cdot})\int K^2(z)dz + O\Big(\frac{1}{n}\Big).
\end{aligned}$$

Next,

$$\begin{aligned}
E\Big\{\iint S_2^2(x,y)dxdy\Big\} &= \int E\big[(1-\widehat{\beta}_{0\cdot})\widehat{g}(x) - E\{(1-\widehat{\beta}_{0\cdot})\widehat{g}(x)\}\big]^2dx\int\big[E\{(1-\widehat{\beta}_{\cdot 0})\widehat{g}(y)\}\big]^2dy\\
&= \frac{1}{nh}(1-\beta_{0\cdot})(1-\beta_{\cdot 0})^2\Big\{\int K^2(z)dz\Big\}\Big\{\int g^2(y)dy\Big\} + O\Big(\frac{1}{n}\Big),
\end{aligned}$$

since

$$\int \big[E\{(1-\widehat{\beta}_{\cdot 0})\widehat{g}(y)\}\big]^2 dy = \frac{1}{h^2}\int \Big[E\Big\{I(Y_i \neq 0)K\Big(\frac{y-Y_i}{h}\Big)\Big\}\Big]^2 dy = (1-\beta_{\cdot 0})^2 \int g^2(y)dy + o(1).$$

Similarly,

$$E\Big\{\iint S_3^2(x,y)dxdy\Big\} = \frac{1}{nh}(1-\beta_{0\cdot})^2(1-\beta_{\cdot 0})\Big\{\int K^2(z)dz\Big\}\Big\{\int g^2(x)dx\Big\} + O\Big(\frac{1}{n}\Big).$$

And

$$\begin{aligned} E\Big\{\iint S_1(x,y)S_2(x,y)dxdy\Big\} = \iint E\Big(&\big[(1-\widehat{\beta}_{0\cdot})\widehat{g}(x) - E\{(1-\widehat{\beta}_{0\cdot})\widehat{g}(x)\}\big]^2 \\ &\big[(1-\widehat{\beta}_{\cdot 0})\widehat{g}(y) - E\{(1-\widehat{\beta}_{\cdot 0})\widehat{g}(y)\}\big]E\{(1-\widehat{\beta}_{\cdot 0})\widehat{g}(y)\}\Big)dxdy = 0, \end{aligned}$$

since $E[(1-\widehat{\beta}_{\cdot 0})\widehat{g}(y) - E\{(1-\widehat{\beta}_{\cdot 0})\widehat{g}(y)\}] = 0$. Similarly,

$$E\Big\{\iint S_1(x,y)S_3(x,y)dxdy\Big\} = E\Big\{\iint S_2(x,y)S_3(x,y)dxdy\Big\} = 0.$$

Now the result follows by adding all terms above. □

<u>Lemma</u> 7. $E(\widehat{T}_4^{(3)}) = E(\widehat{T}_4^{(2)})$.

Proof. First we note that

$$\begin{aligned} \widehat{T}_4^{(3)} &= \iint \big[\widehat{\beta}_{11}\widehat{f}(x,y) - E\{\widehat{\beta}_{11}\widehat{f}(x,y)\}\big]S_1(x,y)dxdy \\ &+ \iint \big[\widehat{\beta}_{11}\widehat{f}(x,y) - E\{\widehat{\beta}_{11}\widehat{f}(x,y)\}\big]S_2(x,y)dxdy \\ &+ \iint \big[\widehat{\beta}_{11}\widehat{f}(x,y) - E\{\widehat{\beta}_{11}\widehat{f}(x,y)\}\big]S_3(x,y)dxdy. \end{aligned}$$

Then, from

$$\begin{aligned} &E\Big(\iint \big[\widehat{\beta}_{11}\widehat{f}(x,y) - E\{\widehat{\beta}_{11}\widehat{f}(x,y)\}\big]S_1(x,y)dxdy\Big) \\ &= \iint E\Big(\frac{1}{nh^2}\sum_{i=1}^n\Big[I(X_i\neq 0, Y_i \neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big) - E\Big\{I(X_i\neq 0, Y_i \neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big)\Big\}\Big]\Big) \\ &\qquad \Big(\frac{1}{nh}\sum_{i=1}^n\Big[I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big) - E\Big\{I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)\Big\}\Big]\Big) \\ &\qquad \Big(\frac{1}{nh}\sum_{i=1}^n\Big[I(Y_i\neq 0)K\Big(\frac{y-Y_i}{h}\Big) - E\Big\{I(Y_i\neq 0)K\Big(\frac{y-Y_i}{h}\Big)\Big\}\Big]\Big)dxdy \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n^2h^4}\iint E\Bigg(\Big[I(X_i\neq 0,Y_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big)-E\Big\{I(X_i\neq 0,Y_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big)\Big\}\Big]\\
&\qquad\qquad\Big[I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)-E\Big\{I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)\Big\}\Big]\\
&\qquad\qquad\Big[I(Y_i\neq 0)K\Big(\frac{y-Y_i}{h}\Big)-E\Big\{I(Y_i\neq 0)K\Big(\frac{y-Y_i}{h}\Big)\Big\}\Big]\Bigg)dxdy\\
&= \frac{1}{n^2h^4}\Bigg(\iint E\Big[I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)-E\Big\{I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)\Big\}\Big]^2\\
&\qquad\qquad\quad\Big[I(Y_i\neq 0)K\Big(\frac{y-Y_i}{h}\Big)-E\Big\{I(Y_i\neq 0)K\Big(\frac{y-Y_i}{h}\Big)\Big\}\Big]^2dxdy\\
&\qquad+\iint E\Big[I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)-E\Big\{I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)\Big\}\Big]^2\\
&\qquad\qquad\quad\Big[I(Y_i\neq 0)K\Big(\frac{y-Y_i}{h}\Big)-E\Big\{I(Y_i\neq 0)K\Big(\frac{y-Y_i}{h}\Big)\Big\}\Big]E\Big\{I(Y_i\neq 0)K\Big(\frac{y-Y_i}{h}\Big)\Big\}dxdy\\
&\qquad+\iint E\Big[I(Y_i\neq 0)K\Big(\frac{y-Y_i}{h}\Big)-E\Big\{I(Y_i\neq 0)K\Big(\frac{y-Y_i}{h}\Big)\Big\}\Big]^2\\
&\qquad\qquad\quad\Big[I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)-E\Big\{I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)\Big\}\Big]E\Big\{I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)\Big\}dxdy\\
&= \frac{1}{nh^2}\int Var\Big\{I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)\Big\}dx\frac{1}{nh^2}\int Var\Big\{I(Y_i\neq 0)K\Big(\frac{y-Y_i}{h}\Big)\Big\}dy=E\Big\{\iint S_1^2(x,y)dxdy\Big\},
\end{aligned}
$$

since under H_0 expectations inside the integral factor out, $E[I(X_i\neq 0)K((x-X_i)/h)-E\{I(X_i\neq 0)K((x-X_i)/h)\}]=E[I(Y_i\neq 0)K((y-Y_i)/h)-E\{I(Y_i\neq 0)K((y-Y_i)/h)\}]=0$ and

$$
\begin{aligned}
&I(X_i\neq 0,Y_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big)-E\Big\{I(X_i\neq 0,Y_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{y-Y_i}{h}\Big)\Big\}\\
&= \Big[I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)-E\Big\{I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)\Big\}\Big]\Big[I(Y_i\neq 0)K\Big(\frac{y-Y_i}{h}\Big)-E\Big\{I(Y_i\neq 0)K\Big(\frac{y-Y_i}{h}\Big)\Big\}\Big]\\
&\quad+\Big[I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)-E\Big\{I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)\Big\}\Big]E\Big\{I(Y_i\neq 0)K\Big(\frac{y-Y_i}{h}\Big)\Big\}\\
&\quad+\Big[I(Y_i\neq 0)K\Big(\frac{y-Y_i}{h}\Big)-E\Big\{I(Y_i\neq 0)K\Big(\frac{y-Y_i}{h}\Big)\Big\}\Big]E\Big\{I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)\Big\}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&E\Big(\iint[\widehat{\beta}_{11}\widehat{f}(x,y)-E\{\widehat{\beta}_{11}\widehat{f}(x,y)\}]S_2(x,y)dxdy\Big)\\
&=\frac{1}{nh^2}\int Var\Big\{I(X_i\neq 0)K\Big(\frac{x-X_i}{h}\Big)\Big\}dx\int[E\{(1-\widehat{\beta}_{\cdot 0})\widehat{g}(y)\}]^2dy=E\Big\{\iint S_2^2(x,y)dxdy\Big\}
\end{aligned}
$$

and

$$E\Big(\iint[\widehat\beta_{11}\widehat f(x,y)-E\{\widehat\beta_{11}\widehat f(x,y)\}]\,S_3(x,y)dxdy\Big)$$
$$=\frac{1}{nh^2}\int Var\Big\{I(Y_i\neq 0)K\Big(\frac{y-Y_i}{h}\Big)\Big\}dy\int\big[E\{(1-\widehat\beta_{\cdot 0})\widehat g(x)\}\big]^2dy=E\Big\{\iint S_3^2(x,y)dxdy\Big\}.$$

□

From Lemmas 4 and 5 of Rosenblatt and Wahlen (1992), we obtain $Var(\widehat T_4^{(2)})=Var(\widehat T_4^{(3)})=O(1/n^2h)=o(1/n^2h^2)$ where $Var(\widehat T_{42}^{(1)})=O(1/n^2h^2)$. Since we assumed that zero proportions are fixed between $(0,1)$, the orders of $Var(\widehat T_4^{(2)})$ and $Var(\widehat T_4^{(3)})$ do not change. Therefore, Theorem 1 (iv) follows by noting

$$\widehat T_4=\frac{1}{nh^2}(1-\beta_{0\cdot})(1-\beta_{\cdot 0})\Big\{\int K^2(z)dz\Big\}^2+(\sqrt2/nh)\sqrt{C_4}Z$$
$$-\frac{1}{nh}(1-\beta_{0\cdot})(1-\beta_{\cdot 0})\Big\{\int K^2(z)dz\Big\}\Big\{(1-\beta_{\cdot 0})\int g^2(x)dx+(1-\beta_{0\cdot})\int g^2(y)dy\Big\}+o_p\Big(\frac{1}{nh}\Big),$$

where Z is a standard normal random variable.

S.1.4 Proof of Corollary 1

Proof. Since $\mathrm{var}(\widehat T_1)=O(n^{-2}),\mathrm{var}(\widehat T_2)=\mathrm{var}(\widehat T_3)=O(n^{-2}h^{-1})$,

$$\widehat T=\frac{1}{n}\mu_1+\frac{1}{nh}\mu_2+\frac{1}{nh}\mu_3+\frac{1}{nh^2}\mu_4+\frac{\sqrt2}{nh}\sigma_4Z+o_p(\frac{1}{nh}),$$

where Z is the standard normal random variable. □

S.1.5 Proof of Theorem 2

Proof. Note that under H_1,

$$nh^2\widehat T_1=nh^2(\widehat\beta_{00}-\beta_{00}+\beta_{00}-\beta_{0\cdot}\beta_{\cdot 0}+\beta_{0\cdot}\beta_{\cdot 0}-\widehat\beta_{0\cdot}\widehat\beta_{\cdot 0})^2\geq nh^2S_1,$$

where $S_1 = (\beta_{00} - \beta_{0\cdot}\beta_{\cdot 0} + \beta_{0\cdot}\beta_{\cdot 0})^2 + 2(\beta_{00} - \beta_{0\cdot}\beta_{\cdot 0})(\widehat{\beta}_{00} - \beta_{00} + \beta_{0\cdot}\beta_{\cdot 0} - \widehat{\beta}_{0\cdot}\widehat{\beta}_{\cdot 0}) \to C_1 > 0$. Therefore $\text{pr}(nh^2\widehat{T}_1 \le t) \le \text{pr}(nh^2 S_1 \le t) \to 0$ for any $t \in (0, \infty)$ since $nh^2 \to \infty$ by assumption. Next,

$$nh^2\widehat{T}_2 = \int \{\widehat{\beta}_{10}\widehat{f}(x) - \widehat{\beta}_{\cdot 0}(1 - \widehat{\beta}_{0\cdot})\widehat{g}(x)\}^2 dx \ge nh^2 S_2,$$

where $S_2 = \int \{\beta_{10} f(x) - \beta_{\cdot 0}(1 - \beta_{0\cdot})g(x)\}^2 dx + 2\int \{\beta_{10} f(x) - \beta_{\cdot 0}(1 - \beta_{0\cdot})g(x)\}\{\widehat{\beta}_{10}\widehat{f}(x) - \beta_{10} f(x) + \beta_{\cdot 0}(1 - \beta_{0\cdot})g(x) - \widehat{\beta}_{\cdot 0}(1 - \widehat{\beta}_{0\cdot})\widehat{g}(x)\}dx \to C_2 > 0$. Therefore $\text{pr}(nh^2\widehat{T}_2 \le t) \le \text{pr}(nh^2 S_2 \le t) \to 0$ for any $t \in (0, \infty)$. Similarly, one can show that $\text{pr}(nh^2\widehat{T}_3 \le t) \to 0$ and $\text{pr}(nh^2\widehat{T}_4 \le t) \to 0$ for any $t \in (0, \infty)$.

Finally,

$$\begin{aligned}
\text{pr}(nh^2\widehat{T} > t_\alpha) &= 1 - \text{pr}(nh^2\widehat{T}_1 + nh^2\widehat{T}_2 + nh^2\widehat{T}_3 + nh^2\widehat{T}_4 \le t_\alpha) \\
&\ge 1 - \min\{\text{pr}(nh^2\widehat{T}_1 \le t_\alpha), \text{pr}(nh^2\widehat{T}_2 \le t_\alpha), \text{pr}(nh^2\widehat{T}_3 \le t_\alpha), \text{pr}(nh^2\widehat{T}_4 \le t_\alpha)\} \\
&\to 1
\end{aligned}$$

as $n \to \infty$. □

S.2 Simulation designs

The following formulae are used to generate simulated samples in Section 4.1. SN: standard null, MN: mixture null, ME: mutually exclusive, TR: triangle, SC: semicircle and DN: donut in Figure 2.

SN $(\beta_{00}, \beta_{10}, \beta_{01}, \beta_{11}) = (0.09, 0.21, 0.21, 0.49)$, $f(x) = f(y) = \text{N}(\cdot|0.4, 0.1^2)$, $f(x, y) = f(x)f(y)$,

MN $(\beta_{00}, \beta_{10}, \beta_{01}, \beta_{11}) = (0.09, 0.21, 0.21, 0.49)$,
$f(x) = f(y) = 0.5\text{N}(\cdot|0.4, 0.1^2) + 0.5\text{N}(\cdot|0.7, 0.1^2), f(x, y) = f(x)f(y)$,

ME $(\beta_{00}, \beta_{10}, \beta_{01}, \beta_{11}) = (0.53, 0.25, 0.13, 0.09)$,
$f(x) = \text{N}(x|0.5, 0.15^2), f(y) = \text{G}(y|5.15, 22.7), f(x, y) = \text{DR}(x, y|12.5, 5.3, 12.2)$,

TR $(\beta_{00}, \beta_{10}, \beta_{01}, \beta_{11}) = (0.07, 0.53, 0.04, 0.36)$, $f(x) = \text{U}(x|0.05, 0.7), f(y) = \text{G}(y|1.71, 16.7)$,
$f(x, y) = 0.56\text{N}\Big(\begin{bmatrix} x \\ y \end{bmatrix} \Big| \begin{bmatrix} 0.518 \\ 0.170 \end{bmatrix}, \begin{bmatrix} 0.0149 & -0.013 \\ -0.013 & 0.0149 \end{bmatrix} \Big) + 0.44\text{N}\Big(\begin{bmatrix} x \\ y \end{bmatrix} \Big| \begin{bmatrix} 0.219 \\ 0.095 \end{bmatrix}, \begin{bmatrix} 0.017 & -0.003 \\ -0.003 & 0.0037 \end{bmatrix} \Big)$,

SC $(\beta_{00}, \beta_{10}, \beta_{01}, \beta_{11}) = (0.25, 0.08, 0.25, 0.42),\ f(x) = \mathrm{U}(x|0.05, 0.25), f(y) = \mathrm{N}(y|0.5, 0.15^2),$

$$f(x,y) = \mathrm{U}(x, y|(x - 0.001)^2/0.248 + (y - 0.405)^2/0.624 \leq 0.25, y > 0),$$

DN $(\beta_{00}, \beta_{10}, \beta_{01}, \beta_{11}) = (0.2, 0.2, 0.2, 0.4),$

$$f(x) = f(y) = 0.8\mathrm{U}(\cdot|0.05, 0.95) + 0.1\mathrm{U}(\cdot|0.15, 0.25) + 0.1\mathrm{U}(\cdot|0.75, 0.85),$$

$$f(x,y) = \mathrm{U}(x, y|0.25^2 \leq (x - 0.5)^2 + (y - 0.5)^2 \leq 0.45^2).$$

Here $\mathrm{N}(\cdot|\mu, \sigma^2)$ denotes the pdf of a normal distribution with mean μ and variance σ^2, $\mathrm{G}(\cdot|\alpha, \beta)$ stands for the pdf of a gamma distribution with shape parameter α and rate parameter β, $\mathrm{DR}(x, y|a_1, a_2, a_3)$ is the pdf of a Dirichlet distribution with parameters (a_1, a_2, a_3), $\mathrm{U}(\cdot|a, b)$ denotes the pdf of an uniform distribution between a and b and $\mathrm{U}(x, y|S)$ is the pdf of the bivariate uniform distribution on a set S.

Note that for the standard and mixture null designs, $g(x) = f(x)$ and $g(y) = f(y)$. However, $g(x) \neq f(x)$ and $g(y) \neq f(y)$ for other designs.

S.3 Normal approximation of the test statistic and bootstarp procedure

On the basis of the standard null design in Section 4 we explored the numerical accuracy of normal distribution approximations in Theorem 1. In addition to the original set up, we increased zero proportions up to $\beta_{0\cdot} = \beta_{\cdot 0} = 0.45$. For this particular simulation, we fixed $h_1 = 0.1n^{-1/5}$ where 0.1 is true standard deviation of non-zero part to remove the effect of bandwidth selection and it satisfies $nh_1^2 \to \infty$.

The entries in Table S.1 are proportions of rejected H_0 at asymptotic 5% level of test defined in (5). To compute the critical value $t_{0.05}$ in (5), we used true values for proportions and $\int g^2(x)dx$. We found that the nominal level for $\widehat{T}_1$ is well-maintained regardless of sample size because it is based on the basic central limit theorem. Although the empirical levels for $\widehat{T}_2, \widehat{T}_3, \widehat{T}_4$ and $\widehat{T}$ increase to the nominal value as sample sizes grow, rejection proportions are still below than 5% even for very large sample size. Comparison between Tables 1 and S.1 clearly demonstrate the benefit of using bootstrap approach to estimate the sampling distribution of the test statistics under H_0.

For bootstrap procedure, first we note that $\boldsymbol{X}$ and $\boldsymbol{Y}$ can be separately generated under the null hypothesis of independence (Wu et al., 2009; Henderson and Parmeter, 2015). Let $\boldsymbol{X}^* = (X_1^*, \ldots, X_n^*)$ and $\boldsymbol{Y}^* = (Y_1^*, \ldots, Y_n^*)$ be two bootstrap sample generated from $\widehat{F}_X$ and $\widehat{F}_Y$, respectively, where $\widehat{F}_X$ is the empirical CDF of marginal distribution for $\boldsymbol{X}$ and $\widehat{F}_Y$ is for $\boldsymbol{Y}$. We then compute test statistic for each bootstrap sample and the estimated p-value is calculated as $\{1+\sum_{b=1}^{B} I(\widehat{T}_b^* > \widehat{T})\}/(1+B)$ where $\widehat{T}_b^*$ is the test statistic for bth bootstrap sample. If the estimated p-value is less than a significance level α, we reject the null hypothesis at the $100\alpha\%$ level.

S.4 The test results for the (2, 17b) pairs in colorectal adenocarcinoma

In Table S.3 for colorectal cancer, Kendall's tau test is much more significant than the SKIT test especially for the (2, 17b) pair. It was pointed out by a referee that Kendall's test should not be that significant since the zero proportions of these two signatures are 94% and 92% respectively so that there are so many ties. In this section, we investigated that pair in detail.

First, we checked the calculation of p-value of Kendall's tau test. Due to a large number of ties (88% of points are both zeros), normal approximation for Kendall's tau test may not be proper; instead, we could calculate p-value of Kendall's tau using bootstrapping. With 1 million bootstrap replications, we obtained bootstrap Kendall's tau p-value 0.00024 (16 times larger than the Kendall's p-value 0.000015 based on the normal approximation). Second, we investigated why bootstrap Kendall's p-value is still smaller than SKIT's p-value (0.00084). Kendall's tau test is based on discordant and concordant pairs. We found that due to extremely high proportions at the origin (88%) for the (2, 17b) pair in colorectal cancer there exist a large number of concordant pairs (3659) but a small number of discordant pairs (808), which leads to the Kendall score Z statistic 4.3. Third, we examined if p-value of Kendall's tau is smaller by chance. Since the pattern of the (2, 17b) pair is similar to a mutually exclusive design, we carried out additional simulations to mimic the pattern of the (2, 17b) pair. We simulated a mutually exclusive design with proportions $(88\%, 6\%, 4\%, 2\%)$ and $n = 500$, and calculated p-values based on 1 million bootstrap replications. The overall powers based on bootstrap are similar, 0.9446 and 0.9332 for SKIT and the Kendall test respectively.

Interestingly, Figure S.4 showed bootstrap p-values between SKIT and the Kendall test, where SKIT has smaller p-value in about half of simulations. The red dot in the scatterplot represents p-values from the (2, 17b) pair in colorectal cancer, implying that it possibly by chance locates in the case when bootstrap Kendall's p-value is smaller. Finally, when the proportion of zeros is extremely high (e.g., $\sim$90% in this case), considering non-parametric pattern by SKIT or concordance by Kendall test may not help much. Indeed, for this case, if we just carry out a Chi-squared test for zero vs. non-zeros for the pair, the p-value would be 0.0003. As demonstrated in Section 4.3, the proposed method is more resilient to power loss due to a larger proportion of zeros than other methods. The resilience is most clear around 50% of zeros or less. However, when zeros are extremely prevalent, the proposed method SKIT would perform similarly to other methods.

S.5 Supplementary Tables and Figures

Table S.1: The entries are proportions of rejected H_0 at asymptotic 5% level of test defined in (5) and corresponding components. $5,000$ simulation samples are generated under design SN.

	$\beta_{0\cdot} = \beta_{\cdot 0} = 0.3$					$\beta_{0\cdot} = \beta_{\cdot 0} = 0.45$				
n	$\widehat{T}$	$\widehat{T}_1$	$\widehat{T}_2$	$\widehat{T}_3$	$\widehat{T}_4$	$\widehat{T}$	$\widehat{T}_1$	$\widehat{T}_2$	$\widehat{T}_3$	$\widehat{T}_4$
500	0.0190	0.0492	0.0224	0.0188	0.0186	0.0262	0.0442	0.0270	0.0308	0.0210
10000	0.0330	0.0514	0.0346	0.0372	0.0320	0.0382	0.0530	0.0354	0.0386	0.0360
50000	0.0366	0.0520	0.0394	0.0424	0.0346	0.0398	0.0574	0.0424	0.0446	0.0372
100000	0.0376	0.0472	0.0452	0.0442	0.0364	0.0434	0.0488	0.0424	0.0402	0.0424
200000	0.0383	0.0487	0.0389	0.0383	0.0389	0.0376	0.0544	0.0480	0.0410	0.0364
240000	0.0429	0.0502	0.0448	0.0439	0.0429	0.0402	0.0430	0.0444	0.0482	0.0368

Table S.2: Complementary table for Table 1 in the main article. The entries present rejection proportions of SKIT statistics and existing tests under H_0 for various designs (described in Figure 2). The simulation study is based on $5,000$ replications at the 5% level. For each simulated data, we estimated p-values for all components of SKIT ($\widehat{T}$, $\widehat{T}_k$ for $k = 1, \ldots, 4$) with $B = 3,000$ bootstrap samples. Other existing independence tests include Pearson correlation test (P), Kendall's τ (K) and Distance covariance (V). The entries corresponding to the standard null and mixture designs show the Type-I error rate and the other designs show powers.

		SKIT					Other Existing Tests		
	n	$\widehat{T}$	$\widehat{T}_1$	$\widehat{T}_2$	$\widehat{T}_3$	$\widehat{T}_4$	P	K	V
Standard	50	0.032	0.060	0.045	0.050	0.028	0.049	0.044	0.054
Null	300	0.047	0.043	0.051	0.050	0.047	0.044	0.045	0.044
	500	0.046	0.051	0.047	0.044	0.047	0.046	0.044	0.046
	1000	0.048	0.044	0.046	0.052	0.048	0.049	0.050	0.048
	2000	0.050	0.050	0.050	0.049	0.050	0.051	0.053	0.053
Mixture	50	0.042	0.060	0.048	0.055	0.038	0.057	0.057	0.058
Null	300	0.046	0.043	0.054	0.048	0.047	0.052	0.053	0.048
	500	0.047	0.052	0.044	0.051	0.050	0.046	0.044	0.050
	1000	0.052	0.046	0.049	0.049	0.051	0.048	0.050	0.049
	2000	0.051	0.052	0.054	0.045	0.052	0.054	0.054	0.053
Mutually	50	0.158	0.102	0.200	0.124	0.149	0.016	0.039	0.061
Exclusiveness	100	0.392	0.123	0.399	0.215	0.396	0.022	0.042	0.072
	200	0.754	0.205	0.715	0.406	0.768	0.035	0.039	0.127
	300	0.929	0.276	0.892	0.581	0.936	0.049	0.040	0.228
Triangle	50	0.103	0.041	0.057	0.037	0.122	0.019	0.032	0.059
	100	0.328	0.053	0.082	0.055	0.417	0.022	0.040	0.123
	200	0.830	0.064	0.119	0.087	0.896	0.031	0.050	0.298
	300	0.985	0.074	0.163	0.131	0.995	0.045	0.068	0.517
Semicircle	50	0.559	0.724	0.702	0.413	0.365	0.059	0.151	0.144
	100	0.928	0.955	0.957	0.788	0.804	0.085	0.272	0.298
	200	0.999	0.999	0.999	0.991	0.995	0.145	0.489	0.639
	300	1.000	1.000	1.000	1.000	1.000	0.192	0.659	0.873
Donut	50	0.261	0.223	0.114	0.117	0.253	0.098	0.137	0.135
	100	0.680	0.385	0.188	0.189	0.754	0.162	0.246	0.238
	200	0.997	0.656	0.366	0.370	1.000	0.289	0.437	0.435
	300	1.000	0.831	0.543	0.543	1.000	0.403	0.605	0.625

Table S.3: Complementary table for Table 2 in the main article. The entries are FDR adjusted p-values of all active pairs between 17a/b and APOBEC-mediated signatures for gastrointestinal cancers. P: Pearson correlation test, K: Kendall's τ and V: Distance covariance. $\widehat{T}_k$: k^{th} component of SKIT, $k = 1, \ldots, 4$.

		SKIT					Other Existing Tests		
Cancer Type	Pairs	$\widehat{T}$	$\widehat{T}_1$	$\widehat{T}_2$	$\widehat{T}_3$	$\widehat{T}_4$	P	K	V
Colorectal	(2, 17b)	0.0019	0.0014	0.0031	0.0009	0.0034	0.3081	0.0001	0.0317
	(13, 17b)	0.0320	0.0356	0.0179	0.0561	0.0387	0.3013	0.0131	0.0559
Esophageal	(2, 17a)	0.0351	0.1310	0.0239	0.1372	0.0486	0.0008	0.0012	0.0005
	(2, 17b)	0.0336	0.1310	0.0377	0.0694	0.0486	0.0024	0.0018	0.0010
	(13, 17a)	0.0081	0.1310	0.0037	0.1372	0.0104	0.0007	0.0003	0.0003
	(13, 17b)	0.0098	0.1310	0.0063	0.0694	0.0176	0.0008	0.0003	0.0003
Stomach	(2, 17a)	0.0354	0.3102	0.0554	0.1457	0.0394	0.3086	0.6069	0.2525
	(2, 17b)	0.0085	0.1804	0.0132	0.0735	0.0097	0.3367	0.5181	0.1745
	(13, 17a)	0.2112	0.3480	0.1839	0.3071	0.2269	0.2182	0.7058	0.3414
	(13, 17b)	0.0423	0.1892	0.1084	0.1605	0.0414	0.3367	0.4754	0.4320

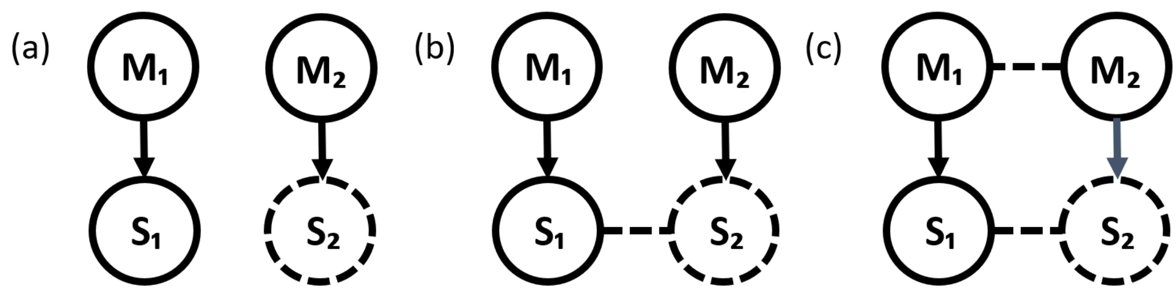

Figure S.1: The schematic outline of the relationship between two signatures described in Section 2.3 where for two signatures S_1 and S_2 the mutational process M_1 is known to cause S_1 but unobserved (represented by an arrow in (a)) and the causal mutational process (M_2) for S_2 is unknown. Since mutational signature S_1 is caused by an mutational process M_1 and itself cannot cause other mutational signatures such as S_2 (blank between S_1 to S_2 in (a)), the null hypothesize would be that S_1 and S_2 are independent if M_1 is not associated with M_2 (blank between M_1 to M_2 in (a)). When the dependence of S_1 and S_2 is observed (represented by a dotted line in (b)), we reject the null hypothesis and conclude M_2 are associated with M_1 as well (represented by a dotted line between M_1 and M_2 in (c)).

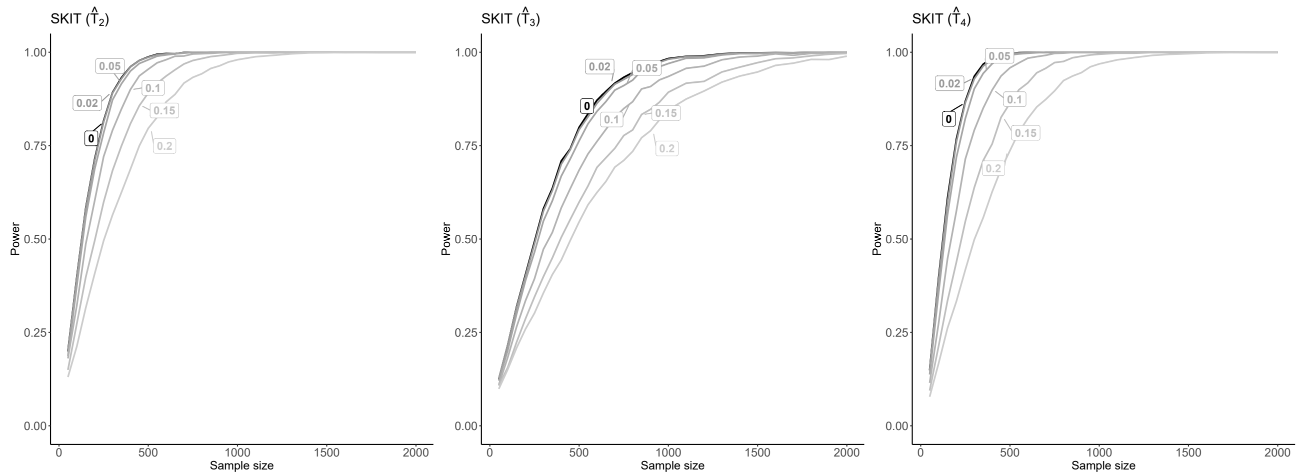

Figure S.2: Simulation results for robustness in Section 4.2 based on 5000 replications for each n and ε. The lighter the color, the larger the ε. The left panel shows the power based on SKIT while the right panel presents the power for distance covariance test.

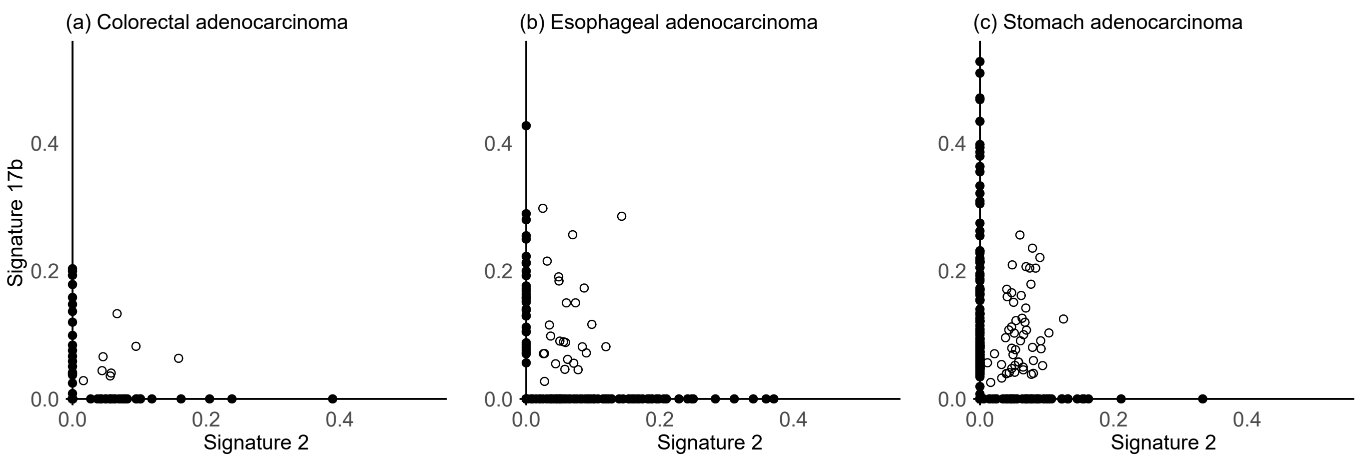

Figure S.3: Scatterplots between signature 2 and signature 17b for (a) colorectal adenocarcinoma, (b) esophageal adenocarcinoma and (c) stomach adenocarcinoma. Solid dots contain at least one zero while circle does not contain any zero.

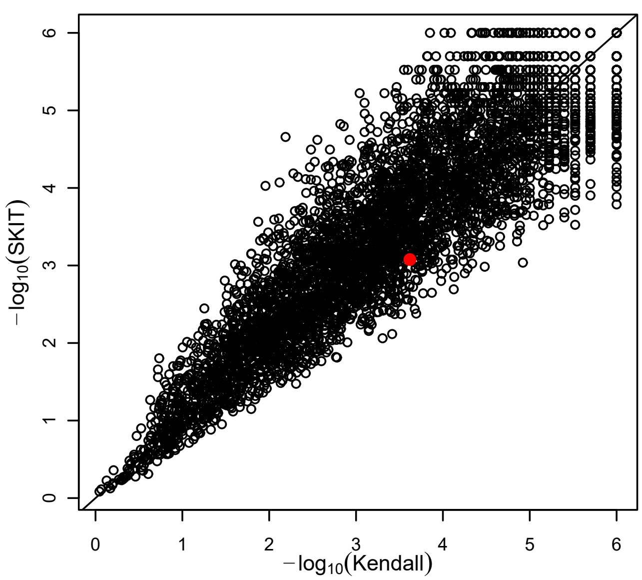

Figure S.4: A scatterplot between bootstrap p-values between SKIT and Kendall test for mutually exclusiveness design with proportions $(88\%, 6\%, 4\%, 2\%)$ and $n = 500$ based on 1 million bootstrap replicates. The red dot represents p-values from the (2, 17b) pair in colorectal cancer.

References

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis 14* (1), 1–16.

Henderson, D. J. and C. F. Parmeter (2015). *Applied Nonparametric Econometrics*. Cambridge University Press.

Rosenblatt, M. and B. E. Wahlen (1992). A nonparametric measure of independence under a hypothesis of independent components. *Statistics and Probability Letters 15* (3), 245–252.

Wahlen, B. E. (1992). *A Nonparametric Measure of Independence*. Ph. D. thesis, University of California, San Diego.

Wu, E. H., L. Philip, and W. K. Li (2009). A smoothed bootstrap test for independence based on mutual information. *Computational Statistics and Data Analysis 53* (7), 2524–2536.