A NEW NONPARAMETRIC EXTENSION OF ANOVA VIA PROJECTION MEAN VARIANCE MEASURE

Jicai Liu[1,2],Yuefeng Si[3], Wenchao Xu[1] and Riquan Zhang[4]

1. Shanghai Lixin University of Accounting and Finance

2. KLATASDS-MOE, East China Normal University

3. University of Hong Kong

4. East China Normal University

Supplementary Material

S1 Proofs of the theoretical results

Proof of Lemma 1. The " $\Rightarrow$ " part is immediate by the elementary properties of independence.

Next we prove the reverse. For any $\mathbf{t}(\neq 0) \in \mathcal{R}^p$ and any $s \in \mathcal{R}$, the characteristic function of $(\mathbf{X}, Y)$ satisfies

$$\begin{aligned}\phi_{(\mathbf{X},Y)}(\mathbf{t}, s) &= E[\exp\{i\mathbf{X}^T\mathbf{t} + isY\}] \\ &= E[\exp\{i\|\mathbf{t}\|\mathbf{X}^T(\mathbf{t}/\|\mathbf{t}\|) + isY\}] \\ &= \phi_{(\tilde{\beta}^T\mathbf{X},Y)}(\|\mathbf{t}\|, s),\end{aligned}$$

where $\tilde{\beta} = \mathbf{t}/\|\mathbf{t}\| \in \mathbb{S}^{p-1}$. Thus, if $\beta^T\mathbf{X} \perp\!\!\!\perp Y$ holds for any $\beta \in \mathbb{S}^{p-1}$, we obtain that

$$\phi_{(\mathbf{X},Y)}(\mathbf{t}, s) = \phi_{\tilde{\beta}^T\mathbf{X}}(\|\mathbf{t}\|)\phi_Y(s) = E[\exp\{i\mathbf{X}^T\mathbf{t}\}]E[\exp\{isY\}] = \phi_{\mathbf{X}}(\mathbf{t})\phi_Y(s).$$

In addition, when $\mathbf{t} = \mathbf{0}$, it is easy to see that $\phi_{(\mathbf{X},Y)}(\mathbf{t}, s) = \phi_{\mathbf{X}}(\mathbf{t})\phi_Y(s)$ also holds. Thus, if $\beta^T\mathbf{X} \perp\!\!\!\perp Y$ holds for any $\beta \in \mathbb{S}^{p-1}$, we have that $\mathbf{X} \perp\!\!\!\perp Y$.

□

Proof of Theorem 1. (i) Note that

$$\begin{aligned}E_Y[F_{\beta^T\mathbf{X}}(u|Y)] &= \sum_{k=1}^{K} \mathrm{pr}\{\beta^T\mathbf{X} \le u|Y = y_k\}\,\mathrm{pr}\{Y = y_k\} \\ &= \sum_{k=1}^{K} \mathrm{pr}\{\beta^T\mathbf{X} \le u, Y = y_k\} = F_{\beta^T\mathbf{X}}(u)\end{aligned}$$

and

$$\mathrm{var}_Y[F_{\beta^T\mathbf{X}}(u|Y)] = \sum_{k=1}^{K} [\mathrm{pr}\{\beta^T\mathbf{X} \le u|Y = y_k\} - F_{\beta^T\mathbf{X}}(u)]^2\,\mathrm{pr}\{Y = y_k\}.$$

Thus, we obtain that

$$\begin{aligned}\mathrm{PMV}(\mathbf{X}|Y) &= \frac{1}{c_p}\int_{\mathbb{S}^{p-1}} E_{\beta^T\mathbf{X}}[\mathrm{var}_Y(F(\beta^T\mathbf{X}|Y))]d\beta \\ &= \frac{1}{c_p}\sum_{k=1}^{K} p_k \int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}[\mathrm{pr}\{\beta^T\mathbf{X}\le u|Y=y_k\} - F_{\beta^T\mathbf{X}}(u)]^2 dF_{\beta^T\mathbf{X}}(u)d\beta.\end{aligned}$$

(ii) By the property (i), we have that $\mathrm{PMV}(\mathbf{X}|Y) = 0$ if and only if, for any $\beta \in \mathbb{S}^{p-1}, u \in \mathcal{R}$ and $k = 1, \cdots, K$,

$$\mathrm{pr}\{\beta^T\mathbf{X} \le u|Y = y_k\} = F_{\beta^T\mathbf{X}}(u), a.s.,$$

which is equivalent to $\beta^T\mathbf{X} \perp\!\!\!\perp Y$ for any $\beta \in \mathbb{S}^{p-1}$. Thus, by (2.3), $\mathrm{PMV}(\mathbf{X}|Y) = 0$ if and only if $\mathbf{X} \perp\!\!\!\perp Y$.

(iii) Let $U = \beta^T\mathbf{X}$ and $U_i = \beta^T\mathbf{X}_i$. After some algebra, we can obtain that

$$\begin{aligned}&\int_{-\infty}^{\infty}[\mathrm{pr}\{\beta^T\mathbf{X}\le u|Y=y_k\} - \mathrm{pr}\{\beta^T\mathbf{X}\le u\}]^2 dF_{\beta^T\mathbf{X}}(u) \\ =\ & p_k^{-2}E[I(U_1\le U_3)I(U_2\le U_3)I(Y_1=y_k,Y_2=y_k)] + E[I(U_1\le U_3)I(U_2\le U_3)] \\ & -2p_k^{-1}E[I(U_1\le U_3)I(U_2\le U_3)I(Y_1=y_k)].\end{aligned}$$

This, together with (2.5), yields that

$$\begin{aligned}
&\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}[\mathrm{pr}\{\beta^T\mathbf{X}\le u|Y=y_k\}-\mathrm{pr}\{\beta^T\mathbf{X}\le u\}]^2dF_{\beta^T\mathbf{X}}(u)d\beta\\
=\ & p_k^{-2}E\Big[I(Y_1=y_k,Y_2=y_k)\int_{\mathbb{S}^{p-1}}I(\beta^T\mathbf{X}_1\le\beta^T\mathbf{X}_3)I(\beta^T\mathbf{X}_2\le\beta^T\mathbf{X}_3)d\beta\Big]\\
&+E\Big[\int_{\mathbb{S}^{p-1}}I(\beta^T\mathbf{X}_1\le\beta^T\mathbf{X}_3)I(\beta^T\mathbf{X}_2\le\beta^T\mathbf{X}_3)d\beta\Big]\\
&-2p_k^{-1}E\Big[I(Y_1=y_k)\int_{\mathbb{S}^{p-1}}I(\beta^T\mathbf{X}_1\le\beta^T\mathbf{X}_3)I(\beta^T\mathbf{X}_2\le\beta^T\mathbf{X}_3)d\beta\Big]\\
=\ & p_k^{-2}E[I(Y_1=y_k,Y_2=y_k)c_p[\pi-\mathrm{ang}(\mathbf{X}_1-\mathbf{X}_3,\mathbf{X}_2-\mathbf{X}_3)]]\\
&+E[c_p[\pi-\mathrm{ang}(\mathbf{X}_1-\mathbf{X}_3,\mathbf{X}_2-\mathbf{X}_3)]]\\
&-2p_k^{-1}E[I(Y_1=y_k)c_p[\pi-\mathrm{ang}(\mathbf{X}_1-\mathbf{X}_3,\mathbf{X}_2-\mathbf{X}_3)]]\\
=\ & -c_pp_k^{-2}E[I(Y_1=y_k,Y_2=y_k)\mathrm{ang}(\mathbf{X}_1-\mathbf{X}_3,\mathbf{X}_2-\mathbf{X}_3)]\\
&-c_pE[\mathrm{ang}(\mathbf{X}_1-\mathbf{X}_3,\mathbf{X}_2-\mathbf{X}_3)]\\
&+2c_pp_k^{-1}E[I(Y_1=y_k)\mathrm{ang}(\mathbf{X}_1-\mathbf{X}_3,\mathbf{X}_2-\mathbf{X}_3)].
\end{aligned}$$

Thus, we obtain that

$$\begin{aligned}
\mathrm{PMV}(\mathbf{X}|Y)\ &=\ \frac{1}{c_p}\sum_{k=1}^{K}p_k\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}[\mathrm{pr}\{\beta^T\mathbf{X}\le u|Y=y_k\}-F_{\beta^T\mathbf{X}}(u)]^2dF_{\beta^T\mathbf{X}}(u)d\beta\\
&=\ -\sum_{k=1}^{K}p_k^{-1}E[I(Y_1=y_k,Y_2=y_k)\mathrm{ang}(\mathbf{X}_1-\mathbf{X}_3,\mathbf{X}_2-\mathbf{X}_3)]\\
&\quad\ +E[\mathrm{ang}(\mathbf{X}_1-\mathbf{X}_3,\mathbf{X}_2-\mathbf{X}_3)].
\end{aligned}$$

(iv) Denote $\mathbf{U}=\mathbf{a}+c\mathbf{A}\mathbf{X}$ and $\mathbf{U}_i=\mathbf{a}+c\mathbf{A}\mathbf{X}_i$. By the definition of

$\text{ang}(\cdot,\cdot)$, we have that

$$\begin{aligned}\text{ang}(\mathbf{U}_1-\mathbf{U}_3,\mathbf{U}_2-\mathbf{U}_3) &= \text{ang}(c\mathbf{A}(\mathbf{X}_1-\mathbf{X}_3), c\mathbf{A}(\mathbf{X}_2-\mathbf{X}_3)) \\ &= \arccos\left\{\frac{(\mathbf{X}_1-\mathbf{X}_3)^T c^2\mathbf{A}^T\mathbf{A}(\mathbf{X}_2-\mathbf{X}_3)}{\|c\mathbf{A}(\mathbf{X}_1-\mathbf{X}_3)\|\|c\mathbf{A}(\mathbf{X}_2-\mathbf{X}_3)\|}\right\} \\ &= \text{ang}(\mathbf{X}_1-\mathbf{X}_3,\mathbf{X}_2-\mathbf{X}_3),\end{aligned}$$

where $\mathbf{A}\in\mathcal{R}^{p\times p}$ is any orthonormal matrix, $\mathbf{a}\in\mathcal{R}^p$ and $c\in\mathcal{R}$. This, together with the property (iii), yields that $\text{PMV}(\mathbf{a}+c\mathbf{A}\mathbf{X}|Y)=\text{PMV}(\mathbf{X}|Y)$.

□

Proof of Corollary 1. For any $v_1, v_2\in\mathcal{R}$, we have the following result

$$\text{ang}(v_1,v_2)=\arccos\left\{\frac{v_1^T v_2}{\|v_1\|\|v_2\|}\right\}=\pi[I(v_1<0)I(v_2>0)+I(v_1>0)I(v_2<0)].$$

Thus, when $\mathbf{X}$ is univariate, we obtain that

$$\text{ang}(X_1-X_3,X_2-X_3)=\pi[I(X_1<X_3)I(X_2>X_3)+I(X_1>X_3)I(X_2<X_3)].$$

This, together with Proposition 4 and the property (iii) of Theorem 1, gives that

$$
\begin{aligned}
\mathrm{PMV}(X|Y) &= E[\mathrm{ang}(X_1 - X_3, X_2 - X_3)] \\
&\quad - \sum_{k=1}^{K} p_k^{-1} E[I(Y_1 = y_k, Y_2 = y_k)\mathrm{ang}(X_1 - X_3, X_2 - X_3)] \\
&= \pi E[I(X_1 < X_3)I(X_2 > X_3)] + \pi E[I(X_1 > X_3)I(X_2 < X_3)] \\
&\quad - \pi \sum_{k=1}^{K} p_k^{-1} E[I(Y_1 = y_k, Y_2 = y_k)I(X_1 < X_3)I(X_2 > X_3)] \\
&\quad - \pi \sum_{k=1}^{K} p_k^{-1} E[I(Y_1 = y_k, Y_2 = y_k)I(X_1 > X_3)I(X_2 < X_3)] \\
&= 2\pi \mathrm{MV}(X|Y).
\end{aligned}
$$

□

Proof of Theorem 2. Note that

$$
g_{U,Y}^n(u; y_k) - g_U^n(u) = \hat{p}_k^{-1} n^{-1} \sum_{i=1}^{n} I(U_i \le u, Y_i = y_k) - n^{-1} \sum_{i=1}^{n} I(U_i \le u).
$$

Thus, we have that

$$
\begin{aligned}
&[g_{U,Y}^n(u; y_k) - g_U^n(u)]^2 \\
= &\frac{1}{n^2} \sum_{i,j=1}^{n} \{\hat{p}_k^{-1} I(U_i \le u, Y_i = y_k) - I(U_i \le u)\}\{\hat{p}_k^{-1} I(U_j \le u, Y_j = y_k) - I(U_j \le u)\}.
\end{aligned}
$$

It follows from (2.5) that we obtain

$$\begin{aligned}
&\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y)\\
=\;& \frac{1}{nc_p}\sum_{k=1}^{K}\hat{p}_k\sum_{i=1}^{n}\int_{\mathbb{S}^{p-1}}[g^n_{U,Y}(\beta^T\mathbf{X}_i;y_k)-g^n_U(\beta^T\mathbf{X}_i)]^2d\beta\\
=\;& \frac{1}{n^3c_p}\sum_{k=1}^{K}\hat{p}_k\sum_{i,j,r=1}^{n}\int_{\mathbb{S}^{p-1}}\Big[\hat{p}_k^{-1}I(\beta^T\mathbf{X}_j\le\beta^T\mathbf{X}_i,Y_j=y_k)-I(\beta^T\mathbf{X}_j\le\beta^T\mathbf{X}_i)\Big]\\
&\times\Big[\hat{p}_k^{-1}I(\beta^T\mathbf{X}_r\le\beta^T\mathbf{X}_i,Y_r=y_k)-I(\beta^T\mathbf{X}_r\le\beta^T\mathbf{X}_i)\Big]d\beta\\
=\;& -\frac{1}{n^3}\sum_{k=1}^{K}\hat{p}_k\sum_{i,j,r=1}^{n}\{\hat{p}_k^{-2}a_{jri}I(Y_j=y_k,Y_r=y_k)-\hat{p}_k^{-1}a_{jri}I(Y_r=y_k)\\
&-\hat{p}_k^{-1}a_{jri}I(Y_j=y_k)+a_{jri}\} \qquad \text{(S1.1)}\\
=\;& -\frac{1}{n^3}\sum_{k=1}^{K}\hat{p}_k^{-1}\sum_{i,j,r=1}^{n}a_{jri}\{b_{jr;k}-\bar{b}_{\cdot r;k}-\bar{b}_{j\cdot;k}+\bar{b}_{\cdot\cdot;k}\}, \qquad \text{(S1.2)}
\end{aligned}$$

where $\bar{b}_{\cdot j;k}=n^{-1}\sum_{i=1}^{n}b_{ij;k}$, $\bar{b}_{i\cdot;k}=n^{-1}\sum_{j=1}^{n}b_{ij;k}$ and $\bar{b}_{\cdot\cdot;k}=n^{-2}\sum_{i,j=1}^{n}b_{ij;k}$.

Using (S1.2), we have that

$$\begin{aligned}
\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y) &= -\frac{1}{n^3}\sum_{k=1}^{K}\hat{p}_k^{-1}\sum_{i,j,r=1}^{n}a_{jri}\{b_{jr;k}-\bar{b}_{\cdot r;k}-\bar{b}_{j\cdot;k}+\bar{b}_{\cdot\cdot;k}\}\\
&= -\frac{1}{n^3}\sum_{k=1}^{K}\hat{p}_k^{-1}\sum_{i,j,r=1}^{n}a_{jri}\widetilde{B}_{jr;k}\\
&= -\frac{1}{n^3}\sum_{k=1}^{K}\hat{p}_k^{-1}\sum_{i,j,r=1}^{n}\widetilde{A}_{jr;i}\widetilde{B}_{jr;k}.
\end{aligned}$$

Thus, we obtain the result in (2.7).

It follows from (S1.1) that

$$\begin{aligned}\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y) &= -\frac{1}{n^3}\sum_{k=1}^{K}\hat{p}_k\sum_{i,j,r=1}^{n}\{\hat{p}_k^{-2}a_{jri}I(Y_j=Y_r=y_k)-2\hat{p}_k^{-1}a_{jri}I(Y_r=y_k)+a_{jri}\}\\ &= -\frac{1}{n^3}\sum_{k=1}^{K}\hat{p}_k^{-1}\sum_{i,j,r=1}^{n}a_{jri}I(Y_j=y_k,Y_r=y_k)+\frac{1}{n^3}\sum_{i,j,r=1}^{n}a_{ijr}.\end{aligned}$$

This completes the proof of (2.8).

□

Proof of Theorem 3. It follows from Theorems 1 and 2 that

$$\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y)=S_{1n}-S_{2n},\quad \mathrm{PMV}(\mathbf{X}|Y)=S_1-S_2,$$

where

$$S_{1n}=\frac{1}{n^3}\sum_{i,j,r=1}^{n}\mathrm{ang}(\mathbf{X}_i-\mathbf{X}_r,\mathbf{X}_j-\mathbf{X}_r),\ S_{2n}=\frac{1}{n^3}\sum_{k=1}^{K}\hat{p}_k^{-1}\sum_{i,j,r=1}^{n}I(Y_i=Y_j=y_k)\mathrm{ang}(\mathbf{X}_i-\mathbf{X}_r,\mathbf{X}_j-\mathbf{X}_r)$$

and

$$S_1=E[\mathrm{ang}(\mathbf{X}_1-\mathbf{X}_3,\mathbf{X}_2-\mathbf{X}_3)],\ S_2=\sum_{k=1}^{K}p_k^{-1}E[I(Y_1=Y_2=y_k)\mathrm{ang}(\mathbf{X}_1-\mathbf{X}_3,\mathbf{X}_2-\mathbf{X}_3)].$$

Thus, we have that

$$\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y)-\mathrm{PMV}(\mathbf{X}|Y)=[S_{1n}-S_1]-[S_{2n}-S_2].\qquad \text{(S1.3)}$$

We shall establish the convergence of each part by the theory of U-statistics.

Step 1: The convergence of S_{1n}. Note that S_{1n} can be expressed as follows:

$$
\begin{aligned}
S_{1n} &= \frac{1}{n^3}\Big\{\sum_{\substack{i\neq r, i\neq j,\\ r\neq j}} \mathrm{ang}(\mathbf{X}_i-\mathbf{X}_r,\mathbf{X}_j-\mathbf{X}_r) + \sum_{\substack{i\neq r, i\neq j,\\ r=j}} \mathrm{ang}(\mathbf{X}_i-\mathbf{X}_r,\mathbf{X}_j-\mathbf{X}_r) \\
&\quad + \sum_{\substack{i\neq r, i=j,\\ r\neq j}} \mathrm{ang}(\mathbf{X}_i-\mathbf{X}_r,\mathbf{X}_j-\mathbf{X}_r) + \sum_{\substack{i=r, i\neq j,\\ r\neq j}} \mathrm{ang}(\mathbf{X}_i-\mathbf{X}_r,\mathbf{X}_j-\mathbf{X}_r)\Big\} \\
&= \frac{1}{n^3}\sum_{\substack{i\neq r, i\neq j,\\ r\neq j}} \mathrm{ang}(\mathbf{X}_i-\mathbf{X}_r,\mathbf{X}_j-\mathbf{X}_r) + \frac{1}{n^3}\sum_{i\neq r, i=j} \mathrm{ang}(\mathbf{X}_i-\mathbf{X}_r,\mathbf{X}_j-\mathbf{X}_r) \\
&= \frac{(n-1)(n-2)}{n^2} T_{1n},
\end{aligned}
$$

where T_{1n} is a U-statistics, defined by

$$
\begin{aligned}
T_{1n} &= \binom{n}{3}^{-1} \sum_{i<r<l} \frac{1}{3}[\mathrm{ang}(\mathbf{X}_i-\mathbf{X}_r,\mathbf{X}_l-X_r)+\mathrm{ang}(\mathbf{X}_r-\mathbf{X}_i,\mathbf{X}_l-\mathbf{X}_i)+\mathrm{ang}(\mathbf{X}_i-\mathbf{X}_l,\mathbf{X}_r-\mathbf{X}_l)] \\
&:= \binom{n}{3}^{-1} \sum_{i<r<l} k_1(\mathbf{X}_i,\mathbf{X}_r,\mathbf{X}_l),
\end{aligned}
$$

with the symmetric kernel function $k_1(\mathbf{X}_i,\mathbf{X}_r,\mathbf{X}_l)$.

For any given $t>0$, we have that

$$
\begin{aligned}
\mathrm{pr}\Big\{|S_{1n}-S_1|\geq t\Big\} &\leq \mathrm{pr}\Big\{\frac{(n-1)(n-2)}{n^2}|T_{1n}-S_1|\geq \frac{t}{2}\Big\} + \mathrm{pr}\Big\{|\frac{3n-2}{n^2}S_1|\geq \frac{t}{2}\Big\} \\
&\leq \mathrm{pr}\Big\{|T_{1n}-S_1|\geq \frac{t}{2}\Big\} + \mathrm{pr}\Big\{|\frac{3n-2}{n^2}S_1|\geq \frac{t}{2}\Big\}. \qquad \text{(S1.4)}
\end{aligned}
$$

Note that the kernel function $k_1(\mathbf{X}_i,\mathbf{X}_r,\mathbf{X}_l)$ satisfies $0\leq |k_1(\mathbf{X}_i,\mathbf{X}_r,\mathbf{X}_l)|\leq \pi$. Thus, we have $|S_1|\leq\pi$ and

$$
\mathrm{pr}\Big\{|\frac{3n-2}{n^2}S_1|\geq t\Big\} = 0, \qquad \text{(S1.5)}
$$

for n large enough, i.e., $n \geq 3\pi/t$.

Note that $E[T_{1n}] = S_1$. Thus, we next study the concentration inequality $\mathrm{pr}\{|T_{1n} - S_1| \geq t\}$. This can be obtained by the decoupling trick in Serfling (1980, Section 5.1.6) and Chernoff method. Denote

$$\Omega\left(\mathbf{X}_1, \cdots, \mathbf{X}_n\right) = \frac{1}{m}\sum_{r=0}^{m-1} k_1(\mathbf{X}_{1+2r}, \mathbf{X}_{2+2r}, \mathbf{X}_{3+2r}),$$

where $m = \lfloor n/3 \rfloor$. As shown by Serfling (1980, Section 5.1.6), we obtain that

$$T_{1n} = \frac{1}{n!}\sum_{n!} \Omega\left(\mathbf{X}_{i_1}, \cdots, \mathbf{X}_{i_n}\right),$$

where $\sum_{n!}$ denote summation over all $n!$ permutations $(i_1, \cdots, i_n)$ of $(1, \cdots, n)$.

By the Markov's inequality, we can obtain that

$$\mathrm{pr}\left\{T_{1n} - S_1 \geq t\right\} \leq \exp\{-\lambda t\} E[\exp\{\lambda[T_{1n} - S_1]\}], \tag{S1.6}$$

for any $\lambda > 0$. It follows from the Jensen's inequality that

$$\begin{aligned}
E\left[\exp\{\lambda T_{1n}\}\right] &= E\Big[\exp\{\frac{\lambda}{n!}\sum_{n!} \Omega\left(\mathbf{X}_{i_1}, \cdots, \mathbf{X}_{i_n}\right)\}\Big] \\
&\leq \frac{1}{n!}\sum_{n!} E\Big[\exp\{\lambda \Omega\left(\mathbf{X}_{i_1}, \cdots, \mathbf{X}_{i_n}\right)\}\Big] \\
&= E\Big[\exp\Big\{\frac{\lambda}{m}\sum_{r=0}^{m-1} k_1(\mathbf{X}_{1+2r}, \mathbf{X}_{2+2r}, \mathbf{X}_{3+2r})\Big\}\Big] \\
&= E^m\Big[\exp\{\lambda k_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)/m\}\Big].
\end{aligned}$$

This, together with (S1.6), yields that

$$\text{pr}\Big\{T_{1n} - S_1 \geq t\Big\} \leq \exp\{-\lambda t\} E^m\Big[\exp\{\lambda[k_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) - S_1]/m\}\Big]. \tag{S1.7}$$

It follows from $0 \leq |k_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) - S_1| \leq 2\pi$ and Hoeffding's Lemma that we have

$$E\Big[\exp\{\lambda[k_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) - S_1]/m\}\Big] \leq \exp\Big\{\frac{\pi^2}{2m^2}\lambda^2\Big\}.$$

This, together with (S1.7), leads to

$$\text{pr}\Big\{T_{1n} - S_1 \geq t\Big\} \leq \exp\Big\{-\lambda t + \frac{\pi^2}{2m}\lambda^2\Big\}.$$

Minimizing the right hand side of the above expression in λ, we can obtain that the optimal choice of λ is $\lambda = mt/\pi^2$. Thus, we can obtain that

$$\text{pr}\Big\{T_{1n} - S_1 \geq t\Big\} \leq \exp\Big\{-\frac{mt^2}{2\pi^2}\Big\}.$$

Repeating this argument for $-[T_{1n} - S_1]$ instead of $T_{1n} - S_1$, we obtain the same bound for $\text{pr}\{-[T_{1n} - S_1] \geq t\}$. Thus, we have that

$$\text{pr}\Big\{|T_{1n} - S_1| \geq t\Big\} \leq 2\exp\Big\{-\frac{mt^2}{2\pi^2}\Big\}. \tag{S1.8}$$

Combining (S1.4), (S1.5) with (S1.8), we obtain that

$$\text{pr}\Big\{|S_{1n} - S_1| \geq t\Big\} \leq 2\exp\Big\{-\frac{nt^2}{24\pi^2}\Big\}, \tag{S1.9}$$

for n large enough.

Step 2: The convergence of S_{2n}. Note that S_{2n} and S_2 can be expressed as follows:

$$\begin{aligned} S_{2n} &= \frac{1}{n^3}\sum_{k=1}^{K}\hat{p}_k^{-1}\sum_{i,j,r=1}^{n} I(Y_i = Y_j = y_k)\text{ang}(\mathbf{X}_i - \mathbf{X}_r, \mathbf{X}_j - \mathbf{X}_r) := \sum_{k=1}^{K}\hat{p}_k^{-1}S_{2n,k}, \\ S_2 &= \sum_{k=1}^{K} p_k^{-1} E[I(Y_1 = Y_2 = y_k)\text{ang}(\mathbf{X}_1 - \mathbf{X}_3, \mathbf{X}_2 - \mathbf{X}_3)] := \sum_{k=1}^{K} p_k^{-1} S_{2,k}, \end{aligned}$$

where

$$S_{2n,k} = \frac{1}{n^3}\sum_{i,j,r=1}^{n} I(Y_i = y_k, Y_j = y_k)\text{ang}(\mathbf{X}_i - \mathbf{X}_r, \mathbf{X}_j - \mathbf{X}_r)$$

and

$$S_{2,k} = E[I(Y_1 = y_k, Y_2 = y_k)\text{ang}(\mathbf{X}_1 - \mathbf{X}_3, \mathbf{X}_2 - \mathbf{X}_3)].$$

Similar to the argument of S_{1n}, we can represent $S_{2n,k}$ as a U-statistics as follows

$$S_{2n,k} = \frac{1}{n^3}\sum_{\substack{i\neq r, i\neq l,\\ r\neq l}} \text{ang}(\mathbf{X}_i - \mathbf{X}_r, \mathbf{X}_l - \mathbf{X}_r) I(Y_i = Y_l = y_k) := \frac{(n-1)(n-2)}{n^2} T_{2n,k},$$

where $T_{2n,k}$ is a U-statistics, given by

$$\begin{aligned} T_{2n,k} &= \binom{n}{3}^{-1} \sum_{i<r<l} \frac{1}{3}[\text{ang}(\mathbf{X}_i - \mathbf{X}_r, \mathbf{X}_l - \mathbf{X}_r) I(Y_i = Y_l = y_k) \\ &\quad + \text{ang}(\mathbf{X}_r - \mathbf{X}_i, \mathbf{X}_l - \mathbf{X}_i) I(Y_r = Y_l = y_k) + \text{ang}(\mathbf{X}_i - \mathbf{X}_l, \mathbf{X}_r - \mathbf{X}_l) I(Y_i = Y_r = y_k)] \\ &:= \binom{n}{3}^{-1} \sum_{i<r<l} k_{2,k}(\mathbf{X}_i, \mathbf{X}_r, \mathbf{X}_l). \end{aligned}$$

By the same argument of S_{1n}, we have

$$\begin{aligned} \mathrm{pr}\Big\{|S_{2n,k}-S_{2,k}|\geq t\Big\} &\leq \mathrm{pr}\Big\{|T_{2n,k}-S_{2,k}|\geq\frac{t}{2}\Big\}+\mathrm{pr}\Big\{|\frac{3n-2}{n^2}S_{2,k}|\geq\frac{t}{2}\Big\} \\ &\leq 2\exp\Big\{-\frac{nt^2}{24\pi^2}\Big\}, \end{aligned} \tag{S1.10}$$

for n large enough. By Hoeffding's inequality, we have

$$\mathrm{pr}\Big\{|\hat{p}_k-p_k|\geq t\Big\}\leq 2\exp\Big\{-2nt^2\Big\}. \tag{S1.11}$$

Note that

$$\begin{aligned} \mathrm{pr}\Big\{|\hat{p}_k^{-1}S_{2n,k}-p_k^{-1}S_{2,k}|\geq t\Big\} &= \mathrm{pr}\Big\{|\hat{p}_k^{-1}[S_{2n,k}-S_{2,k}]+S_{2,k}[\hat{p}_k^{-1}-p_k^{-1}]|\geq t\Big\} \\ &\leq \mathrm{pr}\Big\{|\frac{S_{2n,k}-S_{2,k}}{\hat{p}_k}|\geq\frac{t}{2}\Big\}+\mathrm{pr}\Big\{S_{2,k}|\hat{p}_k^{-1}-p_k^{-1}|\geq\frac{t}{2}\Big\}. \end{aligned}$$

Combining (S1.10), (S1.11) with the condition $c_1/K \leq \min_{1\leq k\leq K} p_k \leq \max_{1\leq k\leq K} p_k \leq c_2/K$, we have

$$\begin{aligned} \mathrm{pr}\Big\{\hat{p}_k^{-1}|S_{2n,k}-S_{2,k}|\geq t\Big\} &\leq \mathrm{pr}\Big\{|\frac{S_{2n,k}-S_{2,k}}{\hat{p}_k}|\geq t,\hat{p}_k\geq\frac{c_1}{2K}\Big\}+\mathrm{pr}\Big\{\hat{p}_k<\frac{c_1}{2K}\Big\} \\ &\leq \mathrm{pr}\Big\{|S_{2n,k}-S_{2,k}|\geq\frac{c_1t}{2K}\Big\}+\mathrm{pr}\Big\{|\hat{p}_k-p_k|\geq\frac{c_1t}{2K}\Big\} \\ &\leq 2\exp\Big\{-n\frac{c_1^2t^2}{48K^2}\Big\}+2\exp\Big\{-2n\frac{c_1^2t^2}{4K^2}\Big\} \\ &\leq 4\exp\Big\{-c_0\frac{nt^2}{K^2}\Big\}, \end{aligned}$$

for some positive constant c_0, and

$$
\begin{aligned}
\mathrm{pr}\left\{S_{2,k}|\hat{p}_k^{-1}-p_k^{-1}|\geq t\right\} &\leq \mathrm{pr}\left\{|\frac{\hat{p}_k-p_k}{\hat{p}_k p_k}|\geq \frac{t}{\pi}, \hat{p}_k \geq \frac{c_1}{2K}\right\} + \mathrm{pr}\left\{\hat{p}_k < \frac{c_1}{2K}\right\} \\
&\leq \mathrm{pr}\left\{|\hat{p}_k-p_k|\geq \frac{c_1^2 t}{2K^2}\right\} + \mathrm{pr}\left\{|\hat{p}_k-p_k|\geq \frac{c_1 t}{2K}\right\} \\
&\leq 2\exp\left\{-2n\frac{c_1^4 t^2}{4K^4}\right\} + 2\exp\left\{-2n\frac{c_1^2 t^2}{4K^2}\right\} \\
&\leq 4\exp\left\{-c_0\frac{nt^2}{K^4}\right\}.
\end{aligned}
$$

These, together with union bound, yield that

$$
\begin{aligned}
\mathrm{pr}\left\{|S_{2n}-S_2|\geq t\right\} &\leq \sum_{k=1}^{K}\mathrm{pr}\left\{|\hat{p}_k^{-1}S_{2n,k}-p_k^{-1}S_{2,k}|\geq \frac{t}{K}\right\} \\
&\leq 4K\left[\exp\left\{-c_0\frac{nt^2}{K^4}\right\}+\exp\left\{-c_0\frac{nt^2}{K^6}\right\}\right] \\
&\leq 8K\exp\left\{-c_0\frac{nt^2}{K^6}\right\}. \qquad \text{(S1.12)}
\end{aligned}
$$

Combining (S1.3), (S1.9) with (S1.12), we obtain that

$$
\begin{aligned}
\mathrm{pr}\left\{|\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y)-\mathrm{PMV}(\mathbf{X}|Y)|\geq t\right\} &\leq \mathrm{pr}\left\{|S_{1n}-S_1|\geq \frac{t}{2}\right\} + \mathrm{pr}\left\{|S_{2n}-S_2|\geq \frac{t}{2}\right\} \\
&\leq 2\exp\left\{-\frac{nt^2}{24\pi^2}\right\} + 8K\exp\left\{-c_0\frac{nt^2}{K^6}\right\} \\
&\leq 9K\exp\left\{-c_0\frac{nt^2}{K^6}\right\},
\end{aligned}
$$

for n large enough. Thus, when $t \geq c_0\frac{K^3}{\sqrt{n}}\log(K/\alpha)$, we have

$$
\mathrm{pr}\left\{|\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y)-\mathrm{PMV}(\mathbf{X}|Y)|\leq c_0\sqrt{\frac{K^6}{n}}\log(K/\alpha)\right\} \geq 1-\alpha,
$$

provided that n is sufficiently large. □

Proof of Theorem 4. By the properties of conditional expectation, we have

$$
\begin{aligned}
& [\mathrm{pr}\{X \le x | Y = y_k\} - \mathrm{pr}\{X \le x\}]^2 \\
= & \Big[E[I(X \le x)|Y = y_k] - E[I(X \le x)]\Big]\Big[E[I(X \le x)|Y = y_k] - E[I(X \le x)]\Big] \\
= & E[I(X_1 \le x)I(X_2 \le x)|Y_1 = y_k, Y_2 = y_k] - 2E[I(X_1 \le x)I(X_2 \le x)|Y_1 = y_k] \\
& + E[I(X_1 \le x)I(X_2 \le x)] \\
= & E[I(X_1 \le x)I(X_2 \le x)|Y_1 = y_k, Y_2 = y_k] + E[I(X_1 \le x)I(X_2 \le x)] \\
& - 2p_k^{-1}E[I(X_1 \le x)I(X_2 \le x)I(Y_1 = y_k)],
\end{aligned}
$$

where (Y_1, X_1) and (Y_2, X_2) are independent and identically distributed. This leads to

$$
\begin{aligned}
\int_{-\infty}^{\infty} [F_k(x) - F(x)]^2 dF(x) = & E[I(X_1 \le X_3)I(X_2 \le X_3)|Y_1 = y_k, Y_2 = y_k] \\
& - 2p_k^{-1}E[I(X_1 \le X_3)I(X_2 \le X_3)I(Y_1 = y_k)] \\
& + E[I(X_1 \le X_3)I(X_2 \le X_3)]. \qquad \text{(S1.13)}
\end{aligned}
$$

By the result in (2.2), we obtain that

$$
\begin{aligned}
& \mathrm{MV}(X|Y) \\
= & \sum_{k=1}^{K} p_k E[I(X_1 \leq X_3)I(X_2 \leq X_3)|Y_1 = y_k, Y_2 = y_k] - E[I(X_1 \leq X_3)I(X_2 \leq X_3)] \\
= & \sum_{k=1}^{K} p_k E\{[1 - I(X_1 > X_3)]I(X_2 \leq X_3)|Y_1 = y_k, Y_2 = y_k] \\
& -E[\{1 - I(X_1 > X_3)\}I(X_2 \leq X_3)] \\
= & E[I(X_1 > X_3)I(X_2 \leq X_3)\} - E[I(X_2 \leq X_3)] + \sum_{k=1}^{K} p_k E[I(X_2 \leq X_3)|Y_2 = y_k] \\
& -\sum_{k=1}^{K} p_k E[I(X_1 > X_3)I(X_2 \leq X_3)|Y_1 = y_k, Y_2 = y_k] \\
= & E[I(X_1 > X_3)I(X_2 \leq X_3)] - \sum_{k=1}^{K} p_k E[I(X_1 > X_3)I(X_2 \leq X_3)|Y_1 = Y_2 = y_k] \\
= & E[I(X_1 > X_3)I(X_2 \leq X_3)] - \sum_{k=1}^{K} p_k^{-1} E[I(X_1 > X_3)I(X_2 \leq X_3)I(Y_1 = Y_2 = y_k)].
\end{aligned}
$$

To prove this theorem, we next need to prove

$$
\mathrm{S_W}(X|Y) = \sum_{k=1}^{K} p_k^{-1} E[I(X_1 > X_3)I(X_2 \leq X_3)I(Y_1 = Y_2 = y_k)].
$$

Note that

$$
\begin{aligned}
& \{E[I(X_1 \leq x)|Y_1 = y_k] - I(X_1 \leq x)\}^2 \\
= & E[I(X_1 \leq x)I(X_2 \leq x)|Y_1 = Y_2 = y_k] + I(X_1 \leq x) - 2E[I(X_1 \leq x)|Y_1 = y_k]I(X_1 \leq x).
\end{aligned}
$$

Then, we have that

$$
\begin{aligned}
& \mathrm{S_W}(X|Y) \\
= & \sum_{k=1}^{K} p_k \int_{-\infty}^{\infty} E[\{E[I(X_1 \le x)|Y_1 = y_k] - I(X_1 \le x)\}^2 | Y_1 = Y_2 = y_k] dF(x) \\
= & \sum_{k=1}^{K} p_k \int_{-\infty}^{\infty} \{E[I(X_1 \le x) I(X_2 \le x)|Y_1 = Y_2 = y_k] + E[I(X_1 \le x)|Y_1 = Y_2 = y_k] \\
& -2E[I(X_1 \le x) I(X_2 \le x)|Y_1 = Y_2 = y_k] dF(x) \\
= & \sum_{k=1}^{K} p_k \int_{-\infty}^{\infty} E[\{I(X_1 \le x) - I(X_1 \le x) I(X_2 \le x)\}|Y_1 = Y_2 = y_k] dF(x) \\
= & \sum_{k=1}^{K} p_k E[I(X_2 \le X_3) - I(X_1 \le X_3) I(X_2 \le X_3)|Y_1 = Y_2 = y_k] \\
= & \sum_{k=1}^{K} p_k^{-1} E[I(Y_1 = Y_2 = y_k) I(X_1 > X_3) I(X_2 \le X_3)].
\end{aligned}
$$

This completes the proof of 4.

□

Proof of Theorem 5. By the definition of $\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y)$, we have

$$
\begin{aligned}
\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y) &= \frac{1}{nc_p} \sum_{k=1}^{K} \hat{p}_k \sum_{i=1}^{n} \int_{\mathbb{S}^{p-1}} \left\{ g_{U,Y}^n(\beta^T \mathbf{X}_i; y_k) - g_U^n(\beta^T \mathbf{X}_i) \right\}^2 d\beta \\
&= \frac{1}{c_p} \int_{\mathbb{S}^{p-1}} \int_{-\infty}^{\infty} \sum_{k=1}^{K} \hat{p}_k^{-1} \left\{ g_{U,Y}^n(u; y_k) \hat{p}_k - g_U^n(u) \hat{p}_k \right\}^2 d\hat{F}_U(u) d\beta \\
&= \frac{1}{c_p} \frac{1}{n} \int_{\mathbb{S}^{p-1}} \int_{-\infty}^{\infty} \sum_{k=1}^{K} W_{n,k}(\beta, u)^2 d\hat{F}_U(u) d\beta, \qquad \text{(S1.14)}
\end{aligned}
$$

where $\hat{F}_U(u) = g_U^n(u) = n^{-1} \sum_{i=1}^{n} I(U_i \le u)$ and

$$
W_{n,k}(\beta, u) := n^{1/2} \Big[g_{U,Y}^n(u; y_k) \hat{p}_k - g_U^n(u) \hat{p}_k \Big] / \sqrt{\hat{p}_k},
$$

for $u \in \mathcal{R}$ and $\beta \in \mathbb{S}^{p-1}$.

By simple calculation, we have that

$$W_{n,k}(\beta, u) = n^{1/2}\Big[\frac{1}{n}\sum_{i=1}^{n} I(U_i \le u, Y_i = y_k) - \frac{1}{n}\sum_{i=1}^{n} I(U_i \le u)\frac{1}{n}\sum_{i=1}^{n} I(Y_i = y_k)\Big]/\sqrt{\hat{p}_k}.$$

Let $\mathbf{W}_n(\beta, u) := (W_{n,1}(\beta, u), \cdots, W_{n,K}(\beta, u))^T$. To study the consistency of $\mathbf{W}_n(\beta, u)$, we consider the following K-dimensional empirical process

$$\mathbf{R}_n(\beta, u) := (R_{n,1}(\beta, u), \cdots, R_{n,K}(\beta, u))^T,$$

where

$$\begin{aligned} R_{n,k}(\beta, u) &= n^{-1/2}\sum_{i=1}^{n} \frac{1}{\sqrt{p_k}}\{I(U_i \le u) - F_{\beta^T\mathbf{X}}(u)\}\{I(Y_i = y_k) - p_k\} \\ &:= n^{-1/2}\sum_{i=1}^{n} \phi_k(\mathbf{X}_i, Y_i; \beta, u) \end{aligned}$$

and $\phi_k(\mathbf{X}_i, Y_i; \beta, u) = \{I(U_i \le u) - F_{\beta^T\mathbf{X}}(u)\}\{I(Y_i = y_k) - p_k\}/\sqrt{p_k}$.

By (S1.14), we can obtain that

$$\begin{aligned} n\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y) &= \frac{1}{c_p}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}\sum_{k=1}^{K} W_{n,k}(\beta, u)^2 d\hat{F}_U(u)d\beta \\ &= \frac{1}{c_p}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}\sum_{k=1}^{K} [W_{n,k}(\beta, u) - R_{n,k}(\beta, u)]^2 d\hat{F}_U(u)d\beta \\ &\quad + \frac{1}{c_p}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}\sum_{k=1}^{K} R_{n,k}(\beta, u)^2 d\hat{F}_U(u)d\beta \\ &\quad + \frac{2}{c_p}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}\sum_{k=1}^{K} [W_{n,k}(\beta, u) - R_{n,k}(\beta, u)]R_{n,k}(\beta, u)d\hat{F}_U(u)d\beta. \end{aligned}$$

By simple algebraic manipulation, we have

$$R_{n,k}(\beta, u) - W_{n,k}(\beta, u) = n^{-1}\sum_{i=1}^{n}\{I(Y_i = y_k) - p_k\}n^{-1/2}\sum_{i=1}^{n}\{I(\beta^T\mathbf{X}_i \le u) - F_{\beta^T\mathbf{X}}(u)\}.$$

By the law of large numbers, we have $n^{-1}\sum_{i=1}^{n}\{I(Y_i = y_k) - p_k\} = o_p(1)$. Using Theorem 2.5.2 in van der Vaart and Wellner (1996), we can show that

$$n^{-1/2}\sum_{i=1}^{n}\Big\{I(\beta^T\mathbf{X}_i \le u) - F_{\beta^T\mathbf{X}}(u)\Big\},$$

converges to a Gaussian process with zero mean and covariance function $P\{\beta^T\mathbf{X} \le u\}(1 - P\{\beta^T\mathbf{X} \le u\})$. This yields that,

$$R_{n,k}(\beta, u) - W_{n,k}(\beta, u) = o_p(1), \tag{S1.15}$$

holds uniformly for $(u, \beta) \in \mathbb{S}^{p-1} \times \mathcal{R}$. By Proposition 7.27 in Kosorok (2008), we have

$$\frac{1}{c_p}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}\sum_{k=1}^{K}[W_{n,k}(\beta, u) - R_{n,k}(\beta, u)]^2 d\hat{F}_U(u)d\beta = o_p(1)$$

and

$$\frac{1}{c_p}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}\sum_{k=1}^{K}[W_{n,k}(\beta, u) - R_{n,k}(\beta, u)]R_{n,k}(\beta, u)d\hat{F}_U(u)d\beta = o_p(1).$$

These lead to

$$n\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y) = c_p^{-1}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}\sum_{k=1}^{K}R_{n,k}(\beta, u)^2 dF_U(u)d\beta + o_p(1). \tag{S1.16}$$

Thus, we only need to establish the convergence of

$$\frac{1}{c_p}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}\sum_{k=1}^{K} R_{n,k}(\beta,u)^2 dF_U(u)d\beta.$$

By the theory of empirical process (van der Vaart and Wellner, 1996), we have

$$\Big\{\mathbf{R}_n(\beta,u) : (\beta,u)\in\mathbb{S}^{p-1}\times\mathcal{R}\Big\} \\ \rightsquigarrow \Big\{\mathbf{R}(\beta,u) = (R_1(\beta,u), R_2(\beta,u),\cdots,R_K(\beta,u))^T : (\beta,u)\in\mathbb{S}^{p-1}\times\mathcal{R}\Big\},$$

where $\rightsquigarrow$ denotes the convergence in distribution and $\{\mathbf{R}(\beta,u) : (\beta,u)\in\mathbb{S}^{p-1}\times\mathcal{R}\}$ is a K-dimensional Gaussian process. If $H_0 : F_1(\mathbf{x}) = F_2(\mathbf{x}) = \cdots = F_K(\mathbf{x}),$ for all $\mathbf{x}\in\mathcal{R}^p$ holds, it is equivalent to that the distributions of $\mathbf{X}$ given $Y = y_k$ are the same and thus $\mathbf{X}\perp\!\!\!\perp Y$. Let

$$\mathrm{Cov}(\mathbf{R}_n(\beta_1,u_1),\mathbf{R}_n(\beta_2,u_2)) = \Sigma = (\sigma_{k_1,k_2})_{K\times K}$$

be a $K\times K$ matrix with each element σ_{k_1,k_2} defined by

$$\sigma_{k_1,k_2} = E[\phi_{k_1}(\mathbf{X},Y;\beta_1,u_1)\phi_{k_2}(\mathbf{X},Y;\beta_2,u_2)].$$

Since $\phi_k(\mathbf{X},Y;\beta,u) = \{I(\beta^T\mathbf{X}\le u) - F_{\beta^T\mathbf{X}}(u)\}\{I(Y = y_k) - p_k\}/\sqrt{p_k}$

under H_0, we obtain that

$$
\begin{aligned}
& E[R_{n,k_1}(\beta_1,u_1)R_{n,k_2}(\beta_2,u_2)] \\
= & E[\phi_{k_1}(\mathbf{X},Y;\beta_1,u_1)\phi_{k_2}(\mathbf{X},Y;\beta_2,u_2)] \\
= & \frac{1}{\sqrt{p_{k_1}p_{k_2}}}E[\{I(\beta_1^T\mathbf{X}\le u_1)-F_{\beta_1^T\mathbf{X}}(u_1)\}\{I(Y=y_{k_1})-p_{k_1}\}\{I(\beta_2^T\mathbf{X}\le u_2)-F_{\beta_2^T\mathbf{X}}(u_2)\} \\
& \times\{I(Y=y_{k_2})-p_{k_2}\}] \\
= & \frac{1}{\sqrt{p_{k_1}p_{k_2}}}E[\{I(\beta_1^T\mathbf{X}\le u_1)-F_{\beta_1^T\mathbf{X}}(u_1)]\}\{I(\beta_2^T\mathbf{X}\le u_2)-F_{\beta_2^T\mathbf{X}}(u_2)\}] \\
& \times E[\{I(Y=y_{k_1})-p_{k_1}\}\{I(Y=y_{k_2})-p_{k_2}\}].
\end{aligned}
$$

Note that

$$
\begin{aligned}
& E[\{I(\beta_1^T\mathbf{X}\le u_1)-F_{\beta_1^T\mathbf{X}}(u_1)]\}\{I(\beta_2^T\mathbf{X}\le u_2)-F_{\beta_2^T\mathbf{X}}(u_2)\}] \\
= & \mathrm{pr}(\beta_1^T\mathbf{X}\le u_1,\beta_2^T\mathbf{X}\le u_2)-F_{\beta_1^T\mathbf{X}}(u_1)F_{\beta_2^T\mathbf{X}}(u_2).
\end{aligned}
$$

If $k_1=k_2$, we have

$$
E[\{I(Y=y_{k_1})-p_{k_1}\}\{I(Y=y_{k_2})-p_{k_2}\}]=p_{k_1}[1-p_{k_1}].
$$

If $k_1\ne k_2$, we have

$$
E[\{I(Y=y_{k_1})-p_{k_1}\}\{I(Y=y_{k_2})-p_{k_2}\}]=-p_{k_1}p_{k_2}.
$$

Thus, we can obtain that

$$
\Sigma=[\mathrm{pr}(\beta_1^T\mathbf{X}\le u_1,\beta_2^T\mathbf{X}\le u_2)-F_{\beta_1^T\mathbf{X}}(u_1)F_{\beta_2^T\mathbf{X}}(u_2)]\left[\mathbf{I}_K-\mathbf{b}_K\ \mathbf{b}_K^T\right],
$$

where $\mathbf{I}_K$ denotes the $K\times K$ identity matrix and $\mathbf{b}_K=(\sqrt{p_1},\sqrt{p_2},\cdots,\sqrt{p_K})^T$.

Let $\Gamma = (\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_K)^T$, where $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{K-1} \in \mathcal{R}^K$ are $K-1$ unit and orthogonal vectors such that Γ is a $K \times K$ orthogonal matrix. Consider the transformed K-dimensional Gaussian process

$$\Gamma \mathbf{R}_n(\beta, u) \rightsquigarrow \mathbf{R}_\Gamma(\beta, u) := \Gamma \mathbf{R}(\beta, u), \text{ for } (\beta, u) \in \mathbb{S}^{p-1} \times \mathcal{R}.$$

Then, we have

$$\begin{aligned}
&\mathrm{Cov}(\Gamma \mathbf{R}_n(\beta_1, u_1), \Gamma \mathbf{R}_n(\beta_2, u_2)) \\
= \; & \Gamma \Sigma \Gamma^T \\
= \; & [\mathrm{pr}(\beta_1^T \mathbf{X} \le u_1, \beta_2^T \mathbf{X} \le u_2) - F_{\beta_1^T \mathbf{X}}(u_1) F_{\beta_2^T \mathbf{X}}(u_2)] \Gamma \left[\mathbf{I}_K - \mathbf{b}_K \ \mathbf{b}_K^T \right] \Gamma^T \\
= \; & [\mathrm{pr}(\beta_1^T \mathbf{X} \le u_1, \beta_2^T \mathbf{X} \le u_2) - F_{\beta_1^T \mathbf{X}}(u_1) F_{\beta_2^T \mathbf{X}}(u_2)] \left[\mathbf{I}_K - \mathrm{diag}(0, \cdots, 0, 1) \right] \\
= \; & [\mathrm{pr}(\beta_1^T \mathbf{X} \le u_1, \beta_2^T \mathbf{X} \le u_2) - F_{\beta_1^T \mathbf{X}}(u_1) F_{\beta_2^T \mathbf{X}}(u_2)] \mathrm{diag}\{1, \cdots, 1, 0\}_{K \times K}.
\end{aligned}$$

It implies that each component of $\Gamma \mathbf{R}_n(\beta, u)$ (or $\mathbf{R}_\Gamma(\beta, u)$) is independent. By applying the continuous mapping theorem, we have

$$\sum_{k=1}^K R_{n,k}(\beta, u)^2 = \left\| \mathbf{R}_n(\beta, u) \right\|^2 = \left\| \Gamma \mathbf{R}_n(\beta, u) \right\|^2 \rightsquigarrow \left\| \mathbf{R}_\Gamma(\beta, u) \right\|^2, \text{ for } (\beta, u) \in \mathbb{S}^{p-1} \times \mathcal{R}.$$

Therefore, we have

$$\begin{aligned}
\frac{1}{c_p} \int_{\mathbb{S}^{p-1}} \int_{-\infty}^{\infty} \sum_{k=1}^K R_{n,k}(\beta, u)^2 dF_U(u) d\beta \quad &= \quad \frac{1}{c_p} \int_{\mathbb{S}^{p-1}} \int_{-\infty}^{\infty} \left\| \mathbf{R}_n(\beta, u) \right\|^2 dF_U(u) d\beta \\
&\xrightarrow{d} \quad \frac{1}{c_p} \int_{\mathbb{S}^{p-1}} \int_{-\infty}^{\infty} \left\| \mathbf{R}_\Gamma(\beta, u) \right\|^2 dF_U(u) d\beta \\
&= \quad \frac{1}{c_p} \sum_{k=1}^{K-1} \int_{\mathbb{S}^{p-1}} \int_{-\infty}^{\infty} \mathbf{R}_{\Gamma,k}(\beta, u)^2 dF_U(u) d\beta.
\end{aligned}$$

According to Kuo (1975, Chapter 1, Section 2), we derive that

$$\frac{1}{c_p}\sum_{k=1}^{K-1}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}\mathbf{R}_{\Gamma,k}(\beta,u)^2 dF_U(u)d\beta \tag{S1.17}$$

follows the same distribution as $\sum_{k=1}^{\infty}\lambda_k\eta_k^2$, where the λ_k depend on the distribution of $(\mathbf{X}, Y)$ and the η_k are independent standard normal random variables. Thus, together with (S1.18), we have

$$\begin{aligned} n\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y) &= c_p^{-1}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}\sum_{k=1}^{K} R_{n,k}(\beta,u)^2 dF_U(u)d\beta + o_p(1) \\ &\xrightarrow{d} \sum_{k=1}^{\infty}\lambda_k\eta_k^2, \quad n\to\infty. \end{aligned} \tag{S1.18}$$

It is easy to obtain that $\mathrm{SS}_T = E[\mathrm{ang}(\mathbf{X}_1-\mathbf{X}_3, \mathbf{X}_2-\mathbf{X}_3)] + o_p(1)$. By Theorem 3, we obtain $\mathrm{SS}_W = \mathrm{SS}_T - \mathrm{SS}_B = \mathrm{PS}_{\mathrm{W}}(\mathbf{X}|Y) + o_p(1)$. By Sultsky's Theorem, we have

$$F_n = \frac{(n-K)\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y)}{(K-1)\mathrm{SS}_W} \xrightarrow{d} \sum_{k=1}^{\infty}\frac{\lambda_k}{(K-1)\mathrm{PS}_{\mathrm{W}}(\mathbf{X}|Y)}\eta_k^2, \quad n\to\infty.$$

□

Proof of Theorem 6. By the proof of Theorem 3, we have

$$
\begin{aligned}
&\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y) - \mathrm{PMV}(\mathbf{X}|Y) \\
= \;& [S_{1n} - S_{2n}] - [S_1 - S_2] \\
= \;& \Big[\frac{(n-1)(n-2)}{n^2} T_{1n} - \frac{(n-1)(n-2)}{n^2} \sum_{k=1}^{K} \hat{p}_k^{-1} T_{2n,k}\Big] - [S_1 - \sum_{k=1}^{K} p_k^{-1} S_{2,k}] \\
= \;& \frac{(n-1)(n-2)}{n^2}[T_{1n} - S_1] - \frac{3n-2}{n^2} S_1 - \frac{(n-1)(n-2)}{n^2} \sum_{k=1}^{K} \hat{p}_k^{-1}[T_{2n,k} - S_{2,k}] \\
& - \frac{(n-1)(n-2)}{n^2} \sum_{k=1}^{K} [\hat{p}_k^{-1} - p_k^{-1}] S_{2,k} + \frac{3n-2}{n^2} \sum_{k=1}^{K} p_k^{-1} S_{2,k} \\
= \;& [T_{1n} - S_1] - \sum_{k=1}^{K} p_k^{-1}[T_{2n,k} - S_{2,k}] + \sum_{k=1}^{K} \Big(\hat{p}_k - p_k\Big) p_k^{-2} S_{2,k} + o_p(n^{-1/2}), \qquad \text{(S1.19)}
\end{aligned}
$$

where

$$
\begin{aligned}
T_{1n} = & \binom{n}{3}^{-1} \sum_{i<r<l} \frac{1}{3}[\mathrm{ang}(\mathbf{X}_i - \mathbf{X}_r, \mathbf{X}_l - X_r) + \mathrm{ang}(\mathbf{X}_r - \mathbf{X}_i, \mathbf{X}_l - \mathbf{X}_i) + \mathrm{ang}(\mathbf{X}_i - \mathbf{X}_l, \mathbf{X}_r - \mathbf{X}_l)] \\
:= & \binom{n}{3}^{-1} \sum_{i<r<l} k_1(\mathbf{X}_i, \mathbf{X}_r, \mathbf{X}_l)
\end{aligned}
$$

and

$$
\begin{aligned}
T_{2n,k} = & \binom{n}{3}^{-1} \sum_{i<r<l} \frac{1}{3}[\mathrm{ang}(\mathbf{X}_i - \mathbf{X}_r, \mathbf{X}_l - \mathbf{X}_r) I(Y_i = Y_l = y_k) \\
& + \mathrm{ang}(\mathbf{X}_r - \mathbf{X}_i, \mathbf{X}_l - \mathbf{X}_i) I(Y_r = Y_l = y_k) + \mathrm{ang}(\mathbf{X}_i - \mathbf{X}_l, \mathbf{X}_r - \mathbf{X}_l) I(Y_i = Y_r = y_k)] \\
:= & \binom{n}{3}^{-1} \sum_{i<r<l} k_{2,k}(\mathbf{X}_i, Y_i; \mathbf{X}_r, Y_r; \mathbf{X}_l, Y_l).
\end{aligned}
$$

To obtain the asymptotic normality, we approximate the U-statistics with their projections (Serfling, 1980)

$$\begin{aligned} T_{1n} - S_1 &= \frac{3}{n}\sum_{i=1}^{n}\{E[k_1(\mathbf{X}_i, \mathbf{X}_r, \mathbf{X}_l)|\mathbf{X}_i] - S_1\} + o_p(n^{-1}) \\ &=: \frac{3}{n}\sum_{i=1}^{n}\tilde{h}_1(\mathbf{X}_i) + o_p(n^{-1}), \\ T_{2n,k} - S_{2,k} &= \frac{3}{n}\sum_{i=1}^{n}\{E[k_{2,k}(\mathbf{X}_i, Y_i; \mathbf{X}_r, Y_r; \mathbf{X}_l, Y_l)|\mathbf{X}_i, Y_i] - S_{2,k}\} + o_p(n^{-1}) \\ &=: \frac{3}{n}\sum_{i=1}^{n}\tilde{h}_{2,k}(\mathbf{X}_i, Y_i) + o_p(n^{-1}), \end{aligned}$$

where

$$\tilde{h}_1(\mathbf{x}) = E[k_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)|\mathbf{X}_1 = \mathbf{x}] - S_1$$

and

$$\tilde{h}_{2,k}(\mathbf{x}, y) = E[k_{2,k}(\mathbf{X}_1, Y_1; \mathbf{X}_2, Y_2; \mathbf{X}_3, Y_3)|\mathbf{X}_1 = \mathbf{x}, Y_1 = y] - S_{2,k}.$$

This, together with (S1.19), yields that

$$\sqrt{n}(\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y) - \mathrm{PMV}(\mathbf{X}|Y)) = n^{-1/2}\sum_{i=1}^{n}\Phi(\mathbf{X}_i, Y_i) + o_p(1),$$

where

$$\Phi(\mathbf{X}_i, Y_i) = 3\tilde{h}_1(\mathbf{X}_i) + 3\sum_{k=1}^{K} p_k^{-1}\tilde{h}_{2,k}(\mathbf{X}_i, Y_i) + \sum_{k=1}^{K} p_k^{-2} S_{2,k}[I(Y_i = y_k) - p_k]. \tag{S1.20}$$

Then by the limit central theorem, we have

$$\sqrt{n}(\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y) - \mathrm{PMV}(\mathbf{X}|Y)) \xrightarrow{d} N\left(0, \sigma^2\right), \tag{S1.21}$$

where $\sigma^2 = \mathrm{Var}\left[\Phi(\mathbf{X}_i, Y_i)\right].$

□

Proof of Corollary 2. For an arbitrary positive constant c, we have

$$\begin{aligned}
&\mathrm{pr}\left\{F_n > c\right\} \\
= \ &\mathrm{pr}\left\{\frac{(n-K)\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y)}{(K-1)\mathrm{SS}_W} > c\right\} \\
= \ &\mathrm{pr}\left\{\sqrt{n}\sigma^{-1}[\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y) - \mathrm{PMV}(\mathbf{X}|Y)] > \frac{c(K-1)\mathrm{SS}_W - (n-K)\mathrm{PMV}(\mathbf{X}|Y)}{(n-K)\sigma/\sqrt{n}}\right\} \\
\to \ &1 - \Phi\Big(\frac{c(K-1)\mathrm{SS}_W - (n-K)\mathrm{PMV}(\mathbf{X}|Y)}{(n-K)\sigma/\sqrt{n}}\Big),
\end{aligned}$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution.

Under the alternative hypothesis, we have $\mathrm{PMV}(\mathbf{X}|Y) > 0$. This indicates that

$$\frac{c(K-1)\mathrm{SS}_W - (n-K)\mathrm{PMV}(\mathbf{X}|Y)}{(n-K)\sigma/\sqrt{n}} \to -\infty, \quad n \to \infty.$$

Thus, we can prove that $\lim_{n\to\infty} \mathrm{pr}\{F_n > c\} = 1.$ □

Proof of Theorem 7. For notation convenience, let $F(u;\beta) = \mathrm{pr}\{\beta^T\mathbf{X} \leq u\}$, $F_{a\cdot}(u;\beta) = \mathrm{pr}\{\beta^T\mathbf{X} \leq u|A = a\}$, $F_{\cdot b}(u;\beta) = \mathrm{pr}\{\beta^T\mathbf{X} \leq u|B = b\}$ and $F_{ab}(u;\beta) = \mathrm{pr}\{\beta^T\mathbf{X} \leq u|A = a, B = b\}$. Define $p_{ab} = 1/(K_AK_B)$, $p_{a\cdot} = 1/K_A$ and $p_{\cdot b} = 1/K_B$.

(i) By similar arguments of (3.5), we consider the following decomposition

$$I(\beta^T\mathbf{X} \le u) - F(u;\beta) = [F_{a\cdot}(u;\beta) - F(u;\beta)] + [F_{\cdot b}(u;\beta) - F(u;\beta)] + [I(\beta^T\mathbf{X} \le u) - F_{a\cdot}(u;\beta) - F_{\cdot b}(u;\beta) + F(u;\beta)].$$

Let $E_{a,b}[\cdot]$ be conditional expectation $E[\cdot|A = a, B = b]$. Then, for any $\beta \in \mathbb{S}^{p-1}$ and $u \in \mathcal{R}$, we have

$$\begin{aligned}
& \mathrm{var}(I(\beta^T\mathbf{X} \le u)) \\
= & \sum_{a=1}^{K_A}\sum_{b=1}^{K_B} p_{ab} E[\{I(\beta^T\mathbf{X} \le u) - F(u;\beta)\}^2 | A = a, B = b] \\
= & \sum_{a=1}^{K_A} p_{a\cdot}\{F_{a\cdot}(u;\beta) - F(u;\beta)\}^2 + \sum_{b=1}^{K_B} p_{\cdot b}\{F_{\cdot b}(u;\beta) - F(u;\beta)\}^2 \\
& + \sum_{a=1}^{K_A}\sum_{b=1}^{K_B} p_{ab} E_{a,b}[\{I(\beta^T\mathbf{X} \le u) - F_{a\cdot}(u;\beta) - F_{\cdot b}(u;\beta) + F(u;\beta)\}^2]. \quad \text{(S1.22)}
\end{aligned}$$

Note that, (S1.22) is obtained by the fact that all of the crossproduct terms in above equation are orthogonal. That is,

$$\sum_{a=1}^{K_A}\sum_{b=1}^{K_B} p_{ab}\left\{F_{a\cdot}(u;\beta) - F(u;\beta)\right\}\left\{F_{\cdot b}(u;\beta) - F(u;\beta)\right\} = 0, \qquad \text{(S1.23)}$$

$$\sum_{a=1}^{K_A}\sum_{b=1}^{K_B} p_{ab}\{F_{a\cdot}(u;\beta) - F(u;\beta)\} E_{a,b}[I(\beta^T\mathbf{X} \le u) - F_{a\cdot}(u;\beta) - F_{\cdot b}(u;\beta) + F(u;\beta)] = 0 \qquad \text{(S1.24)}$$

and

$$\sum_{a=1}^{K_A}\sum_{b=1}^{K_B} p_{ab}\{F_{\cdot b}(u;\beta)-F(u;\beta)\}E_{a,b}[I(\beta^T\mathbf{X}\le u)-F_{a\cdot}(u;\beta)-F_{\cdot b}(u;\beta)+F(u;\beta)] = 0. \tag{S1.25}$$

We here only check (S1.24). The results in (S1.23) and (S1.25) can be proved similarly. By the properties of conditional expectation, we obtain that

$$\begin{aligned}
& \sum_{a=1}^{K_A}\sum_{b=1}^{K_B} p_{ab}\{F_{a\cdot}(u;\beta)-F(u;\beta)\}E_{a,b}[I(\beta^T\mathbf{X}\le u) - F_{a\cdot}(u;\beta) - F_{\cdot b}(u;\beta) + F(u;\beta)] \\
= & \sum_{a=1}^{K_A}\{F_{a\cdot}(u;\beta)-F(u;\beta)\}\sum_{b=1}^{K_B} p_{a,b}[E_{a,b}[I(\beta^T\mathbf{X}\le u)]-F_{a\cdot}(u;\beta)-F_{\cdot b}(u;\beta)+F(u;\beta)] \\
= & \sum_{a=1}^{K_A}\{F_{a\cdot}(u;\beta)-F(u;\beta)\}[E[I(\beta^T\mathbf{X}\le u)I(A=a)]-F_{a\cdot}(u;\beta)p_{a\cdot}-F(u;\beta)p_{a\cdot}+F(u;\beta)p_{a\cdot}] \\
= & \ 0.
\end{aligned}$$

By (S1.22) and the definition of $\mathrm{PMV}(\cdot|\cdot)$, we have that

$$\begin{aligned}
& c_p^{-1}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}\mathrm{var}(I(\beta^T\mathbf{X}\le u))dF(u;\beta)d\beta \\
= & \sum_{a=1}^{K_A} p_{a\cdot}c_p^{-1}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}\{F_{a\cdot}(u;\beta) - F(u;\beta)\}^2 dF(u;\beta)d\beta \\
& +\sum_{b=1}^{K_B} p_{\cdot b}c_p^{-1}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}\{F_{\cdot b}(u;\beta) - F(u;\beta)\}^2 dF(u;\beta)d\beta \\
& +\sum_{a=1}^{K_A}\sum_{b=1}^{K_B} p_{ab}c_p^{-1}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty} E_{a,b}[\{I(\beta^T\mathbf{X}\le u) - F_{a\cdot}(u;\beta) - F_{\cdot b}(u;\beta) + F(u;\beta)\}^2]dF(u;\beta)d\beta \\
= & \ \mathrm{PMV}(\mathbf{X}|A) + \mathrm{PMV}(\mathbf{X}|B) + \sigma_{\mathrm{E},1}^2,
\end{aligned}$$

where

$$\begin{aligned}\sigma^2_{\mathrm{E},1} &= \sum_{a=1}^{K_A}\sum_{b=1}^{K_B} p_{ab} c_p^{-1} \int_{\mathbb{S}^{p-1}} \int_{-\infty}^{\infty} E_{a,b}[\{I(\beta^T \mathbf{X} \le u) - F_{a\cdot}(u;\beta) - F_{\cdot b}(u;\beta) \\ &\quad + F(u;\beta)\}^2] dF(u;\beta) d\beta. \qquad \text{(S1.26)}\end{aligned}$$

(ii) For model $\mathbf{X} \sim A + B + A * B$, we consider the decomposition

$$\begin{aligned}I(\beta^T \mathbf{X} \le u) - F(u;\beta) &= [F_{a\cdot}(u;\beta) - F(u;\beta)] + [F_{\cdot b}(u;\beta) - F(u;\beta)] \\ &\quad + [F_{a,b}(u;\beta) - F_{a\cdot}(u;\beta) - F_{\cdot b}(u;\beta) + F(u;\beta)] \\ &\quad + [I(\beta^T \mathbf{X} \le u) - F_{a,b}(u;\beta)].\end{aligned}$$

Then, we have

$$\begin{aligned}\mathrm{var}(I(\beta^T \mathbf{X} \le u)) &= \sum_{a=1}^{K_A}\sum_{b=1}^{K_B} p_{ab} E[\{I(\beta^T \mathbf{X} \le u) - F(u;\beta)\}^2 | A = a, B = b] \\ &= \sum_{a=1}^{K_A} p_{a\cdot}\{F_{a\cdot}(u;\beta) - F(u;\beta)\}^2 + \sum_{b=1}^{K_B} p_{\cdot b}\{F_{\cdot b}(u;\beta) - F(u;\beta)\}^2 \\ &\quad + \sum_{a=1}^{K_A}\sum_{b=1}^{K_B} p_{ab} E_{a,b}[\{F_{ab}(u;\beta) - F_{a\cdot}(u;\beta) - F_{\cdot b}(u;\beta) + F(u;\beta)\}^2] \\ &\quad + \sum_{a=1}^{K_A}\sum_{b=1}^{K_B} p_{ab} E_{a,b}[\{I(\beta^T \mathbf{X} \le u) - F_{ab}(u;\beta)\}^2]. \qquad \text{(S1.27)}\end{aligned}$$

Here, we can prove that all of the crossproduct terms in above equation are orthogonal using the similar arguments of (S1.24). Then, we have that

$$
\begin{aligned}
& c_p^{-1}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}\mathrm{var}(I(\beta^T\mathbf{X} \le u))dF(u;\beta)d\beta \\
= \ & \sum_{a=1}^{K_A} p_{a\cdot}c_p^{-1}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}\{F_{a\cdot}(u;\beta) - F(u;\beta)\}^2 dF(u;\beta)d\beta \\
& + \sum_{b=1}^{K_B} p_{\cdot b}c_p^{-1}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}\{F_{\cdot b}(u;\beta) - F(u;\beta)\}^2 dF(u;\beta)d\beta \\
& + \sum_{a=1}^{K_A}\sum_{b=1}^{K_B} p_{ab}c_p^{-1}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty} E_{a,b}[\{F_{ab}(u;\beta) - F_{a\cdot}(u;\beta) - F_{\cdot b}(u;\beta) + F(u;\beta)\}^2]dF(u;\beta)d\beta \\
& + \sum_{a=1}^{K_A}\sum_{b=1}^{K_B} p_{ab}c_p^{-1}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty} E_{a,b}[\{I(\beta^T\mathbf{X} \le u) - F_{ab}(u;\beta)\}^2]dF(u;\beta)d\beta \\
= \ & \mathrm{PMV}(\mathbf{X}|A) + \mathrm{PMV}(\mathbf{X}|B) + \mathrm{PMV}(\mathbf{X}|A*B) + \sigma^2_{\mathrm{E},2},
\end{aligned}
$$

where

$$
\sigma^2_{\mathrm{E},2} = \sum_{a=1}^{K_A}\sum_{b=1}^{K_B} p_{ab}c_p^{-1}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty} E_{a,b}[\{I(\beta^T\mathbf{X} \le u) - F_{ab}(u;\beta)\}^2]dF(u;\beta)d\beta \qquad \text{(S1.28)}
$$

and

$$
\begin{aligned}
\mathrm{PMV}(\mathbf{X}|A*B) \ = \ & \sum_{a=1}^{K_A}\sum_{b=1}^{K_B} p_{ab}c_p^{-1}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty} E_{a,b}[\{F_{ab}(u;\beta) - F_{a\cdot}(u;\beta) \\
& - F_{\cdot b}(u;\beta) + F(u;\beta)\}^2]dF(u;\beta)d\beta. \qquad \text{(S1.29)}
\end{aligned}
$$

(iii) Note that

$$
\begin{aligned}
& E_{a,b}[\{F_{ab}(u;\beta) - F_{a\cdot}(u;\beta) - F_{\cdot b}(u;\beta) + F(u;\beta)\}^2] \\
& = E_{a,b}[\{[F_{ab}(u;\beta) - F(u;\beta)] - [F_{a\cdot}(u;\beta) - F(u;\beta)] - [F_{\cdot b}(u;\beta) - F(u;\beta)]\}^2].
\end{aligned}
$$

By the similar arguments of (S1.24), we have

$$\begin{aligned}
& \mathrm{PMV}(\mathbf{X}|A*B) \\
= & \sum_{a=1}^{K_A}\sum_{b=1}^{K_B} p_{ab} c_p^{-1} \int_{\mathbb{S}^{p-1}} \int_{-\infty}^{\infty} E_{a,b}[\{F_{ab}(u;\beta) - F(u;\beta)\}^2] dF(u;\beta) d\beta \\
& + \sum_{a=1}^{K_A}\sum_{b=1}^{K_B} p_{ab} c_p^{-1} \int_{\mathbb{S}^{p-1}} \int_{-\infty}^{\infty} E_{a,b}[\{F_{a\cdot}(u;\beta) - F(u;\beta)\}^2] dF(u;\beta) d\beta \\
& + \sum_{a=1}^{K_A}\sum_{b=1}^{K_B} p_{ab} c_p^{-1} \int_{\mathbb{S}^{p-1}} \int_{-\infty}^{\infty} E_{a,b}[\{F_{\cdot b}(u;\beta) - F(u;\beta)\}^2] dF(u;\beta) d\beta \\
= & \mathrm{PMV}(\mathbf{X}|A:B) - \mathrm{PMV}(\mathbf{X}|A) - \mathrm{PMV}(\mathbf{X}|B).
\end{aligned}$$

□

Bibliography

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer.

Kuo, H.-H. (1975). *Gaussian measures in Banach spaces*. Springer-Verlag.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. New York: Wiley.

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