Proof of Proposition 1. 1. By the LLN $I^2/MN \to 1/(E(M)E(N))$ almost surely and by Lemma 2 $\sum_{i,j} \psi_{ij}/I^2 \to E\psi_{12}$ almost surely. Conditioning on the sample,

$$\begin{aligned} E\psi(\xi_I, \eta_I) &= E(E(\psi(\xi_I, \eta_I) \mid (X_1, Y_1, M_1, N_1), \ldots, (X_I, Y_I, M_I, N_I)) \\ &= E\left(\frac{\sum_{1 \le i,j \le I} \sum_{1 \le k \le M_i, 1 \le l \le N_j} \psi(X_{ik}, Y_{jl})}{\sum_{i=1}^I M_i \sum_{i=1}^I N_i} \right) \\ &= E\left(\frac{\sum_{1 \le i,j \le I} \psi_{ij}}{\sum_{i=1}^I M_i \sum_{i=1}^I N_i} \right) \to \frac{E\psi_{12}}{E(M)E(N)} = \theta_{12}. \end{aligned}$$

The limit is justified since $\sum_{i,j} \psi_{i,j}/(\sum_i M_i \sum_i N_i) \le 1$.

2. The second part follows on showing that $(\xi_I, \eta_I) \to (\xi_\infty, \eta_\infty)$ setwise. For $a, b \in \mathbb{R}$, by a similar argument as above,

$$\begin{aligned} P(\xi_I < a, \eta_I < b) &= E\left(\frac{\sum_{1 \le i,j \le I} \sum_{1 \le k \le M_i, 1 \le l \le N_j} \{X_{ik} < a, Y_{jl} < b\}}{\sum_{i=1}^I M_i \sum_{i=1}^I N_i} \right) \\ &\to \frac{E\left(\sum_{k=1}^{M_1} \{X_{1k} < a\}\right)}{E(M)} \frac{E\left(\sum_{l=1}^{N_1} \{Y_{1l} < b\}\right)}{E(N)}. \end{aligned}$$

The probability of sampling an element from a cluster of size $M = m$ given an initial segment of I samples $(X_1, Y_1, M_1, N_1), \ldots, (X_I, Y_I, M_I, N_I)$, is $\frac{m \sum_{i=1}^I \{M_i = m\}}{\sum_{i=1}^I M_i}$. Along almost any sequence of samples as $I \to \infty$ this relative frequency tends to $\frac{mP(M=m)}{E(M)}$.

Therefore

$$
\begin{aligned}
P(\xi_\infty < a) = \sum_{m=1}^{\infty} & P(\xi_\infty < a \mid \xi_\infty \text{ is sampled from a cluster of size } m) \cdot \\
& P(\xi_\infty \text{ is sampled from a cluster of size } m) \\
= & \sum_{m=1}^{\infty} \frac{1}{m} \sum_{k=1}^{m} P(X_{1k} < a \mid M = m) \frac{mP(M=m)}{E(M)} \\
= & \frac{1}{EM} \sum_{m=1}^{\infty} \sum_{k=1}^{m} P(X_{1k} < a \mid M = m) P(M = m) \\
= & \frac{1}{E(M)} E\left(\sum_{k=1}^{M} \{X_{1k} < a\} \right).
\end{aligned}
$$

Analogously,

$$
P(\eta_\infty < a) = \frac{1}{E(N)} E\left(\sum_{l=1}^{N} \{X_{1l} < a\} \right).
$$

The product is the limit of $P(\xi_I < a, \eta_I < b)$ given above.

□

The following lemma gives a convergence result for a two-sample U-statistic with kernel of degree $(1,1)$ where the data is paired. The corresponding definitions and result for independent samples is given in, e.g., Lee (2019). Let V denote the space of finite sequences.

Lemma 1. *Given a sample $(X_0, Y_0), (X_1, Y_1), \ldots, (X_I, Y_I)$ on $V \times V$ IID according to P and a function $\psi : V \times V \to \mathbb{R}$ in $L^2(P)$, define*

$$
U_I = I^{-2} \sum_{\substack{1 \le i,j \le I \\ i \ne j}} \psi(X_i, Y_j), \qquad V_I = I^{-2} \sum_{1 \le i,j \le I} \psi(X_i, Y_j),
$$

and

$$
\hat{U}_I = I^{-1} \sum_{i=1}^{I} \left(E(\psi(X_i, Y_0) \mid X_i, Y_i) + E(\psi(X_0, Y_i) \mid X_i, Y_i) \right) - 2E\psi(X_1, Y_2).
$$

Then

$$
E(U_I - EU_I - \hat{U}_I)^2 = O(I^{-2}) \text{ and } E(V_I - EV_I - \hat{U}_I)^2 = O(I^{-2}).
$$

Proof of Lemma 1. Define

$$\overline{\psi}_{ij} = \psi(X_i, Y_j) - E(\psi(X_i, Y_0) \mid X_i, Y_i) - E(\psi(X_0, Y_j) \mid X_j, Y_j) + E\psi(X_0, Y_0).$$

Then, for $i \neq j$, $E(\overline{\psi}_{ij} \mid (X_i, Y_i)) = E(\overline{\psi}_{ij} \mid (X_j, Y_j)) = 0$, implying

$$\begin{aligned} E(U_I - EU_I - \hat{U}_I)^2 &= E\left((I)_2^{-1} \sum_{i \neq j} \overline{\psi}_{ij}\right) \\ &= (I)_2^{-2} \sum_{i \neq j} E\overline{\psi}_{ij}^2 + O(I^{-2}) \\ &= O(I^{-2}). \end{aligned}$$

For the second equation,

$$\begin{aligned} E(U_I - EU_I - V_I + EV_I)^2 &= I^{-2} E\left((I)_2^{-1} \sum_{i \neq j} \psi_{ij} - E\psi_{11} + E\psi_{12}\right)^2 \\ &\leq I^{-2}\left((I)_2^{-1} \sum_{i \neq j} E(\psi_{ij} - E\psi_{11} + E\psi_{12})^2\right) \\ &= O(I^{-2}). \end{aligned}$$

□

Corollary 2. *With the same setup as Lemma 1, $U_I - EU_I \to 0$ a.s. and $\sqrt{I}(U_I - EU_I)/\sqrt{\mathrm{Var}(U_I)} \to \mathcal{N}(0, 1)$ in distribution.*

Proof of Corollary 2. By Lemma 1, $U_I - EU_I \to \hat{U}_I$ a.s. and $\sqrt{I}(U_I - EU_I - \hat{U}_I) \to 0$ in quadratic mean, and $\hat{U}_I$ is an IID sum subject to the usual LLN and CLT. □

Proof of Proposition 2.

$$\begin{aligned}
\theta_{11}(P) &= E\left(\frac{\sum_{k=1}^{M}\sum_{l=1}^{N}\psi(X_{1k},Y_{1l})}{MN}\right)\\
&= E\left(\frac{1}{MN}E\left(\sum_{k=1}^{M}\sum_{l=1}^{N}\psi(X_{1k},Y_{1l}) \mid M,N\right)\right)\\
&= E\left(\frac{1}{MN}MNE(\psi(X_{11},Y_{11} \mid M,N))\right) = E\psi(X_{11},Y_{11}).
\end{aligned}$$

Similar to the above,

$$\begin{aligned}
\theta_{12}(P) &= \frac{E\left(\sum_{k=1}^{M_1}\sum_{l=1}^{N_2}\psi(X_{1k},Y_{2l})\right)}{E(M)E(N)}\\
&= \frac{E(M)E(N)E\psi(X_{11},Y_{21})}{E(M)E(N)} = E\psi(X_{11},Y_{21}).
\end{aligned}$$

$\square$

Lemma 3. *Given integrable random variables $M, V, X_1, X_2, \ldots,$ such that $M \in \{1, 2, \ldots\}$ and $\sum_{i=1}^{\infty} E(|X_i|; M \geq i) < \infty$,*

$$E\left(\sum_{i=1}^{M} X_i \middle| M, V\right) = \sum_{i=1}^{M} E(X_i \mid M, V)$$

Proof of Lemma 3.

$$\begin{aligned}
E\left(\sum_{i=1}^{M} X_i \middle| M, V\right) &= E\left(\sum_{m=1}^{\infty}\{M=m\}\sum_{i=1}^{m} X_i \middle| M, V\right)\\
&= \sum_{m=1}^{\infty} E\left(\{M=m\}\sum_{i=1}^{m} X_i \middle| M, V\right)\\
&= \sum_{m=1}^{\infty}\sum_{i=1}^{m}\{M=m\}E(X_i \mid M, V)\\
&= \sum_{i=1}^{M} E(X_i \mid M, V),
\end{aligned}$$

the interchange in the second equality allowed since $E\left|\sum_{i=1}^{M} X_i\right| \leq \sum_{i=1}^{\infty} E(|X_i|; M \geq i) < \infty$. $\square$

Proof of Lemma 4. Define for $n \in \mathbb{N}$ approximations to θ_{11} and θ_{12} by

$$
\begin{aligned}
A_{ij}^{(n)} &= \left\{(x,y) : \frac{i}{2^n} \leq x < \frac{i+1}{2^n}, \frac{j}{2^n} \leq y < \frac{j+1}{2^n}\right\}, \quad -2^{2n} \leq i,j < 2^{2n}-1 \\
\theta_{11}^{(n)} &= \sum_{i=-2^{2n}}^{2^{2n}-1} \sum_{j=i+1}^{2^{2n}-1} P(A_{ij}^{(n)}) + \frac{1}{2} \sum_{i=-2^{2n}}^{2^{2n}-1} P(A_{ii}^{(n)}) \\
\theta_{12}^{(n)} &= \sum_{i=-2^{2n}}^{2^{2n}-1} \sum_{j=i+1}^{2^{2n}-1} P_{\perp\perp}(A_{ij}^{(n)}) + \frac{1}{2} \sum_{i=-2^{2n}}^{2^{2n}-1} P_{\perp\perp}(A_{ii}^{(n)}).
\end{aligned}
$$

Since $\bigcup_n \bigcup_i \bigcup_{j>i} A_{ij}^{(n)} = \{x < y\}$ and $\bigcap_n \bigcup_i A_{ii}^{(n)} = \{x = y\}$, by continuity of measure $\theta_{11}^{(n)} \to \theta_{11}$ and $\theta_{12}^{(n)} \to \theta_{12}$. Therefore, it is enough to establish the inequality (4) for $\theta_{11}^{(n)}$ and $\theta_{12}^{(n)}$.

Fixing n,

$$
\begin{aligned}
&\sum_{i=-2^{2n}}^{2^{2n}-2} \sum_{j=i+1}^{2^{2n}-1} P_{\perp\perp}(A_{ij}^{(n)}) = \sum_{i=-2^{2n}}^{2^{2n}-2} \sum_{j=i+1}^{2^{2n}-1} P_{\perp\perp}(A_{ij}^{(n)}) \\
&= \sum_{i=-2^{2n}}^{2^{2n}-2} \sum_{j=i+1}^{2^{2n}-1} P_{\perp\perp}(\frac{i}{2^n} \leq x < \frac{i+1}{2^n}) P_{\perp\perp}(\frac{j}{2^n} \leq y < \frac{j+1}{2^n}) \\
&\geq \sum_{i=-2^{2n}}^{2^{2n}-2} \sum_{j=i+1}^{2^{2n}-1} (P(A_{ii}^{(n)}) + \sum_{k=i+1}^{2^{2n}-1} P(A_{ik}^{(n)}))(P(A_{jj}^{(n)}) + \sum_{l=-2^{2n}}^{j-1} P(A_{lj}^{(n)})) \\
&= \sum_{i=-2^{2n}}^{2^{2n}-2} \sum_{j=i+1}^{2^{2n}-1} \left(\sum_{k=i+1}^{2^{2n}-1} P(A_{ik}^{(n)}) \sum_{l=-2^{2n}}^{j-1} P(A_{lj}^{(n)}) + P(A_{ii}^{(n)}) \sum_{l=-2^{2n}}^{j-1} P(A_{lj}^{(n)}) \right. \\
&\left. + P(A_{jj}^{(n)}) \sum_{k=i+1}^{2^{2n}-1} P(A_{ik}^{(n)}) + P(A_{ii}^{(n)}) P(A_{jj}^{(n)}) \right).
\end{aligned}
$$

We lower bound the first three terms in parentheses.

First term:

$$\sum_{i=-2^{2n}}^{2^{2n}-2}\sum_{j=i+1}^{2^{2n}-1}\sum_{k=i+1}^{2^{2n}-1} P(A_{ik}^{(n)}) \sum_{l=-2^{2n}}^{j-1} P(A_{lj}^{(n)})$$

$$= \sum_{i=-2^{2n}}^{2^{2n}-2}\sum_{k=i+1}^{2^{2n}-1} P(A_{ik}^{(n)}) \sum_{j=i+1}^{2^{2n}-1}\sum_{l=-2^{2n}}^{j-1} P(A_{lj}^{(n)})$$

$$\geq \sum_{i=-2^{2n}}^{2^{2n}-2}\sum_{k=i+1}^{2^{2n}-1} P(A_{ik}^{(n)}) \sum_{j=i+1}^{2^{2n}-1}\sum_{l=i}^{j-1} P(A_{lj}^{(n)})$$

$$= \sum_{i=-2^{2n}}^{2^{2n}-2}\sum_{k=i+1}^{2^{2n}-1} P(A_{ik}^{(n)}) \sum_{l=i}^{2^{2n}-2}\sum_{j=l+1}^{2^{2n}-1} P(A_{lj}^{(n)})$$

$$= \sum_{i=-2^{2n}}^{2^{2n}-2}\sum_{k=i+1}^{2^{2n}-1} P(A_{ik}^{(n)}) \sum_{j=i+1}^{2^{2n}-1} P(A_{ij}^{(n)}) + \sum_{i=-2^{2n}}^{2^{2n}-2}\sum_{k=i+1}^{2^{2n}-1} P(A_{ik}^{(n)}) \sum_{l=i+1}^{2^{2n}-2}\sum_{j=l+1}^{2^{2n}-1} P(A_{lj}^{(n)})$$

$$\geq \sum_{i=-2^{2n}}^{2^{2n}-2}\sum_{j=i+1}^{2^{2n}-1} P(A_{ij}^{(n)})^2 + \sum_{i=-2^{2n}}^{2^{2n}-2}\sum_{k=i+1}^{2^{2n}-2}\sum_{j=k+1}^{2^{2n}-1} P(A_{ij}^{(n)})P(A_{ik}^{(n)}) + \sum_{i=-2^{2n}}^{2^{2n}-2}\sum_{k=i+1}^{2^{2n}-1} P(A_{ik}^{(n)}) \sum_{l=i+1}^{2^{2n}-2}\sum_{j=l+1}^{2^{2n}-1} P(A_{lj}^{(n)})$$

$$= \sum\sum\sum\sum_{\substack{i\neq k \text{ or } j\neq l \\ j>i \text{ and } l>k}} P(A_{ij}^{(n)})P(A_{kl}^{(n)}) + \sum_{i=-2^{2n}}^{2^{2n}-2}\sum_{j=i+1}^{2^{2n}-1} P(A_{ij}^{(n)})^2$$

$$= \frac{1}{2}\left(\sum_{i=-2^{2n}}^{2^{2n}-2}\sum_{j=i+1}^{2^{2n}-1} P(A_{ij}^{(n)})\right)^2 + \frac{1}{2}\sum_{i=-2^{2n}}^{2^{2n}-2}\sum_{j=i+1}^{2^{2n}-1} P(A_{ij}^{(n)})^2.$$

Middle two terms:

$$\begin{aligned}
&\sum_{i=-2^{2n}}^{2^{2n}-2}\sum_{j=i+1}^{2^{2n}-1}\left(P(A_{ii}^{(n)})\sum_{l=-2^{2n}}^{j-1}P(A_{lj}^{(n)})+P(A_{jj}^{(n)})\sum_{k=i+1}^{2^{2n}-1}P(A_{ik}^{(n)})\right)\\
&=\sum_{i=-2^{2n}}^{2^{2n}-2}P(A_{ii}^{(n)})\sum_{l=i}^{2^{2n}-2}\sum_{j=l+1}^{2^{2n}-1}P(A_{lj}^{(n)})+\sum_{j=-2^{2n}+1}^{2^{2n}-1}P(A_{jj}^{(n)})\sum_{i=-2^{2n}}^{j-1}\sum_{k=i+1}^{2^{2n}-1}P(A_{ik}^{(n)})\\
&=\sum_{i=-2^{2n}}^{2^{2n}-2}P(A_{ii}^{(n)})\sum_{l=i}^{2^{2n}-2}\sum_{j=l+1}^{2^{2n}-1}P(A_{lj}^{(n)})+\sum_{i=-2^{2n}+1}^{2^{2n}-1}P(A_{ii}^{(n)})\sum_{l=-2^{2n}}^{i-1}\sum_{j=l+1}^{2^{2n}-1}P(A_{lj}^{(n)})\\
&=\left(\sum_{i=-2^{2n}}^{2^{2n}-1}P(A_{ii}^{(n)})\right)\left(\sum_{l=-2^{2n}}^{2^{2n}-2}\sum_{j=l+1}^{2^{2n}-1}P(A_{lj}^{(n)})\right).
\end{aligned}$$

The second-to-last equality is just renaming indices.

With these lower bounds,

$$\begin{aligned}
\theta_{12}^{(n)} &= \sum_{i=-2^{2n}}^{2^{2n}-1}\sum_{j=i+1}^{2^{2n}-1}P_{\perp\!\!\!\perp}(A_{ij}^{(n)})+\frac{1}{2}\sum_{i=-2^{2n}}^{2^{2n}-1}P_{\perp\!\!\!\perp}(A_{ii}^{(n)})\\
&\geq \frac{1}{2}\left(\sum_{i=-2^{2n}}^{2^{2n}-2}\sum_{j=i+1}^{2^{2n}-1}P(A_{ij}^{(n)})\right)^2+\left(\sum_{i=-2^{2n}}^{2^{2n}-1}P(A_{ii}^{(n)})\right)\left(\sum_{l=-2^{2n}}^{2^{2n}-2}\sum_{j=l+1}^{2^{2n}-1}P(A_{lj}^{(n)})\right)+\\
&\sum_{i=-2^{2n}}^{2^{2n}-2}\sum_{j=i+1}^{2^{2n}-1}P(A_{ii}^{(n)})P(A_{jj}^{(n)})+\frac{1}{2}\sum_{i=-2^{2n}}^{2^{2n}-1}P(A_{ii}^{(n)})^2\\
&=\frac{1}{2}\left(\sum_{i=-2^{2n}}^{2^{2n}-2}\sum_{j=i+1}^{2^{2n}-1}P(A_{ij}^{(n)})+\sum_{i=-2^{2n}}^{2^{2n}-1}P(A_{ii}^{(n)})\right)^2\\
&=\frac{1}{2}\left(\theta_{11}^{(n)}+\frac{1}{2}\sum_{i=-2^{2n}}^{2^{2n}-1}P(A_{ii}^{(n)})\right)^2.\\
&=\frac{1}{2}\left(\theta_{11}^{(n)}+\frac{1}{2}P(X=Y)\right)^2+o(1).
\end{aligned}$$

The upper bound then follows by the same symmetry argument as given in Section 4.

□

Proof of Theorem 3. With

$$\theta_{11} = \frac{1}{mn}E(\psi_{11}) = \frac{1}{mn}\sum_{i,j}(P(X_{1i} < Y_{1j}) + \frac{1}{2}P(X_{1i} = Y_{1j}))$$

Lemma 4 gives

$$\begin{aligned}
\theta_{12} &= \frac{1}{mn}E(\psi_{12}) = \frac{1}{mn}\sum_{i,j}(P(X_{1i} < Y_{2j}) + \frac{1}{2}P(X_{1i} = Y_{2j})) \\
&\geq \frac{1}{mn}\sum_{i,j}\frac{1}{2}(P(X_{1i} < Y_{1j}) + P(X_{1i} = Y_{1j}))^2 \\
&\geq \frac{1}{2}\left(\frac{1}{mn}\sum_{i,j}(P(X_{1i} < Y_{1j}) + P(X_{1i} = Y_{1j}))\right)^2 \\
&= \frac{1}{2}\left(\theta_{11} + \frac{1}{2mn}\sum_{i,j}P(X_{1i} = Y_{1j})\right)^2.
\end{aligned}$$

The second inequality is Jensen's inequality, which is tight when the pairwise AUCs are all equal. The other bound follows similarly. $\square$

Proof of Theorem 5. By Lemma 1,

$$\sqrt{I}\left(\frac{(I)_2^{-1}\sum_{i\neq j}\psi_{ij} - E\psi_{12}}{\mathrm{sd}(\sqrt{I}(I)_2^{-1}\sum_{i\neq j}\psi_{ij})}, \frac{I^{-2}\sum_{i,j}M_iN_j - E(M)E(N)}{\mathrm{sd}(I^{-3/2}\sum_{i,j}M_iN_j)}, \frac{I^{-1}\sum_i\psi_{ii}/(M_iN_i) - E(\psi_{11}/M_1N_1)}{\mathrm{sd}(\psi_{11}/M_1N_1)}\right)$$

converges to

$$\begin{aligned}
I^{-1/2}\sum_{i=1}^{I}\Bigg(&\frac{E(\psi_{i0} \mid W_i) + E(\psi_{0i} \mid W_i) - 2E\psi_{12}}{\mathrm{sd}(E(\psi_{10} \mid W_1) + E(\psi_{01} \mid W_1))}, \frac{M_iE(N) + N_iE(M) - 2E(M)E(N)}{\mathrm{sd}(M_1E(N) + N_1E(M))}, \\
&\frac{\psi_{ii}/(M_iN_i) - E(\psi_{11}/M_1N_1)}{\mathrm{sd}(\psi_{11}/M_1N_1)}\Bigg)
\end{aligned}$$

in mean-square. The latter is an IID sum with finite covariance matrix and is asymptotically normal by the usual CLT. Applying the delta method with the function $(x, y, z) \mapsto (x/y, z)$, with derivative

$$\left.\begin{pmatrix} 1/y & -x/y^2 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right|_{(x,y)=(\theta_{12},E(M)E(N))}$$

for $y \neq 0$, i.e., $E(M) \neq 0, E(N) \neq 0$, gives the asymptotic normality of $(\theta_{11}, \theta_{12})$. The asymptotic covariance matrix is given by delta method.

$\square$

References

Lee, A. J. (2019). *U-statistics: Theory and Practice.* Routledge.