Web-based Supplementary Materials for Warsserstein Tests for Equality of Several Groups for Distributional Data

Sanghun Jeong[1], Jeffrey S. Morris[2], Hojin Yang[1*]

[1]Department of Statistics, Pusan National University, 2 Busan-daehak-ro 63 Beon-gil, Geumjeong-gu, Busan, 46241, South Korea.
[2]Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, 423 Guardian Drive, Philadelphia, 19104, Pennsylvania, United States.

*Corresponding author(s). E-mail(s): hjyang@pusan.ac.kr;

1 Web Appendix A: Technical details of mathematical results

Proof of Theorem 3.1 Suppose the total number of observed points N_i is a random integer number. Let $\widetilde{f}_i$ be an empirical measure of the form such that $\widetilde{f}_i = N_i^{-1}\sum_{j=1}^{N_i}\delta_{X_{ij}}$, where X_{ij} is independent with the distribution associated with f_i, and $\delta_{X_{ij}}$ is a Dirac measure at X_{ij}. For any $x \in \mathscr{X}$, the kernel density estimator is given as the convolution measure $\widehat{f}_i^*(x) = (\widetilde{f}_i * K_{a_i})(x) = \frac{1}{N_i}\sum_{j=1}^{N_i} K_{a_i}(x - X_{ij})$. The Nadaraya-Watson estimator for the normalization issue can be represented by $\widehat{f}_i(x) = \widehat{f}_i^*(x)/\int \widehat{f}_i^*(x)dx$. This yields

$$
\begin{aligned}
d_W^2(\widetilde{f}_i, \widehat{f}_i) &\leq \frac{1}{N_i}\sum_{j=1}^{N_i} d_W\left(\frac{\delta_{X_{ij}} * K_{a_i}}{\int \widehat{f}_i^*(x)dx}, \delta_{X_{ij}}\right) \\
&= \frac{1}{N_i}\sum_{j=1}^{N_i}\frac{1}{\int \widehat{f}_i^*(x)dx}\int_{\mathscr{X}^2}\|x-z\|^2[\delta_{X_{ij}} * K_{a_i}](x)dx\delta_{X_{ij}}(z)dz \\
&= \frac{1}{N_i}\sum_{j=1}^{N_i}\frac{1}{\int \widehat{f}_i^*(x)dx}\int \|x - X_{ij}\|^2 K_{a_i}(x - X_{ij})dx \leq \frac{a_i^2}{\int \widehat{f}_i^*(x)dx}.
\end{aligned}
$$

Let μ_K be a measure corresponding to the kernel density K and μ_L be the Lebesgue measure. Under the assumption of Section 4.4.1 in **?**, $a_i \leq 1$, we compare two probabilities by using the variable change to derive the lower bound of the denominator in the above equation as the followings

$$\begin{aligned}\int \widehat{f}_i^*(x)dx &= \int_{\mathscr{X}} \frac{1}{N_i}\sum_{j=1}^{N_i} K_{a_i}(x - X_{ij})dx = \frac{1}{N_i}\sum_{j=1}^{N_i}\int_{(\mathscr{X}-X_{ij})/a_i} K(x)dx \\ &= \mu_K\{(\mathscr{X} - X_{ij})/a_i\} \geq \mu_K\{\mathscr{X} - X_{ij}\} = \int_{\mathscr{X}-X_{ij}} K(x)dx \\ &\geq \int_{\mathscr{X}-x_j} K(d_{\text{sup}})dx = K(d_{\text{sup}})\mu_L(\mathscr{X}) > 0,\end{aligned}$$

where it is assumed that the diameter of the compact set $\mathscr{X}$ denoted by $d_{\text{sup}} = \sup_{x,y\in\mathscr{X}} \|x-y\|^2$ is finite. Defining the constant $C_K = K(d_{\text{sup}})\mu_L(\mathscr{X})$ depending on the kernel, we have

$$d_W^2(\widetilde{f}_i, \widehat{f}_i) \leq \frac{1}{N_i}\sum_{j=1}^{N_i} d_W\left(\frac{\delta_{X_{ij}} * K_{a_i}}{\int \widehat{f}_i^*(x)dx}, \delta_{X_{ij}}\right) \leq C_K a_i^2 \quad \text{for } a_i \leq 1. \tag{1}$$

Suppose that the total number of observed points for the ith subejct consists of $N_i = N_i^{(1)}+, \ldots, N_i^{(\eta)}$, where each $N_i^{(1)}$ has the same distribution as η independent copies from π_i, for instance, a Poisson distribution. Then, the empirical measure $\widetilde{f}_i$ similarly consists of $\widetilde{f}_i = N_i^{-1}(\sum_{j=1}^{N_i^{(1)}} \delta_{X_{ij}^{(1)}} + \cdots + \sum_{j=1}^{N_i^{(\eta)}} \delta_{X_{ij}^{(\eta)}})$. Following Karr (2017), we have

$$\frac{1}{\eta}\sum_{k=1}^{\eta}\sum_{j=1}^{N_i^{(\eta)}} \delta_{X_{ij}^{(k)}} \xrightarrow{p} \mu_i.$$

It follows from Condition 3.2 $\mu_i \in \mathcal{W}$ that $\widetilde{f}_i$ converges μ_i in the Wassertein distance, i.e., $d_W(\frac{1}{\eta}\sum_{k=1}^{\eta}\sum_{j=1}^{N_i^{(\eta)}} \delta_{X_{ij}^{(k)}}, \mu_i) \xrightarrow{p} 0$, which leads to the weak convergence because that the weak convergence is equivalent to Wasserstein convergence when Condition 3.3 holds, i.e.,

$$\int_{\mathscr{X}} hd\frac{1}{\eta}\sum_{k=1}^{\eta}\sum_{j=1}^{N_i^{(\eta)}} \delta_{X_{ij}^{(k)}} \xrightarrow{p} \int_{\mathscr{X}} hd\mu_i \quad \text{for all } \; h \in C(\mathscr{X}),$$

where $C(\mathscr{X})$ denotes the class of the functions satisfying the boundedness and continuity on $\mathscr{X}$. It follows from Condition 3.6 $\mathbb{E}[N_i(A)|T_i] = \kappa\mu_i(A)$, for any Borel set

$A \subseteq B(\mathscr{X})$, and Slutsky's theorem that

$$\frac{1}{N_i} \sum_{k=1}^{\eta} \sum_{j=1}^{N_i^{(\eta)}} \delta_{X_{ij}^{(k)}} \xrightarrow{p} \mu_i.$$

Although it is an abused notation, we denote it as $\widetilde{f}_i \xrightarrow{p} f_i$ for the brevity. The result of (1) and the triangle inequality lead to

$$d_W(\widehat{f}_i, f_i) \leq d_W(\widehat{f}_i, \widetilde{f}_i) + d_W(\widetilde{f}_i, f_i) \leq \sqrt{C}_K a_i + d_W(\frac{1}{N_i} \sum_{k=1}^{\eta} \sum_{j=1}^{N_i^{(\eta)}} \delta_{X_{ij}^{(k)}}, f_i) \to 0.$$

This completes the proof for the first statement in Theorem 3.1.
Let us define risk functionals on the $\mathcal{W}(\mathscr{X})$

$$M(f) = \frac{1}{2}\mathbb{E}[d_W^2(f_1, f)], \quad M_n(f) = \frac{1}{2n} \sum_{i=1}^{n} d_W^2(f_i, f)$$
$$\widetilde{M}_n(f) = \frac{1}{2n} \sum_{i=1}^{n} d_W^2(\widetilde{f}_i, f), \quad \widehat{M}_n(f) = \frac{1}{2n} \sum_{i=1}^{n} d_W^2(\widehat{f}_i, f)$$

as the populational, empirical, conditional, and smoothed risk functionals, respectively. We note the property of the Wassertein distance

$$d_W(f, g) \leq \sqrt{\sup_{h \in \mathcal{P}(\mathscr{X}^2)} \int_{\mathscr{X}^2} \|x - y\|^2 dh(x, y)} \leq \sqrt{\sup_{x,y \in \mathscr{X}} \|x - y\|^2} = d_{\sup} < \infty.$$

The key idea of our proof is to show the uniform convergence of $\widehat{M}_n(f)$ to $M(f)$ over the Wasserstein space. We consider the difference between the conditional and smoothed empirical risk functionals, and factorize this difference into two parts as the followings:

$$\begin{aligned} |\widehat{M}(f) - \widetilde{M}_n(f)| &= |d_W(\widetilde{f}_i, f) - d_W(\widehat{f}_i, f)||d_W(\widetilde{f}_i, f) + d_W(\widehat{f}_i, f)| \\ &\leq \frac{2}{n} d_{\sup} \sum_{i=1}^{n} d_W(\widetilde{f}_i, \widehat{f}_i), \end{aligned}$$

where the reverse triangle inequality for the first factor and the property of the Wassertein distance are used for the inequality. Hence we have

$$\sup_{\mu \in W_2(\mathcal{X})} |\widehat{M}(f) - \widetilde{M}_n(f)| \leq d_{\sup} \sqrt{C_K} \frac{1}{n} \sum_{i=1}^{n} a_i \ , \tag{2}$$

where it vanishes under the assumption of Theorem 3.1. We similarly have

$$\sup_{\mu \in W_2(\mathcal{X})} |\widetilde{M}_n(f) - M_n(f)| \leq \frac{1}{n} \sum_{i=1}^{n} d_W(\widetilde{f}_i, f_i).$$

By the weak law of large numbers and the equivalence of the weak convergence to the Wasserstein convergence with Condition 3.3, We have

$$\frac{1}{n} \sum_{i=1}^{n} d_W(\widetilde{f}_i, f_i) \xrightarrow{p} 0 \quad \text{and} \quad \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} d_W(\widetilde{f}_i, f_i)\right] \xrightarrow{p} 0. \tag{3}$$

By the markov inequality, the concentration inequality is given by

$$\mathbb{P}\left[\frac{1}{n} \sum_{i=1}^{n} \left(d_W(\widetilde{f}_i, f_i) - \mathbb{E} d_W(\widetilde{f}_i, f_i)\right)^2 > \epsilon\right] \leq \frac{\mathbb{E}\left(d_W(\widetilde{f}_i, f_i) - \mathbb{E} d_W(\widetilde{f}_i, f_i)\right)^2}{\epsilon^2 n} \leq \frac{4 d_{\sup}}{\epsilon^2 n}, \tag{4}$$

Notice that $d_W(\widetilde{f}_1, f_1), \ldots, d_W(\widetilde{f}_n, f_n)$ are independent and identically distributed. According to the results of (3) and the Borel-Cantelli lemma by taking $\epsilon = n^{-1/3}$ in (4), we obtain

$$\left|\frac{1}{n} \sum_{i=1}^{n} d_W(\widetilde{f}_i, f_i)\right| \leq \left|\frac{1}{n} \sum_{i=1}^{n} d_W(\widetilde{f}_i, f_i) - \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} d_W(\widetilde{f}_i, f_i)\right)\right| + \left|\mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} d_W(\widetilde{f}_i, f_i)\right)\right| \xrightarrow{p} 0 \;.$$

It follows from the weak law of large numbers that for a fixed f

$$M_n(f) \xrightarrow{p} M(f).$$

For the fixed $\epsilon > 0$, because of the boundedness of $\mathcal{W}(\mathscr{X})$ from Condition 3.2, there exist finite ϵ-cover $\gamma_1, \ldots, \gamma_l$ satisfying that for any $f \in \mathcal{W}(\mathscr{X})$, there exists j such that $d_W(f, \gamma_j) \leq \epsilon$, where l depends on the size of ϵ. We decompose $M_n(f) - M(f)$ into three parts:

$$\begin{aligned}
|M_n(f) - M(f)| &\leq |M_n(f) - M(\gamma_j)| + |M_n(\gamma_j) - M(\gamma_j)| + |M_n(\gamma_j) - M(f)| \\
&\leq 2 d_W(f, f_j) \cdot d_{\sup} + |M_n(f_j) - M(f_j)| \\
&\leq 2 d_{\sup} \epsilon + |M_n(f_j) - M(f_j)|
\end{aligned}$$

Thus, it yields

$$\sup_{\mu \in W_2(\mathscr{X})} |M_n(f) - M(f)| \leq 2 d_{\sup} \epsilon.$$

Since ϵ is an arbitrary, we conclude

$$\sup_{f \in W_2(\mathscr{X})} |\widehat{M}_n(f) - M(f)| \xrightarrow{p} 0.$$

The uniform convergence of $\widehat{M}_n(\mu)$ to $M(f)$, and the continuity of $M(f)$ on the compact set $W_2(\mathscr{X})$ lead to $d_W(\widehat{f}_\oplus, f_\oplus) \xrightarrow{p} 0$, where $\widehat{f}_\oplus$ and $f_\oplus$ are minimums for $\widehat{M}_n(f)$ and $M(f)$, respectively. This completes the proof for second statement in Theorem 3.1.

It can be shown that

$$\sqrt{n}(\widehat{F}_\oplus^{-1} - F_\oplus^{-1}) = \sqrt{n}(\widehat{F}_\oplus^{-1} - \bar{F}_\oplus^{-1}) + \sqrt{n}(\bar{F}_\oplus^{-1} - F_\oplus^{-1}) = I + II.$$

For $I = \sqrt{n}(\widehat{F}_\oplus^{-1} - \bar{F}_\oplus^{-1})$, we observe that

$$\begin{aligned}
\|\widehat{F}_\oplus^{-1} - \bar{F}_\oplus^{-1}\| &= \|\frac{1}{\sqrt{n}}\sum_{i=1}^n \widehat{F}_i^{-1} - \frac{1}{\sqrt{n}}\sum_{i=1}^n F_i^{-1}\| \\
&\leq \frac{1}{\sqrt{n}}\sum_{i=1}^n \|\widehat{F}_i^{-1} - F_i^{-1}\| = \frac{1}{n}\sum_{i=1}^n d_W(\widehat{f}_i, f_i).
\end{aligned}$$

It follows from the triangle inequality and the upper bound aforementioned that

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^n d_W(\widehat{f}_i, f_i) &\leq \frac{1}{\sqrt{n}}\sum_{i=1}^n d_W(\widehat{f}_i, \widetilde{f}_i) + \frac{1}{\sqrt{n}}\sum_{i=1}^n d_W(\widetilde{f}_i, f_i) \\
&\leq \frac{1}{\sqrt{n}}\sum_{i=1}^n d_W(\widetilde{f}_i, f_i) + \sqrt{C_K n}\sigma_n.
\end{aligned}$$

The markov inequality yields

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^n d_W(\widetilde{f}_i, f_i) > \epsilon\right] \leq \frac{\sum_{i=1}^n \mathbb{E}d_W(\widetilde{f}_i, f_i)}{\epsilon n} \leq \frac{\mathbb{E}d_W(\widetilde{f}_i, f_i)}{\epsilon}.$$

We consider the upper bound for the expectation of the Wasserstein distance between $\widetilde{f}_i$ and f_i. It can be easily shown that

$$\begin{aligned}
\mathbb{E}d_W(\widetilde{f}_i, f_i) &\leq d_{\sup} \cdot d_{W_1}(\widetilde{f}_i, f_i) \\
&\leq d_{\sup} \cdot \int_{\mathscr{X}} \mathbb{E}\left|F_1(x) - \frac{1}{N_1}\sum_{k=1}^{\eta}\sum_{j=1}^{N_1^{(\eta)}} \delta_{X_{1j}^{(k)}}(x)\right| dx \\
&\equiv d_{\sup} \cdot \int_{\mathscr{X}} \mathbb{E}|Y(x)| dx, \qquad (5)
\end{aligned}$$

where for the first inequality holds by the holder inequality, and the second inequality holds by the 1-Wasserstein distance property with the application of Fubini's Theorem. Note that when the total number of the observed points N_1 is a fixed case along with the Poisson distribution assumption, the random variable $Y(x)$ is defined as $Y(x) = X(x)/N_1 - F_1(x)$, for which $X(x)$ follows a Binomial distribution, i.e., $X(x) \sim \text{Binomial}(N_1, F_1(x))$. Thereby, we have the conditional expectation of $Y(x)$ given on N_1 as $\mathbb{E}Y(x)|N_1 = 0$ and the second moment for it as

$$\mathbb{E}Y(x)^2|N_1 = \frac{N_1 F_1(x)(1 - F_1(x))}{N_1^2} \leq \frac{1}{4N_1}.$$

Assuming that N_1 follows the Poisson distribution, i.e., $N_1 \sim \text{Poisson}(\eta)$, we compute the unconditional second moment as the following

$$\mathbb{E}Y(x)^2 \leq \mathbb{E}\frac{1}{N_1} = \sum_{j=1}^{\infty} \frac{1}{j} e^{-\eta} \frac{\eta^j}{j\,!} \leq \sum_{j=1}^{\infty} 2e^{-\eta} \frac{\eta^j}{(j+1)\,!} = 2\kappa^{-1} \sum_{j=1}^{\infty} e^{-\eta} \frac{\eta^{j+1}}{(j+1)\,!} \leq \frac{2}{\eta}$$

By applying this result into (5), we obtain

$$\mathbb{E}d_W(\widetilde{f}_i, f_i) \leq d_{\text{sup}}^{1/2} \int_{\mathscr{X}} \mathbb{E}|Y(x)|dx \leq d_{\text{sup}}^{1/2} \int_{\mathscr{X}} \mathbb{E}|Y^2(x)|^{1/2} dx \leq \frac{C}{\sqrt{\eta}}$$

Hence, we obtain the convergence rate of this quantity to zero given by

$$\frac{1}{n} \sum_{i=1}^{n} d_W(\widetilde{f}_i, f_i) = o_p(\frac{1}{\sqrt{\eta}}),$$

which implies

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} d_W(\widehat{f}_i, f_i) = o_p(\frac{1}{\sqrt{\eta n}}) + o_p(\sqrt{n}\sigma_n) \tag{6}$$

Since the Conditions 3.1-3.5 yields that the quantity in (6) converges to zero, we have

$$\sqrt{n}(\widehat{F}_{\oplus}^{-1} - \bar{F}_{\oplus}^{-1}) = o_p(1). \tag{7}$$

For $II = \sqrt{n}(\bar{F}_{\oplus}^{-1} - F_{\oplus}^{-1})$, we observe that

$$\sqrt{n}\left(\frac{1}{n} \sum_{i=1}^{n} F_i^{-1}(s) - F_{\oplus}^{-1}(s)\right)$$

Since $F_1, \dots, F_n$ are independent and identically distributed, and the second moment of the term II, since the Condition 3.2, is the finite value,

$$\mathbb{E}\left[\int_0^1 \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n F_i^{-1}(s) - F_\oplus^{-1}(s)\right)^2 ds\right] = \mathrm{Var}_\oplus[F] < \infty.$$

the second term II converges to a Gaussian process $Z(s)$ in $L_2([0,1])$ with the covariance function given by $\Sigma_0(u,v) = \mathbb{E}[Z(u)Z(v)]$ as n goes to infinity.
Next, by the continuous mapping theorem,

$$\begin{aligned}\sqrt{n}(\frac{1}{n}\sum_{i=1}^n \mathbf{t}_{\mu_\oplus}^{\widehat{\mu_i}} - \mathbf{i}) &= \sqrt{n}(\widehat{F}_\oplus^{-1} - F_\oplus^{-1}) \circ F_\oplus \\ &= \sqrt{n}(\widehat{F}_\oplus^{-1} - \bar{F}_\oplus^{-1}) \circ F_\oplus + \sqrt{n}(\bar{F}_\oplus^{-1} - F_\oplus^{-1}) \circ F_\oplus \end{aligned} \tag{8}$$

The first term of the right-hand side in (8) is asymptotically negligible as the result of (7). For the second term of the right-hand side in (8), we have

$$\begin{aligned}\sqrt{n}(\bar{F}_\oplus^{-1} - F_\oplus^{-1}) \circ F_\oplus &= \sqrt{n}(\frac{1}{n}\sum_{i=1}^n F_i^{-1} \circ F_\oplus - F_\oplus^{-1} \circ F_\oplus) \\ &= \sqrt{n}(\frac{1}{n}\sum_{i=1}^n \mathbf{t}_{\mu_\oplus}^{\mu_i} - \mathbf{i}),\end{aligned}$$

which converges to a Gaussian process in $L_2(\mathscr{X})$, i.e., $\sqrt{n}(\frac{1}{n}\sum_{i=1}^n \mathbf{t}_{\mu_\oplus}^{\mu_i} - \mathbf{i}) \xrightarrow{\mathscr{D}} Z \circ F_\oplus$ with the covariance function given by

$$\begin{aligned}\Sigma_{\mu_\oplus}(u,v) &= \mathbb{E}[(\mu^{-1} - \mu_\oplus^{-1}) \circ \mu_\oplus(u) \cdot (\mu^{-1} - \mu_\oplus^{-1}) \circ \mu_\oplus(v)] \\ &= \mathbb{E}[(\mathbf{t}_{\mu_\oplus}^{\mu}(u) - u)(\mathbf{t}_{\mu_\oplus}^{\mu}(v) - v)]\end{aligned}$$

This completes the proof for Theorem 3.1. □

Proof of Theorem 3.2 Under the null hypothesis with the given Conditions, recall that $(\nu_\oplus^{-1} - \mu_\oplus^{-1}) \circ \mu_\oplus$ is zero function for all $y \in \mathscr{Y}$. By the asymptotic expansion the proposed statistics, $U_{n,m}$ is represented by

$$\begin{aligned}&\sqrt{\frac{nm}{n+m}}\left(\frac{1}{m}\sum_{j=1}^m \widehat{\nu}_j^{-1} - \frac{1}{n}\sum_{i=1}^n \widehat{\mu}_i^{-1}\right) \circ \left(\frac{1}{n}\sum_{i=1}^n \widehat{\mu}_i^{-1}\right)^{-1} \\ &\quad = \sqrt{\frac{nm}{n+m}}\left(\frac{1}{m}\sum_{j=1}^m \widehat{\nu}_j^{-1} - \frac{1}{n}\sum_{i=1}^n \widehat{\mu}_i^{-1}\right) \circ \mu_\oplus + o_p(1) \\ &\quad = \sqrt{\frac{nm}{n+m}}\left(\frac{1}{m}\sum_{j=1}^m \widehat{\nu}_j^{-1} - \nu_\oplus^{-1} - \frac{1}{n}\sum_{i=1}^n \widehat{\mu}_i^{-1} + \mu_\oplus^{-1} + \nu_\oplus^{-1} - \mu_\oplus^{-1}\right) \circ \mu_\oplus + o_p(1)\end{aligned}$$

$$= \sqrt{\frac{m}{n+m}}\sqrt{m}\left(\frac{1}{m}\sum_{j=1}^{m}\widehat{\nu}_j^{-1} - \nu_\oplus^{-1}\right)\circ\mu_\oplus + \sqrt{\frac{n}{n+m}}\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n}\widehat{\mu}_j^{-1} - \mu_\oplus^{-1}\right)\circ\mu_\oplus + o_p(1)$$

$$\xrightarrow{\mathscr{D}} \sqrt{\theta}(\mathbf{t}^{\nu}_{\mu_\oplus} - \nu_\oplus^{-1}\mu_\oplus) + \sqrt{(1-\theta)}(\mathbf{t}^{\mu}_{\mu_\oplus} - \mathbf{i}),$$

where the first equality holds from the result of (7) and the Slustky, and Theorem 3.1 are applied to obtain the asymptotic distribution. Denoting Γ as the asymptotic random function, i.e., $\Gamma = \sqrt{(1-\theta)}Z\circ F_\oplus + \sqrt{\theta}Z\circ F_\oplus$, the Γ follows the Gaussian process with the mean $\mathbb{E}(\Gamma(y)) = \mu(y)$ for all $y \in \mathscr{Y}$ and covariance function,

$$\begin{aligned}\Sigma(y,y') &= (1-\theta)\mathbb{E}[(\mathbf{t}^{\mu}_{\mu_\oplus}(y) - y)(\mathbf{t}^{\mu}_{\mu_\oplus}(y') - y')] + \theta\mathbb{E}[(\mathbf{t}^{\nu}_{\mu_\oplus}(y) - y)(\mathbf{t}^{\nu}_{\mu_\oplus}(y') - y')] \\ &= (1-\theta)\Sigma_{\mu_\oplus}(y,y') + \theta\Lambda(\Sigma_{\nu_\oplus}(y,y')).\end{aligned}$$

It follows from $\Sigma(y,y')$ has finite trace, i.e., $\mathrm{tr}(\Sigma) = \int_{\mathscr{Y}}\Sigma(y,y)d\mu_\oplus(y) < \infty$, it allows the singluar value decomposition (SVD) $\Sigma(y,y') = \sum_{k=1}^{\infty}\lambda_k\phi_k(y)\phi_k(y')$, where λ_k and $\phi_k(\cdot)$ are the kth eigenvalue and function for the covariance function such that $\int\phi_k(y)^2 d\mu_\oplus(y) = 1$ for all k and $\int\phi_k(y)\phi_h(y)d\mu_\oplus(y) = 0$ for $k \neq h$. By Karhunen–Loève expansion in $L(\mu_\oplus)$,

$$\Gamma(y) = \mu(y) + \sum_{k=1}^{\infty}\langle\Gamma,\phi_k\rangle_{\mu_\oplus}\phi_k(y)$$

where $\mathbb{E}\langle\Gamma,\phi_k\rangle_{\mu_\oplus} = 0$ and $\mathbb{E}\langle\Gamma,\phi_k\rangle^2_{\mu_\oplus} = \mathbb{E}\langle\Gamma,\phi_k\rangle_{\mu_\oplus}\langle\Gamma,\phi_k\rangle_{\mu_\oplus} = \langle\mathbb{E}\langle\Gamma,\phi_k\rangle_{\mu_\oplus}\Gamma,\phi_k\rangle_{\mu_\oplus} = \langle\lambda_k\phi_k,\phi_k\rangle_{\mu_\oplus} = \lambda_k$. It is easy to see $\mathbb{E}\Gamma(y) = \mu(y)$ and $\mathrm{cov}(\Gamma(y),\Gamma(y')) = \Sigma(y,y')$

$$\begin{aligned}U_{n,m} &= \int\mu^2(y)d\mu_\oplus(y) + 2\sum_{k=1}^{\infty}\langle\Gamma,\phi_k\rangle\int\mu(y)\phi_k(y)d\mu_\oplus(y) + \int(\sum_{k=1}^{\infty}\langle\Gamma,\phi_k\rangle\phi_k(y))^2 d\mu_\oplus(y) + o_p(1) \\ &= \sum_{k=1}^{\infty}\langle\mu,\phi_k\rangle^2 + 2\sum_{k=1}^{\infty}\langle\Gamma,\phi_k\rangle\langle\mu,\phi_k\rangle + \sum_{k=1}^{\infty}\langle\Gamma,\phi_k\rangle^2 + o_p(1) \\ &= \sum_{k=1}^{K}\Big(\langle\mu,\phi_k\rangle + \langle\Gamma,\phi_k\rangle\Big)^2 + \sum_{k=K+1}^{\infty}\langle\mu,\phi_k\rangle^2 + o_p(1) \\ &= \sum_{k=1}^{K}\lambda_k\left[\frac{(\langle\mu,\phi_k\rangle + \langle\Gamma,\phi_k\rangle)^2}{\lambda_k}\right] + \sum_{k=K+1}^{\infty}\langle\mu,\phi_k\rangle^2 + o_p(1)\end{aligned}$$

Under the Gaussian property, we have $(\langle\mu,\phi_k\rangle + \langle\Gamma,\phi_k\rangle)/\sqrt{\lambda_k} \sim N(\langle\mu,\phi_k\rangle, 1)$, which results in

$$U_{n,m} \xrightarrow{\mathscr{D}} \sum_{k=1}^{K} w_k\chi^2_{k,\delta_k,(1)} + \sum_{k=K+1}^{\infty}\langle\mu,\phi_k\rangle^2$$

Hence, under H_0, we have $U_{n,m} \xrightarrow{\mathscr{D}} \sum_{k=1}^{K} w_k\chi^2_{k,(1)}$. This completes the proof for the first statement.

For the notational convenience, denote

$$\widehat{\Sigma}_{\widehat{\mu}_\oplus} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{t}_{\widehat{\mu}_\oplus}^{\widehat{\mu}_i} - \mathbf{i})(\mathbf{t}_{\widehat{\mu}_\oplus}^{\widehat{\mu}_i} - \mathbf{i}), \quad \widehat{\Lambda}(\widehat{\Sigma}_{\widehat{\nu}_\oplus}) = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{t}_{\widehat{\nu}_\oplus}^{\widehat{\nu}_i} - \widehat{\nu}_\oplus^{-1}(\widehat{\mu}_\oplus))(\mathbf{t}_{\widehat{\nu}_\oplus}^{\widehat{\nu}_i} - \widehat{\nu}_\oplus^{-1}(\widehat{\mu}_\oplus))$$
$$\Sigma_{\mu_\oplus} = \mathbb{E}[(\mathbf{t}_{\mu_\oplus}^{\mu_i} - \mathbf{i})(\mathbf{t}_{\mu_\oplus}^{\mu_i} - \mathbf{i})], \quad \Lambda(\Sigma_{\nu_\oplus}) = \mathbb{E}[(\mathbf{t}_{\nu_\oplus}^{\nu_i} - \nu_\oplus^{-1}(\mu_\oplus))(\mathbf{t}_{\nu_\oplus}^{\nu_i} - \nu_\oplus^{-1}(\mu_\oplus))]$$

and $\widetilde{\Sigma}[(y_1, y_1'), (y_2, y_2')] = \mathbb{E}[(\mathbf{t}_{\mu_\oplus}^{\mu}(y_1) - y_1)(\mathbf{t}_{\mu_\oplus}^{\mu}(y_1') - y_1')(\mathbf{t}_{\nu_\oplus}^{\nu}(y_2) - y_2)(\mathbf{t}_{\nu_\oplus}^{\nu}(y_2') - y_2')] - \mathbb{E}[(\mathbf{t}_{\mu_\oplus}^{\mu}(y_1) - y_1)(\mathbf{t}_{\mu_\oplus}^{\mu}(y_1') - y_1')]\mathbb{E}[(\mathbf{t}_{\nu_\oplus}^{\nu}(y_2) - y_2)(\mathbf{t}_{\nu_\oplus}^{\nu}(y_2') - y_2')]$. Under the null hypothesis, denote $\Sigma_{\mu_\oplus}^0$ as $\Sigma_{\mu_\oplus} = \Lambda(\Sigma_{\nu_\oplus}) \equiv \Sigma_{\mu_\oplus}^0$. We first show a Gaussian process for the discrepancy between two covariance kernels based on the the asymptotic expansion

$$\begin{aligned}
&\sqrt{\frac{nm}{n+m}}\Big(\widehat{\Sigma}_{\widehat{\mu}_\oplus}(y,y') - \widehat{\Lambda}(\widehat{\Sigma}_{\widehat{\nu}_\oplus})(y,y')\Big) \\
&\quad = \sqrt{\frac{nm}{n+m}}\Big(\widehat{\Sigma}_{\widehat{\mu}_\oplus}(y,y') - \Sigma_{\mu_\oplus}^0(y,y')\Big) - \sqrt{\frac{nm}{n+m}}\Big(\widehat{\Lambda}(\widehat{\Sigma}_{\widehat{\nu}_\oplus})(y,y') - \Sigma_{\mu_\oplus}^0(y,y')\Big) \\
&\quad = \sqrt{1-\theta}\sqrt{n}\Big(\widehat{\Sigma}_{\widehat{\mu}_\oplus}(y,y') - \Sigma_{\mu_\oplus}^0(y,y')\Big) - \sqrt{\theta}\sqrt{m}\Big(\widehat{\Lambda}(\widehat{\Sigma}_{\widehat{\nu}_\oplus})(y,y') - \Sigma_{\mu_\oplus}^0(y,y')\Big) + o_p(1) \\
&\quad \xrightarrow{\mathscr{D}} \sqrt{(1-\theta)}B_1(y,y') - \sqrt{\theta}B_2(y,y')
\end{aligned}$$

where we use the arguments applied in Theorem 3.1.
Consequently, we have

$$\sqrt{n}\Big(\widehat{\Sigma}_{\widehat{\mu}_\oplus}(y,y') - \Sigma_{\mu_\oplus}^0(y,y')\Big) \xrightarrow{\mathscr{D}} B_1(y,y') \quad \text{and} \quad \sqrt{m}\Big(\widehat{\Lambda}(\widehat{\Sigma}_{\widehat{\nu}_\oplus})(y,y') - \Sigma_{\mu_\oplus}^0(y,y')\Big) \xrightarrow{\mathscr{D}} B_2(y,y'),$$

respectively, where $B(y,y')$ is a Gaussian process with the covariance process, $\widetilde{\Sigma}[(y_1, y_1'), (y_2, y_2')]$. Note that the independence of $\widehat{\Sigma}_{\widehat{\mu}_\oplus}$ and $\widehat{\Lambda}(\widehat{\Sigma}_{\widehat{\nu}_\oplus})$ and the unit sphere condition of the sample proportion, $\sqrt{1-\theta}^2 + \sqrt{\theta}^2 = 1$, we obtain the following

$$\sqrt{\frac{nm}{n+m}}\Big(\widehat{\Sigma}_{\widehat{\mu}_\oplus}(y,y') - \widehat{\Lambda}(\widehat{\Sigma}_{\widehat{\nu}_\oplus})(y,y')\Big) \xrightarrow{\mathscr{D}} B(y,y') + o_p(1).$$

By the asymptotic expansion the proposed statistics, $V_{n,m}$ is represented by

$$\begin{aligned}
V_{n,m} &= \int_{\mathscr{Y}}\int_{\mathscr{Y}} \frac{nm}{n+m}\Big(\widehat{\Sigma}_{\widehat{\mu}_\oplus}(y,y') - \widehat{\Lambda}(\widehat{\Sigma}_{\widehat{\nu}_\oplus})(y,y')\Big)^2 d\mu_\oplus(y)d\mu_\oplus(y') \\
&= \int_{\mathscr{Y}}\int_{\mathscr{Y}} \frac{nm}{n+m}\Big(\widehat{\Sigma}_{\widehat{\mu}_\oplus}(y,y') - \Sigma_{\mu_\oplus}^0(y,y')\Big)^2 + \frac{nm}{n+m}\Big(\widehat{\Lambda}(\widehat{\Sigma}_{\widehat{\nu}_\oplus})(y,y') - \Sigma_{\mu_\oplus}^0(y,y')\Big)^2 \\
&\qquad - 2\frac{nm}{n+m}\Big(\widehat{\Sigma}_{\widehat{\mu}_\oplus}(y,y') - \Sigma_{\mu_\oplus}^0(y,y')\Big)\Big(\widehat{\Lambda}(\widehat{\Sigma}_{\widehat{\nu}_\oplus})(y,y') - \Sigma_{\mu_\oplus}^0(y,y')\Big)d\mu_\oplus(y)d\mu_\oplus(y') \\
&= \int_{\mathscr{Y}}\int_{\mathscr{Y}} \Big(\sqrt{n}(\widehat{\Sigma}_{\widehat{\mu}_\oplus}(y,y') - \Sigma_{\mu_\oplus}^0(y,y')), \sqrt{m}(\widehat{\Lambda}(\widehat{\Sigma}_{\widehat{\nu}_\oplus})(y,y') - \Sigma_{\mu_\oplus}^0(y,y'))\Big)\Big(I - P\Big)
\end{aligned}$$

$$\times \left(\sqrt{n}(\widehat{\Sigma}_{\widehat{\mu}_\oplus}(y,y') - \Sigma^0_{\mu_\oplus}(y,y')), \sqrt{m}(\widehat{\Lambda}(\widehat{\Sigma}_{\widehat{\nu}_\oplus})(y,y') - \Sigma^0_{\mu_\oplus}(y,y')) \right)^T d\mu_\oplus(y) d\mu_\oplus(y'),$$

where I is the identity matrix in $\mathbb{R}^2$, $P = JJ^T/(n+m)$, and $J = (n,m)^T$. We see that $(\sqrt{n_1}(\widehat{\Sigma}_{\widehat{\mu}_\oplus}(y,y') - \Sigma^0_{\mu_\oplus}(y,y')), \sqrt{n_2}(\widehat{\Lambda}(\widehat{\Sigma}_{\widehat{\nu}_\oplus})(y,y') - \Sigma^0_{\mu_\oplus}(y,y')))$ converges to a bivariate Gaussian process $(B_1(y,y'), B_2(y,y'))$ and $I-P$ is a symmetric and idempotent matrix, where $\mathrm{tr}(I-P)=1$. Therefore, by the theorem of the quadratic form for the Gaussian process, there exists a Gaussian process such that

$$\begin{aligned}
V_{n,m} &= \sum_{r=1}^{\mathrm{rank}(I-P)} \int_0^1 \int_0^1 W_r(u,v)^T W_r(u,v) du dv \\
&= \sum_{r=1}^{\mathrm{rank}(I-P)} \left(\sum_{k=1}^{K} w_k \chi^2_{k,\delta_k,(1)} + \sum_{k=K+1}^{\infty} \langle \mu, \phi_k \rangle^2 \right) \\
&= \sum_{k=1}^{K} w_k \chi^2_{k,\delta_k,(\mathrm{rank}(I-P))} + \sum_{r=1}^{\mathrm{rank}(I-P)} \sum_{k=K+1}^{\infty} \langle \mu, \phi_k \rangle^2,
\end{aligned}$$

where the process $W_r(u,v)$ is defined as $W_r(u,v) = (I-P)(B_F(u,v), B_G(u,v))^T$. Hence, we conclude

$$V_{n,m} \xrightarrow{p} \sum_{k=1}^{K} w_k \chi^2_{k,(\mathrm{rank}(I-P))},$$

where ω_k is the kth eigenvalue of $\widetilde{\Sigma}[(u_1,v_1),(u_2,v_2)] = \mathbb{E}[(\mathbf{t}^{\mu}_{\mu_\oplus}(u_1) - u_1)(\mathbf{t}^{\mu}_{\mu_\oplus}(v_1) - v_1)(\mathbf{t}^{\nu}_{\nu_\oplus}(u_2) - u_2)(\mathbf{t}^{\nu}_{\nu_\oplus}(v_2) - v_2)] - \mathbb{E}[(\mathbf{t}^{\mu}_{\mu_\oplus}(u_1) - u_1)(\mathbf{t}^{\mu}_{\mu_\oplus}(v_1) - v_1)]\mathbb{E}[(\mathbf{t}^{\nu}_{\nu_\oplus}(u_2) - u_2)(\mathbf{t}^{\nu}_{\nu_\oplus}(v_2) - v_2)]$ under H_0. This completes the proof for the second statement. □

Proof of Theorem 3.3 We only show the power performance for the variance difference test while it is not difficulty in showing the power performance for the mean difference test. Let us define $R_{n,m,\alpha} = \{V_{n,m} > q_\alpha\}$, and $H_{n,m} = \{\Delta : \Delta \geq a_{n,m}\}$, $\beta_{H_{n,m}} = \inf_{V \in H_{n,m}} \mathbb{P}(R_{n,m,\alpha})$ as the rejection region, the local alternative hypothesis, and the power for the test at the significance level of α. It can be shown that

$$\begin{aligned}
V_{n,m} &= \int_{\mathscr{Y}} \int_{\mathscr{Y}} \frac{nm}{n+m} \left(\widehat{\Sigma}_{\widehat{\mu}_\oplus}(y,y') - \widehat{\Lambda}(\widehat{\Sigma}_{\widehat{\nu}_\oplus})(y,y') \right)^2 d\mu_\oplus(y) d\mu_\oplus(y') \\
&= \int_{\mathscr{Y}} \int_{\mathscr{Y}} \frac{nm}{n+m} \Bigg[\left\{ \left(\widehat{\Sigma}_{\widehat{\mu}_\oplus}(y,y') - \Sigma_{\mu_\oplus}(y,y') \right) - \left(\widehat{\Lambda}(\widehat{\Sigma}_{\widehat{\nu}_\oplus})(y,y') - \Lambda(\Sigma_{\nu_\oplus})(y,y') \right) \right\}^2 \\
&\quad + 2\left(\Sigma_{\mu_\oplus}(y,y') - \Lambda(\Sigma_{\nu_\oplus})(y,y') \right) \left\{ \left(\widehat{\Sigma}_{\widehat{\mu}_\oplus}(y,y') - \Sigma_{\mu_\oplus}(y,y') \right) - \left(\widehat{\Lambda}(\widehat{\Sigma}_{\widehat{\nu}_\oplus})(y,y') - \Lambda(\Sigma_{\nu_\oplus})(y,y') \right) \right\} \\
&\quad + \left(\Sigma_{\mu_\oplus}(y,y') - \Lambda(\Sigma_{\nu_\oplus})(y,y') \right)^2 \Bigg] d\mu_\oplus(y) d\mu_\oplus(y') \\
&= \frac{nm}{n+m} \int_{\mathscr{Y}} \int_{\mathscr{Y}} (\widetilde{V}_{n,m} + \Delta_{n,m}) d\mu_\oplus(y) d\mu_\oplus(y'),
\end{aligned}$$

where denote $\widetilde{V}_{n,m}(y,y') = [(\widehat{\Sigma}_{\widehat{\mu_\oplus}}(y,y') - \Sigma_{\mu_\oplus}(y,y')) - (\widehat{\Lambda}(\widehat{\Sigma}_{\widehat{\nu_\oplus}})(y,y') - \Lambda(\Sigma_{\nu_\oplus})(y,y')]^2$ and $\Delta_{n,m}(y,y') = 2(\Sigma_{\mu_\oplus}(y,y') - \Lambda(\Sigma_{\nu_\oplus})(y,y'))\{(\widehat{\Sigma}_{\widehat{\mu_\oplus}}(y,y') - \Sigma_{\mu_\oplus}(y,y')) - (\widehat{\Lambda}(\widehat{\Sigma}_{\widehat{\nu_\oplus}})(y,y') - \Lambda(\Sigma_{\nu_\oplus})(y,y'))\} + [\Sigma_{\mu_\oplus}(y,y') - \Lambda(\Sigma_{\nu_\oplus})(y,y')]^2$. Under H_1, we observe

$$\Delta_n(y,y') \xrightarrow{p} \Big(\Sigma_{\mu_\oplus}(y,y') - \Lambda(\Sigma_{\nu_\oplus})(y,y')\Big)^2 = \Delta(y,y').$$

Let q_α be $(1-\alpha)$the quantile of the weighted sum of the chi-squared distribution, i.e., $\sum_{k=1}^K w_k \chi^2_{k,(\mathrm{rank}(I-P))}$. For the rejection region, we obtain the probability given by

$$\begin{aligned}
\mathbb{P}[R_{n,m,\alpha}] &= \mathbb{P}\Big[\int_{\mathscr{Y}}\int_{\mathscr{Y}}\Big(\widetilde{V}_{n,m}(y,y') + \Delta_{n,m}(y,y')\Big)d\mu_\oplus(y)d\mu_\oplus(y') > \frac{n+m}{nm}q_\alpha\Big] \\
&\geq \mathbb{P}\Big[\int_{\mathscr{Y}}\int_{\mathscr{Y}}\Big(\widetilde{V}_{n,m}(y,y')d\mu_\oplus(y)d\mu_\oplus(y') + \frac{1}{2}\Delta_{n,m}(y,y')\Big) > \frac{n+m}{nm}q_\alpha\Big] \\
&\geq \mathbb{P}\Big[\frac{nm}{n+m}\int_{\mathscr{Y}}\int_{\mathscr{Y}}\widetilde{V}_{n,m}(y,y')d\mu_\oplus(y)d\mu_\oplus(y') > q_\alpha - \frac{nm}{2(n+m)}\int_{\mathscr{Y}}\int_{\mathscr{Y}}\Delta_{n,m}(y,y')d\mu_\oplus(y)d\mu_\oplus(y')\Big].
\end{aligned}$$

Thus, we have the power performance given by

$$\begin{aligned}
\lim_{n,m\to\infty}\beta_{H_{n,m}} &= \lim_{n,m\to\infty}\mathbb{P}[R_{n,m,\alpha}] \\
&\geq \mathbb{P}\Big[\frac{nm}{n+m}\int_{\mathscr{Y}}\int_{\mathscr{Y}}\widetilde{V}_{n,m}(y,y')d\mu_\oplus(y)d\mu_\oplus(y') > q_\alpha - \frac{nm}{2(n+m)}\int_{\mathscr{Y}}\int_{\mathscr{Y}}\Delta_{n,m}(y,y')d\mu_\oplus(y)d\mu_\oplus(y')\Big] \\
&\geq \mathbb{P}\Big[\frac{nm}{n+m}\int_{\mathscr{Y}}\int_{\mathscr{Y}}\widetilde{V}_{n,m}(y,y')d\mu_\oplus(y)d\mu_\oplus(y') > q_\alpha - \frac{\theta(1-\theta)}{2}\int_{\mathscr{Y}}\int_{\mathscr{Y}}(n+m)\Delta(y,y')d\mu_\oplus(y)d\mu_\oplus(y')\Big] \\
&\geq \mathbb{P}\Big[\frac{nm}{n+m}\int_{\mathscr{Y}}\int_{\mathscr{Y}}\widetilde{V}_{n,m}(y,y')d\mu_\oplus(y)d\mu_\oplus(y') > q_\alpha - \frac{\theta(1-\theta)}{2}(n+m)a_n\Big]
\end{aligned}$$

Since we already have the result from Theorem 3.2,

$$\frac{nm}{n+m}\int_{\mathscr{Y}}\int_{\mathscr{Y}}\widetilde{V}_{n,m}(y,y')d\mu_\oplus(y)d\mu_\oplus(y') \xrightarrow{\mathscr{D}} \sum_{k=1}^K w_k\chi^2_{k,(\mathrm{rank}(I-P))},$$

if we have $a_{n,m}$ such that $(n+m)a_n \to \infty$, then we conclude $\lim_{n,m\to\infty}\beta_{H_n} = 1$, for all $\alpha > 0$. This completes the proof for Theorem 3.3. □

References

Karr, A.: Point Processes and Their Statistical Inference. CRC Press, Routledge (2017)