Consistency part for Fréchet change point detection

1 Preliminary

Assumption 1. *Given a heterogeneous sequence of independent observations* $\{Z_1, \ldots, Z_T\}$ *taking values in general metric space* (Ω, d) *with a change point denoted by* τ*, we let* $Z_1, \ldots, Z_{\lfloor \gamma T \rfloor} \overset{i.i.d.}{\sim} F_L$ *and* $Z_{\lfloor \gamma T \rfloor + 1}, \ldots, Z_T \overset{i.i.d.}{\sim} F_R$*, where* $\gamma = \frac{\tau}{T}$*. For ease of notation, we denote* $\mu_{XX} = E(d^\alpha(X, X'))$*,* $\mu_{XY} = E(d^\alpha(X, Y))$*,* $\mu_{YY} = E(d^\alpha(Y, Y'))$ *where* $X, X' \overset{i.i.d.}{\sim} F_L$*,* $Y, Y' \overset{i.i.d.}{\sim} F_R$*,* X, X', Y, Y' *are mutually independent, and* $d^\alpha(.)$ *is strictly negative definite kernel. Let* $\{\delta_T\}$ *be a sequence of positive numbers with property of* $\delta_T \to 0$ *and* $T\delta_T \to \infty$*, as* $T \to \infty$*.*

Lemma 1. *Under Assumption 1,*

$$\lim_{T\to\infty} \left\{ \sup_{\gamma \in [\delta_T, 1-\delta_T]} \left| \binom{T}{2}^{-1} \sum_{1 \le i < j \le T} d^\alpha(Z_i, Z_j) - \left(\gamma^2 \mu_{XX} + (1-\gamma)^2 \mu_{YY} + 2\gamma(1-\gamma)\mu_{XY}\right) \right| \right\} = 0, a.s.$$

Proof. To show the convergence almost surely uniformly in γ, we split $\binom{T}{2}^{-1} \sum_{1 \le i < j \le T} d^\alpha(Z_i, Z_j)$ into three terms: $\binom{\lfloor \gamma T \rfloor}{2}^{-1} \sum_{1 \le i < j \le \lfloor \gamma T \rfloor} d^\alpha(Z_i, Z_j)$, $\binom{T - \lfloor \gamma T \rfloor}{2}^{-1} \sum_{\lfloor \gamma T \rfloor + 1 \le i < j \le T} d^\alpha(Z_i, Z_j)$, and $2\binom{T - \lfloor \gamma T \rfloor}{1}^{-1} \binom{\lfloor \gamma T \rfloor}{1}^{-1} \sum_{i=1}^{\lfloor \gamma T \rfloor} \sum_{j=\lfloor \gamma T \rfloor + 1}^{T} d^\alpha(Z_i, Z_j)$, then prove the three terms converge to $\gamma^2 \mu_{XX}$, $(1-\gamma)^2 \mu_{YY}$, and $2\gamma(1-\gamma)\mu_{XY}$ almost surely uniformly in γ respectively. By continuity theorem, Lemma 1 is proved.

The detail of proof is shown in Appendix. □

Lemma 1 is saying that in a sequence of length 1 with two distributions truncated at γ, the average pairwise deviation $\binom{T}{2}^{-1} \sum_{1 \le i < j \le T} d^\alpha(Z_i, Z_j)$ converges almost surely to a function of distributions mean in a form of $\gamma^2 \mu_{XX} + (1-\gamma)^2 \mu_{YY} + 2\gamma(1-\gamma)\mu_{XY}$, uniformly in γ. This result will be frequently used when we prove the test statistics $\hat{Q}$ is approximate to a scaled energy distance in both single change point detection case and multiple change point detection case.

2 Single Change Point

Under Assumption 1, since distributions F_L and F_R are truncated at γ, i.e, γ is the TRUE fraction of change point in the sequence, we can calculate a scaled energy distance denoted as $\xi(\tilde{\gamma})$ for any candidate $\tilde{\gamma} \in [\delta_T, 1-\delta_T]$ in terms of γ. Thus, we propose the following Proposition 1 that shows the scaled energy distance $\xi(\tilde{\gamma})$ is maximized when $\tilde{\gamma}$ takes value of γ.

Proposition 1. *Under Assumption 1,*

$$\underset{\tilde{\gamma}\in[\delta_T,1-\delta_T]}{\operatorname{argmax}}\ \xi(\tilde{\gamma}) = \gamma, \tag{1}$$

where

$$\begin{aligned}\xi(\tilde{\gamma}) =& \tilde{\gamma}(1-\tilde{\gamma})\left(\left(\frac{\gamma}{\tilde{\gamma}}\right)^2 \mathbb{1}_{\tilde{\gamma}>\gamma} + \left(\frac{1-\gamma}{1-\tilde{\gamma}}\right)^2 \mathbb{1}_{\tilde{\gamma}\leq\gamma}\right)[2\mu_{XY} - \mu_{XX} - \mu_{YY}] \\ =& \left(\gamma^2 \frac{1-\tilde{\gamma}}{\tilde{\gamma}} \mathbb{1}_{\tilde{\gamma}>\gamma} + (1-\gamma)^2 \frac{\tilde{\gamma}}{1-\tilde{\gamma}} \mathbb{1}_{\tilde{\gamma}\leq\gamma}\right)[2\mu_{XY} - \mu_{XX} - \mu_{YY}]\,. \end{aligned} \tag{2}$$

Proof. The detail of proof is shown in Appendix. □

The scaled energy distance $\xi(\tilde{\gamma})$ defined in Equation 2 is a piecewise function based on the location of $\tilde{\gamma}$ and γ. We then prove $\hat{Q}(\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}, \mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^T)$ is convergent to the scaled energy distance $\xi(\tilde{\gamma})$ almost surely uniformly in $\tilde{\gamma} \in [\delta_T, 1-\delta_T]$.

Lemma 2. *Under Assumption 1,*

$$\lim_{T\to\infty}\left\{\sup_{\tilde{\gamma}\in[\delta_T,1-\delta_T]}\left|\frac{1}{T}\hat{Q}(\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}, \mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^T) - \xi(\tilde{\gamma})\right|\right\} = 0, a.s,$$

where $\xi(\tilde{\gamma})$ is defined in Equation (2).

Proof. The detail of proof is shown in Appendix. □

Finally, we define the estimated change point $\hat{\tau}$ in Equation 3 in terms of δ_T.

$$\hat{\tau} = \underset{t\in\{\lceil T\delta_T\rceil, \lceil T\delta_T\rceil+1,\ldots,\lfloor T(1-\delta_T)\rfloor\}}{\operatorname{argmax}} \hat{Q}(\mathbf{Z}_1^t, \mathbf{Z}_{t+1}^T). \tag{3}$$

Since we proved that $\frac{1}{T}\hat{Q}(\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}, \mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^T)$ converges to $\xi(\tilde{\gamma})$ almost surely uniformly in $\tilde{\gamma}$, furthermore, $\hat{\tau}$ and γ are argmax of $\hat{Q}(\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}, \mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^T)$ and $\xi(\tilde{\gamma})$ respectively, we can naturally have the result that $\frac{\hat{\tau}}{T}$ converges to γ in probability by Consistency of M-estimators. However, we here propose a stronger statement that $\frac{\hat{\tau}}{T}$ converges to γ almost surely in Theorem 1 due to one of the property of $\xi(\tilde{\gamma})$.

Theorem 1. *Under Assumption 1,* $\forall \epsilon > 0$,

$$P(\lim_{T\to\infty}\left|\frac{\hat{\tau}}{T}-\gamma\right| < \epsilon) = 1$$

Proof. The detail of proof is shown in Appendix. □

3 Multiple Change Point

Without lose of generality, we first consider there exist two change points in a sequence.

Assumption 2. *Given a heterogeneous sequence of independent observations* $\{Z_1, \ldots, Z_T\}$ *taking values in general metric space* (Ω, d) *with two change points denoted by* τ_1 *and* τ_2, *we let* $Z_1, \ldots, Z_{\lfloor \gamma_1 T\rfloor} \overset{i.i.d.}{\sim} F_L$, $Z_{\lfloor \gamma_1 T\rfloor+1}, \ldots, Z_{\lfloor \gamma_2 T\rfloor} \overset{i.i.d.}{\sim} F_C$ *and* $Z_{\lfloor \gamma_2 T\rfloor+1}, \ldots, Z_T \overset{i.i.d.}{\sim} F_R$, *where* $\gamma_1 = \frac{\tau_1}{T}$ *and* $\gamma_2 = \frac{\tau_2}{T}$. *For ease of notation, we denote* $\mu_{XX} = E(d^\alpha(X, X'))$, $\mu_{YY} = E(d^\alpha(Y', Y'))$, $\mu_{UU} = E(d^\alpha(U, U'))$, $\mu_{XY} = E\left(d^\alpha\left(X, Y\right)\right)$, $\mu_{XU} = E\left(d^\alpha\left(X, U\right)\right)$ *and* $\mu_{UY} = E\left(d^\alpha\left(U, Y\right)\right)$, *where* $X, X' \sim F_L$, $U, U' \sim F_C$, $Y, Y' \sim F_R$, X, X', U, U', Y', Y' *are mutually independent, and* $d^\alpha(.)$ *is strictly negative definite kernel. Let* $\{\delta_T\}$ *be a sequence of positive numbers with property of* $\delta_T \to 0$ *and* $T\delta_T \to \infty$, *as* $T \to \infty$.

Lemma 3. *Under Assumption 2,*

$$\lim_{T\to\infty}\left\{\sup_{\substack{\gamma_1,\gamma_2\in[\delta_T,1-\delta_T]\\ \gamma_1\le\gamma_2}}\left|\binom{T}{2}^{-1}\sum_{1<i<j<T} d^\alpha(Z_i, Z_j) - \Big({\gamma_1}^2\mu_{XX} + (\gamma_2-\gamma_1)^2\mu_{UU} + (1-\gamma_2)^2\mu_{YY} \right.\right.$$
$$\left.\left. +2\gamma_1(\gamma_2-\gamma_1)\mu_{XY} + 2\gamma_1(1-\gamma_2)\mu_{XY} + 2(1-\gamma_2)(\gamma_2-\gamma_1)\mu_{UY}\Big)\right|\right\} = 0.$$

Proof. The proof of Lemma 3 is shown in Appendix. □

Same as what we have proved in Lemma 1, Lemma 3 shows that when a sequence of length 1 with three distributions truncated at γ_1 and γ_2, the average pairwise deviation $\binom{T}{2}^{-1}\sum_{1\le i<j\le T} d^\alpha(Z_i, Z_j)$ converges almost surely to a function of distributions mean in a form of ${\gamma_1}^2\mu_{XX} + (\gamma_2-\gamma_1)^2\mu_{UU} + (1-\gamma_2)^2\mu_{YY} + 2\gamma_1(\gamma_2-\gamma_1)\mu_{XY} + 2\gamma_1(1-\gamma_2)\mu_{XY} + 2(1-\gamma_2)(\gamma_2-\gamma_1)\mu_{UY}$, uniformly in $\gamma_1, \gamma_2 \in [\delta_T, 1-\delta_T]$. We will apply Lemma 1 and this result to prove the test statistics $\hat{Q}$ converges to a scaled energy distance almost surely in the two change points case.

Since γ_1 and γ_2 truncate the sequence into three distributions F_L, F_C, and F_R, it means they both are TRUE change points in the sequence. Thus, for any candidate

$\tilde{\gamma} \in [\delta_T, 1-\delta_T]$, we can calculate a scaled energy distance denoted as $f(\tilde{\gamma})$ in terms of γ_1 and γ_2. To prove the scaled energy distance $f(\tilde{\gamma})$ is maximized at γ_1 or γ_2, we propose the following two assumptions: Assumption 3 and Assumption 4.

Assumption 3. *let $\eta(\gamma_1,\gamma_2) > 0$, where $\eta(\gamma_1,\gamma_2)$ is defined as below:*

$$\begin{aligned}\eta(\gamma_1,\gamma_2) =&(2\mu_{XY}-\mu_{XX}-\mu_{YY}) + \gamma_1^2(2\mu_{XU}-\mu_{UU}-\mu_{XX})\\&+\gamma_2^2(2\mu_{UY}-\mu_{UU}-\mu_{YY}) + 2\gamma_1(\mu_{UY}+\mu_{XX}-\mu_{XU}-\mu_{XY})\\&+2\gamma_2(\mu_{XU}+\mu_{YY}-\mu_{UY}-\mu_{XY}) + 2\gamma_1\gamma_2(\mu_{UU}+\mu_{XY}-\mu_{XU}-\mu_{UY}).\end{aligned}$$

Assumption 4. *let $\varphi(\gamma_1,\gamma_2) > 0$, where $\eta(\gamma_1,\gamma_2)$ is defined as below:*

$$\begin{aligned}\varphi(\gamma_1,\gamma_2) =&\gamma_1^2(2\mu_{XU}-\mu_{UU}-\mu_{XX}) + \gamma_2^2(2\mu_{UY}-\mu_{UU}-\mu_{YY})\\&+2\gamma_1\gamma_2(\mu_{UU}+\mu_{XY}-\mu_{XU}-\mu_{UY}).\end{aligned}$$

Assumption 3 and Assumption 4 mainly contribute to the property of convex of $f(\tilde{\gamma})$ when $\tilde{\gamma} < \gamma_1$ and $\tilde{\gamma} > \gamma_2$. To explain the reasonableness of those two assumptions, we propose Lemma 4 and 5. Lemma 4 shows those two assumptions hold when $\Omega = \mathcal{R}$ and $d^\alpha(p,q) = (p-q)^2$. Lemma 5 shows those two assumptions hold when Ω is a set of one-dimensional probability distributions and metric d is 2-Wasserstein distance.

Lemma 4. *Given $E(X) = \mu_X$, $Var(X) = \sigma_X^2$, $E(Y) = \mu_Y$, $Var(Y) = \sigma_Y^2$, $E(U) = \mu_U$, and $Var(U) = \sigma_U^2$, Assumption 3 and 4 are holds when $\Omega = \mathcal{R}$ and $d^\alpha(p,q) = (p-q)^2$.*

Proof. The proof of Lemma 4 is shown in Appendix. □

Lemma 5. *When Ω is a set of one-dimensional probability distributions and distance is 2-Wasserstein distance which is define as $d^\alpha(\mathcal{G},\mathcal{H}) = W_2^2(\mathcal{G},\mathcal{H}) = \int_0^1 \left(\mathcal{G}^{-1}(t) - \mathcal{H}^{-1}(t)\right)^2 dt$ with $\alpha = 2$, such that for any $\mathcal{G}$, $\mathcal{H} \in \Omega$, $\int_0^1 \left(\mathcal{G}^{-1}(t) - \mathcal{H}^{-1}(t)\right)^2 dt < \infty$, Assumption 3 and 4 are holds.*

Proof. The proof of Lemma 5 is shown in Appendix. □

Thus, we can prove the scaled energy distance $f(\tilde{\gamma})$ is maximized at γ_1 and γ_2.

Proposition 2. *Under Assumption 2, 3 and 4, The scaled energy distance $f(\tilde{\gamma})$ that* *defined in Equation (4) is maximized when* $\tilde{\gamma} = \gamma_1$ *or* $\tilde{\gamma} = \gamma_2$*, where*

$$f(\tilde{\gamma}) = \tilde{\gamma}(1-\tilde{\gamma})\Bigg(\mathbb{1}_{\tilde{\gamma}\le\gamma_1<\gamma_2}\mathcal{E}(F_L, \frac{\gamma_1-\tilde{\gamma}}{1-\tilde{\gamma}}F_L + \frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}}F_R)$$
$$+\mathbb{1}_{\gamma_1<\tilde{\gamma}\le\gamma_2}\mathcal{E}(\frac{\gamma_1}{\tilde{\gamma}}F_L + \frac{\tilde{\gamma}-\gamma_1}{\tilde{\gamma}}F_C, \frac{\gamma_2-\tilde{\gamma}}{1-\tilde{\gamma}}F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}}F_R)$$
$$+\mathbb{1}_{\gamma_1<\gamma_2<\tilde{\gamma}}\mathcal{E}(\frac{\gamma_1-\tilde{\gamma}}{1-\tilde{\gamma}}F_L + \frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}}F_R, F_R)\Bigg). \tag{4}$$

Proof. The proof of Proposition 2 is shown in Appendix. □

Lemma 6. *Under Assumption 2,*

$$\lim_{T\to\infty}\left\{\sup_{\tilde{\gamma}\in[\delta_T,1-\delta_T]}\left|\frac{1}{T}\hat{Q}(\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}, \mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^T) - f(\tilde{\gamma})\right|\right\} = 0, a.s.$$

Proof. The proof of Lemma 6 is shown in Appendix. □

Theorem 2. *Under Assumption 2, 3 and 4,* $\forall\epsilon > 0$,

$$\lim_{T\to\infty} P(\left|\frac{\hat{\tau}}{T} - \gamma_{true}\right| > \epsilon) = 0,$$

where $\hat{\tau}$ *is defined in Equation 3 and* γ_{true} *is one of* $\{\gamma_1, \gamma_2\}$ *which maximized* $f(\tilde{\gamma})$*.*

Proof. When

1. $\lim_{T\to\infty} P(sup_{\tilde{\gamma}\in[\delta_T,1-\delta_T]}\left|\frac{1}{T}\hat{Q}(\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}, \mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^T) - f(\tilde{\gamma})\right| > \epsilon) = 0$, by Lemma 6,
2. $f(\tilde{\gamma})$ continuous and uniquely maximized at γ_{true}, proved in Proposition 2,
3. $\hat{\tau} = \underset{t\in\{\lceil T\delta_T\rceil,\lceil T\delta_T\rceil+1,\ldots,\lfloor T(1-\delta_T)\rfloor\}}{\operatorname{argmax}} \hat{Q}(\mathbf{Z}_1^t, \mathbf{Z}_{t+1}^T)$, defined in Equation 3,

it is evident that $\frac{\hat{\tau}}{T} \xrightarrow{p} \hat{\gamma}$, as $T \to \infty$, by Consistency in M-estimator. □

Finally, let's consider the most general case that there exist k change points in the sequence.

Assumption 5. *Suppose there are* $k+1$ *distributions in the heterogeneous sequence* $\{Z_1, \ldots, Z_T\}$*, denoted by* $F_1, \ldots, F_{k+1}$*, such that there exists* k *change points, denoted by* $\tau_1, \ldots, \tau_k$*, where* $1 < \tau_1 < \tau_k < T$ *and* $k \ge 2$*. For any two change point* τ_i *and* τ_j

where $1 \le i < j \le k$, we assume the observations $\{Z_1, \ldots, Z_{\lfloor T\gamma^{(i)} \rfloor}\}$ follow the mixture distribution of $F_1, \ldots, F_i$, denoted by F_L; observations $\{Z_{\lfloor T\gamma^{(j+1)} \rfloor}, \ldots, Z_T\}$ follow the mixture distribution of $F_{j+1}, \ldots, F_{k+1}$, denoted by F_R; the remaining observations follow the mixture distribution $F_{i+1}, \ldots, F_j$, denoted by F_C, in which $\gamma^{(i)} = \frac{\tau_i}{T}$ and $\gamma^{(j)} = \frac{\tau_j}{T}$.

Under Assumption 5, the multiple change point case can be simplified to two change points case. Like we did in two change point case, see expression of $f(\tilde{\gamma})$ in Equation 4, we can define a scale energy distance respect to $\gamma^{(i)}$ and $\gamma^{(j)}$ denoted by $g_{i,j}(\tilde{\gamma})$, which is shown as below:

$$\begin{aligned} g_{i,j}(\tilde{\gamma}) = \tilde{\gamma}(1-\tilde{\gamma}) \Bigg[& \mathbb{1}_{\tilde{\gamma} \le \gamma^{(i)}} \mathcal{E}(F_L, \frac{\gamma^{(i)} - \tilde{\gamma}}{1-\tilde{\gamma}} F_L + \frac{\gamma^{(j)} - \gamma^{(i)}}{1-\tilde{\gamma}} F_C + \frac{1-\gamma^{(j)}}{1-\tilde{\gamma}} F_R) \\ & + \mathbb{1}_{\gamma^{(i)} \le \tilde{\gamma} \le \gamma^{(j)}} \mathcal{E}(\frac{\gamma^{(i)}}{\tilde{\gamma}} F_L + \frac{\tilde{\gamma} - \gamma^{(i)}}{\tilde{\gamma}} F_C, \frac{\gamma^{(j)} - \tilde{\gamma}}{1-\tilde{\gamma}} F_C + \frac{1-\gamma^{(j)}}{1-\tilde{\gamma}} F_R) \\ & + \mathbb{1}_{\gamma^{(j)} \le \tilde{\gamma}} \mathcal{E}(\frac{\gamma^{(i)} - \tilde{\gamma}}{1-\tilde{\gamma}} F_L + \frac{\gamma^{(j)} - \gamma^{(i)}}{1-\tilde{\gamma}} F_C + \frac{1-\gamma^{(j)}}{1-\tilde{\gamma}} F_R, F_R) \Bigg] \end{aligned}$$

Proposition 3. *Under Assumptions 2, 3, 4 and 5. $\forall \epsilon > 0$,*

$$\lim_{T \to \infty} P(\left| \frac{\hat{\tau}}{T} - \gamma_{true} \right| > \epsilon) = 0,$$

where $\hat{\tau}$ is defined in Equation (3) and γ_{true} is one of $\{\gamma^{(i)}, \gamma^{(j)}\}$ which maximized $f(\tilde{\gamma})$.

Proof. Under Assumptions 5, we can regard $\gamma^{(i)}$ and $\gamma^{(j)}$ of $g_{i,j}(\tilde{\gamma})$ as γ_1 and γ_2 of $f(\tilde{\gamma})$. Thus, based on Theorem 2 , the proof of Proposition 3 is clear. We omit the details here. □

Appendix

Proof of Lemma 1. Let $\binom{T}{2}^{-1} \sum_{1<i<j<T} d^{\alpha}(Z_i, Z_j)$ be split into three sets, which are $\Pi_1 = \{(i,j) : i < j, Z_i, Z_j \sim F_L\}$, $\Pi_2 = \{(i,j) : i < j, Z_i \sim F_L, Z_j \sim F_R\}$ and $\Pi_3 = \{(i,j) : i < j, Z_i, Z_j \sim F_R\}$. Denote $l = \lfloor \gamma T \rfloor$ and $s = T - \lfloor \gamma T \rfloor$ for calculation convenience. By the strong law of large numbers for U-statistics Hoeffding (1961), we have that with probability 1: $\forall \epsilon > 0$, $\exists N_1, N_2, N_3, N_4 \in \mathbf{N}$, such that for $l > N_1$, $s \cdot l > N_2$, $s > N_3$ and

$T > N_4$,

$$\left| \binom{l}{2}^{-1} \sum_{\Pi_1} d^\alpha(Z_i, Z_j) - \mu_{XX} \right| < \epsilon; \tag{5a}$$

$$\left| \binom{s}{1}^{-1} \binom{l}{1}^{-1} \sum_{\Pi_2} d^\alpha(Z_i, Z_j) - \mu_{XY} \right| < \epsilon; \tag{5b}$$

$$\left| \binom{s}{2}^{-1} \sum_{\Pi_3} d^\alpha(Z_i, Z_j) - \mu_{YY} \right| < \epsilon; \tag{5c}$$

$$\frac{1}{T-1} < \frac{\epsilon}{2}. \tag{5d}$$

by the Inequality (5d), quantities that

$$\left| \frac{l}{T} - \gamma \right| < \epsilon, \tag{5e}$$

$$\left| \frac{l-1}{T-1} - \gamma \right| < \epsilon, \tag{5f}$$

$$\left| \frac{s}{T} - (1-\gamma) \right| < \epsilon, \tag{5g}$$

$$\left| \frac{s-1}{T-1} - (1-\gamma) \right| < \epsilon, \tag{5h}$$

$$\left| \frac{l}{T-1} - \gamma \right| < \epsilon \tag{5i}$$

can be easily proved. Considering the set Π_1, if we multiply inequality (5e) and (5f), as shown below

$$\left| \frac{l}{T} - \gamma \right| \left| \frac{l-1}{T-1} - \gamma \right| < \epsilon^2,$$

some tedious manipulation yields

$$\left| \frac{l}{T} \frac{l-1}{T-1} - \gamma^2 \right| < \epsilon^2 + 2\gamma\epsilon. \tag{5j}$$

Continuously, multiply Inequality (5a) and (5j)

$$\left| \frac{l}{T} \frac{l-1}{T-1} - \gamma^2 \right| \left| \binom{l}{2}^{-1} \sum_{\Pi_1} d^\alpha(Z_i, Z_j) - \mu_{XX} \right| < \epsilon^3 + 2\gamma\epsilon^2,$$

after rearranging terms, we will have the following conclusion

$$\begin{aligned}&\left|\binom{T}{2}^{-1}\sum_{\Pi_1} d^\alpha(Z_i,Z_j)-\gamma^2\mu_{XX}\right|\\&<\epsilon^3+(2\gamma+(1+2\gamma)\mu_{XX})\epsilon^2+\gamma^2\epsilon.\end{aligned}\tag{5k}$$

The argument of set Π_2 and Π_3 is analogous to that in set Π_1. Multiplying inequality (5g) & (5h) and (5g) & (5i) and doing some transformations, we will have

$$\left|\frac{s}{T}\frac{s-1}{T-1}-(1-\gamma)^2\right|<\epsilon^2+2(1-\gamma)\epsilon,\tag{5l}$$

$$\left|\frac{s}{T}\frac{l}{T-1}-\gamma(1-\gamma)\right|<\epsilon^2+\epsilon\tag{5m}$$

respectively, continuously multiply inequality (5b) & (5l) and (5c) & (5m) , we will get the following results

$$\begin{aligned}&\left|\frac{1}{T}\frac{1}{T-1}\sum_{\Pi_2} d^\alpha(Z_i,Z_j)-\gamma(1-\gamma)\mu_{XY}\right|\\&<\epsilon^3+(\mu_{XY}+1)\epsilon^2+(\gamma(1-\gamma)+\mu_{XY})\epsilon,\end{aligned}\tag{5n}$$

$$\begin{aligned}&\left|\binom{T}{2}^{-1}\sum_{\Pi_3} d^\alpha(Z_i,Z_j)-(1-\gamma)^2\mu_{YY}\right|\\&<\epsilon^3+(2(1-\gamma)+\mu_{YY})\epsilon^2+((1-\gamma)^2+2(1-\gamma)\mu_{YY})\epsilon.\end{aligned}\tag{5o}$$

Continuously, if we add Inequalities (5k), (5o) and twice (5n), and by triangle inequality rule, for $T>N_1\vee N_2\vee N_3\vee N_4$,

$$\begin{aligned}&\left|\binom{T}{2}^{-1}\sum_{1<i<j<T} d^\alpha(Z_i,Z_j)\right.\\&\left.-\left(\gamma^2\mu_X+(1-\gamma)^2\mu_Y+2\gamma(1-\gamma)\mu_{XY}\right)\right|<g(\epsilon)\end{aligned}$$

where $g(\epsilon)$ is a function of ϵ. We omit the expression of $g(\epsilon)$ here due to the arbitrariness of ϵ. Hence with probability 1, $\forall\epsilon>0$, $\exists N\in\mathbf{N}$, such that for $T\delta_T>N$, and every

$\gamma \in [\delta_T, 1-\delta_T]$,

$$\left| \binom{T}{2}^{-1} \sum_{1<i<j<T} d^{\alpha}(Z_i, Z_j) \right.$$
$$\left. - (\gamma^2 \mu_X + (1-\gamma)^2 \mu_Y + 2\gamma(1-\gamma)\mu_{XY}) \right| < \epsilon.$$

Such that, we have proved $\binom{T}{2}^{-1} \sum_{1<i<j<T} d^{\alpha}(Z_i, Z_j) \xrightarrow{a.s.} (\gamma^2 \mu_X + (1-\gamma)^2 \mu_Y + 2\gamma(1-\gamma)\mu_{XY})$, uniformly in γ. $\square$

Proof of Proposition 1. Let's show how to get the scaled energy distance $\xi(\tilde{\gamma})$ first.
If $\tilde{\gamma} \leq \gamma$, the random object from cluster $\mathbf{Z}_1^{\lfloor \tilde{\gamma} T \rfloor}$ follows the distribution F_L while the random object from set $\mathbf{Z}_{\lfloor \tilde{\gamma} T \rfloor+1}^{T}$ follows the distribution F_L with probability $\frac{\gamma-\tilde{\gamma}}{1-\tilde{\gamma}}$ and the distribution of F_R with probability $\frac{1-\gamma}{1-\tilde{\gamma}}$, which can be considered as following a mixture distribution $\frac{\gamma-\tilde{\gamma}}{1-\tilde{\gamma}} F_L + \frac{1-\gamma}{1-\tilde{\gamma}} F_R$. Hence, the energy distance $\mathcal{E}(F_L, \frac{\gamma-\tilde{\gamma}}{1-\tilde{\gamma}} F_L + \frac{1-\gamma}{1-\tilde{\gamma}} F_R)$ is calculated to be

$$\begin{aligned}
&\mathcal{E}(F_L, \frac{\gamma-\tilde{\gamma}}{1-\tilde{\gamma}} F_L + \frac{1-\gamma}{1-\tilde{\gamma}} F_R) \\
=&2\big(\frac{\gamma-\tilde{\gamma}}{1-\tilde{\gamma}} \mu_{XX} + \frac{1-\gamma}{1-\tilde{\gamma}} \mu_{XY}\big) && \text{(6a)} \\
&- \mu_{XX} && \text{(6b)} \\
&- \left((\frac{\gamma-\tilde{\gamma}}{1-\tilde{\gamma}})^2 \mu_{XX} + 2\frac{\gamma-\tilde{\gamma}}{1-\tilde{\gamma}} \frac{1-\gamma}{1-\tilde{\gamma}} \mu_{XY} + (\frac{1-\gamma}{1-\tilde{\gamma}})^2 \mu_{YY} \right) && \text{(6c)} \\
=&(\frac{1-\gamma}{1-\tilde{\gamma}})^2 \left[2\mu_{XY} - \mu_{XX} - \mu_{YY} \right].
\end{aligned}$$

If $\tilde{\gamma} > \gamma$, the random object from set $\mathbf{Z}_1^{\lfloor \tilde{\gamma} T \rfloor}$ follows a mixture distribution expressed as $\frac{\gamma}{\tilde{\gamma}} F_L + \frac{\tilde{\gamma}-\gamma}{\tilde{\gamma}} F_R$ and the random object from set $\mathbf{Z}_{\lfloor \tilde{\gamma} T \rfloor+1}^{T}$ follows distribution F_R, such that the energy distance $\mathcal{E}(\frac{\gamma}{\tilde{\gamma}} F_L + \frac{\tilde{\gamma}-\gamma}{\tilde{\gamma}} F_R, F_R)$ is calculated to be $(\frac{\gamma}{\tilde{\gamma}})^2 \left[2\mu_{XY} - \mu_{XX} - \mu_{YY} \right]$.

Combing two situations above, the scaled energy distance denoted by $\xi(\tilde{\gamma})$ is shown in Equation 2.

It can easily be seen that $\gamma^2 \frac{1-\tilde{\gamma}}{\tilde{\gamma}} \mathbb{1}_{\tilde{\gamma}>\gamma} + (1-\gamma)^2 \frac{\tilde{\gamma}}{1-\tilde{\gamma}} \mathbb{1}_{\tilde{\gamma}\leq\gamma}$ is maximized when $\tilde{\gamma} = \gamma$. Also, using the property of negative definite kernel (Rachev et al. (2013)), $2E(d^{\alpha}(X, Y)) - E(d^{\alpha}(X, X')) - E(d^{\alpha}(Y, Y')) \geq 0$. Hence $\xi(\tilde{\gamma})$ is maximized when $\tilde{\gamma} = \gamma$. Proposition 1 is proved.

$\square$

Proof of Lemma 2. Split $\hat{Q}(\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}, \mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^T)$ into three terms: $\binom{\lfloor\tilde{\gamma}T\rfloor}{2}^{-1}\sum_{1\le i<j\le\lfloor\tilde{\gamma}T\rfloor} d^\alpha(Z_i, Z_j)$, $\binom{T-\lfloor\tilde{\gamma}T\rfloor}{2}^{-1}\sum_{\lfloor\tilde{\gamma}T\rfloor+1\le i<j\le T} d^\alpha(Z_i, Z_j)$, and $\frac{2}{\lfloor\tilde{\gamma}T\rfloor(T-\lfloor\tilde{\gamma}T\rfloor)}\sum_{i=1}^{\lfloor\tilde{\gamma}T\rfloor}\sum_{j=\lfloor\tilde{\gamma}T\rfloor+1}^{T} d^\alpha(Z_i, Z_j)$.

If $\tilde{\gamma} \le \gamma$, the energy distance $\mathcal{E}(F_L, \frac{\gamma-\tilde{\gamma}}{1-\tilde{\gamma}}F_L + \frac{1-\gamma}{1-\tilde{\gamma}}F_R)$ can be split into three terms shown at 6b,6c, and 6a. According to Lemma 1, it can be proved that $\hat{Q}(\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}, \mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^T)$ converges to $\mathcal{E}(F_L, \frac{\gamma-\tilde{\gamma}}{1-\tilde{\gamma}}F_L + \frac{1-\gamma}{1-\tilde{\gamma}}F_R)$ almost surely uniformly in $\tilde{\gamma}$.

Similarly, if $\gamma < \tilde{\gamma}$, $\hat{Q}(\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}, \mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^T)$ converges to $\mathcal{E}(\frac{\gamma}{\tilde{\gamma}}F_L + \frac{\tilde{\gamma}-\gamma}{\tilde{\gamma}}F_R, F_R)$ almost surely uniformly in $\tilde{\gamma}$.

Combining situations of $\tilde{\gamma} > \gamma$ and $\tilde{\gamma} \le \gamma$, $\frac{1}{T}\hat{Q}(\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}, \mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^T) \xrightarrow{a.s.} \xi(\tilde{\gamma})$ uniformly in $\tilde{\gamma}$. □

Proof of Theorem 1. For ease of notation, we define

$$h_T(\omega, \tilde{\gamma}) = \frac{1}{T}\hat{Q}(\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}, \mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^T),\ where\ \omega \in \Omega,$$

and

$$\hat{\gamma_T} = \frac{\hat{\tau}}{T}.$$

Assume $\hat{\gamma_T} \overset{a.s.}{\not\to} \gamma$, such that, $\exists\ \epsilon > 0$, for all $T_1 \in \mathbb{N}$, $\exists\ T > T_1$,

$$|\hat{\gamma_T} - \gamma| > \epsilon. \tag{7a}$$

By Proposition 1 that $\gamma = \underset{\tilde{\gamma}}{\operatorname{argmax}}\, \xi(\tilde{\gamma})$ and the property of $\xi(\tilde{\gamma})$ that it increases when $\tilde{\gamma} \in [\delta_T, \gamma)$ and decreases when $\tilde{\gamma} \in [\gamma, 1-\delta_T]$, we define

$$\varepsilon := min\{\xi(\tilde{\gamma}) - \xi(\tilde{\gamma} - \epsilon), \xi(\tilde{\gamma}) - \xi(\tilde{\gamma} + \epsilon)\}.$$

With inequality 7a, we have

$$\xi(\gamma) - \xi(\hat{\gamma_T}) > \varepsilon. \tag{7b}$$

As we define in Equation 3 that $\hat{\tau}$ is argmax of $h_T(\omega, \tilde{\gamma})$,

$$h_T(\omega, \hat{\gamma_T}) > h_T(\omega, \gamma). \tag{7c}$$

Choosing T_1 large enough, we have

$$|h_T(\omega, \hat{\gamma_T}) - \xi(\hat{\gamma_T})| < \frac{\varepsilon}{100}, \tag{7d}$$

$$|h_T(\omega, \gamma) - \xi(\gamma)| < \frac{\varepsilon}{100}, \tag{7e}$$

according to Lemma 2. Such that, according to Equation 7c, 7d and 7b, we have

$$\begin{aligned} h_T(\omega, \gamma) + \frac{99}{100}\varepsilon &< h_T(\omega, \hat{\gamma_T}) + \frac{99}{100}\varepsilon \\ &< \xi(\hat{\gamma_T}) + \frac{\varepsilon}{100} + \frac{99}{100} \\ &< \xi(\hat{\gamma_T}) + \varepsilon \\ &< \xi(\gamma), \end{aligned}$$

which contradicts with inequality 7e. Thus, we can conclude $\hat{\gamma}_T \overset{a.s.}{\to} \gamma$, i.e, $\frac{\hat{\tau}}{T} \overset{a.s.}{\to} \gamma$. $\square$

Proof of Lemma 3. The proof of Lemma 3 is analogous to the proof of Lemma 1. Denote the length of sequence before γ_1 as $m = \lfloor \gamma_1 T \rfloor$, the length of sequence after γ_2 as $n = T - \lfloor \gamma_2 T \rfloor$, the length of sequence between γ_1 and γ_2 as $g = \lfloor \gamma_2 T \rfloor - \lfloor \gamma_1 T \rfloor$. The combinations of $\binom{T}{2}^{-1} \sum_{1<i<j<T} d(Z_i, Z_j)$ can be split into six sets, which are $\Pi_1 = \{(i,j) : i < j, Z_i, Z_j \sim F_L\}$, $\Pi_2 = \{(i,j) : i < j, Z_i, Z_j \sim F_C\}$, $\Pi_3 = \{(i,j) : i < j, Z_i, Z_j \sim F_R\}$, $\Pi_4 = \{(i,j) : i < j, Z_i \sim F_L, Z_j \sim F_C\}$, $\Pi_5 = \{(i,j) : i < j, Z_i \sim F_L, Z_j \sim F_R\}$, $\Pi_6 = \{(i,j) : i < j, Z_i \sim F_C, Z_j \sim F_R\}$. By the strong law of large numbers for U-statistics, in the metric space (Ω, d), we have with probability 1: $\forall \epsilon > 0$, $\exists N_1, N_2, N_3, N_4, N_5, N_6, N_7 \in \mathbf{N}$, such that for $m > N_1$, $g > N_2$, $n > N_3$ and $m \cdot n > N_4$,

$m \cdot g > N_5$, $g \cdot n > N_6$, $T > N_7$,

$$\left| \binom{m}{2}^{-1} \sum_{\Pi_1} d^\alpha(Z_i, Z_j) - \mu_{XX} \right| < \epsilon; \tag{8a}$$

$$\left| \binom{g}{2}^{-1} \sum_{\Pi_2} d^\alpha(Z_i, Z_j) - \mu_{UU} \right| < \epsilon; \tag{8b}$$

$$\left| \binom{n}{2}^{-1} \sum_{\Pi_3} d^\alpha(Z_i, Z_j) - \mu_{YY} \right| < \epsilon; \tag{8c}$$

$$\left| \binom{m}{1}^{-1} \binom{n}{1}^{-1} \sum_{\Pi_4} d^\alpha(Z_i, Z_j) - \mu_{XY} \right| < \epsilon; \tag{8d}$$

$$\left| \binom{m}{1}^{-1} \binom{g}{1}^{-1} \sum_{\Pi_5} d^\alpha(Z_i, Z_j) - \mu_{XU} \right| < \epsilon; \tag{8e}$$

$$\left| \binom{g}{1}^{-1} \binom{n}{1}^{-1} \sum_{\Pi_6} d^\alpha(Z_i, Z_j) - \mu_{UY} \right| < \epsilon; \tag{8f}$$

$$\frac{1}{T-1} < \frac{\epsilon}{2}. \tag{8g}$$

by the inequality (8g), the quantities

$$\left| \frac{m}{T} - \gamma_1 \right| < \epsilon, \tag{8h}$$

$$\left| \frac{m-1}{T-1} - \gamma_1 \right| < \epsilon, \tag{8i}$$

$$\left| \frac{m}{T-1} - \gamma_1 \right| < \epsilon, \tag{8j}$$

$$\left| \frac{n}{T} - (1-\gamma_2) \right| < \epsilon, \tag{8k}$$

$$\left| \frac{n}{T-1} - (1-\gamma_2) \right| < \epsilon, \tag{8l}$$

$$\left| \frac{n-1}{T-1} - (1-\gamma_2) \right| < \epsilon, \tag{8m}$$

$$\left| \frac{g}{T} - (\gamma_2-\gamma_1) \right| < \epsilon, \tag{8n}$$

$$\left| \frac{g-1}{T-1} - (\gamma_2-\gamma_1) \right| < \epsilon \tag{8o}$$

can be easily proved. Multiplying every two of the inequalities above: (8h) &(8i), (8k) &(8l),(8n) &(8o),(8k) &(8j),(8n) &(8j) as well as (8n)&(8l) and doing some transformation, we will get

$$\left|\frac{m}{T}\frac{m-1}{T-1}-\gamma_1^2\right|<\epsilon^2+2\gamma_1\epsilon, \tag{9a}$$

$$\left|\frac{n}{T}\frac{n-1}{T-1}-(1-\gamma_2)^2\right|<\epsilon^2+2(1-\gamma_2)\epsilon, \tag{9b}$$

$$\left|\frac{g}{T}\frac{g-1}{T-1}-(\gamma_2-\gamma_1)^2\right|<\epsilon^2+2(\gamma_2-\gamma_1)\epsilon, \tag{9c}$$

$$\left|\frac{n}{T}\frac{m}{T-1}-\gamma_1(1-\gamma_2)\right|<\epsilon^2+(1-\gamma_2)\epsilon+\gamma_1\epsilon, \tag{9d}$$

$$\left|\frac{g}{T}\frac{m}{T-1}-\gamma_1(\gamma_2-\gamma_1)\right|<\epsilon^2+(\gamma_2-\gamma_1)\epsilon+\gamma_1\epsilon, \tag{9e}$$

$$\begin{aligned}&\left|\frac{g}{T}\frac{n}{T-1}-(1-\gamma_2)(\gamma_2-\gamma_1)\right|\\ &<\epsilon^2+(\gamma_2-\gamma_1)\epsilon+(1-\gamma_2)\epsilon\end{aligned} \tag{9f}$$

respectively.

Moreover, multiplying (8a)&(9a), (8b)&(9b), (8c)&(9c), (8d)&(9d), (8e)&(9e) and (8f)&(9f), and rearranging terms, we will get the following solutions

$$\left|\binom{T}{2}^{-1}\sum_{\Pi_1}d^\alpha(Z_i,Z_j)-\gamma_1^2\mu_{XX}\right|<\epsilon^3+(\mu_{XX}+2\gamma_1)\epsilon^2+(2\gamma_1\mu_{XX}+\gamma_1^2)\epsilon, \tag{10a}$$

$$\begin{aligned}\left|\binom{T}{2}^{-1}\sum_{\Pi_2}d^\alpha(Z_i,Z_j)-(\gamma_2-\gamma_1)^2\mu_{UU}\right|&<\epsilon^3+(\mu_{UU}+2(\gamma_2-\gamma_1))\epsilon^2\\ &+(2(\gamma_2-\gamma_1)\mu_{UU}+(\gamma_2-\gamma_1)^2)\epsilon,\end{aligned} \tag{10b}$$

$$\begin{aligned}&\left|\binom{T}{2}^{-1}\sum_{\Pi_3}d^\alpha(Z_i,Z_j)-(1-\gamma_2)^2\mu_{YY}\right|\\ &\qquad<\epsilon^3+(\mu_{Y+Y}2(1-\gamma_2))\epsilon^2\\ &\qquad\qquad+(2(1-\gamma_2)\mu_{YY}+(1-\gamma_2)^2)\epsilon,\end{aligned} \tag{10c}$$

$$\left|\frac{1}{T(T-1)}\sum_{\Pi_4} d^\alpha(Z_i,Z_j)-\gamma_1(1-\gamma_2)\mu_{XY}\right| \\ <\epsilon^3+(\mu_{XY}+\gamma_1+(1-\gamma_2))\epsilon^2 \\ +(\gamma_1\mu_{XY}+(1-\gamma_2)\mu_{XY}+\gamma_1(1-\gamma_2))\epsilon, \tag{10d}$$

$$\left|\frac{1}{T(T-1)}\sum_{\Pi_5} d^\alpha(Z_i,Z_j)-\gamma_1(\gamma_2-\gamma_1)\mu_{XU}\right| \\ <\epsilon^3+(\mu_{XU}+\gamma_1+(\gamma_2-\gamma_1))\epsilon^2 \\ +(\gamma_1\mu_{XU}+(\gamma_2-\gamma_1)\mu_{XU}+\gamma_1(\gamma_2-\gamma_1))\epsilon, \tag{10e}$$

$$\left|\frac{1}{T(T-1)}\sum_{\Pi_6} d^\alpha(Z_i,Z_j)-(1-\gamma_2)(\gamma_2-\gamma_1)\mu_{UY}\right| \\ <\epsilon^3+(\mu_{UY}+(1-\gamma_2)+(\gamma_2-\gamma_1))\epsilon^2+ \\ ((1-\gamma_2)\mu_{UY}+(\gamma_2-\gamma_1)\mu_{UY}+(1-\gamma_2)(\gamma_2-\gamma_1))\epsilon. \tag{10f}$$

At final, adding the inequalities (10a),(10b),(10c) and twice of (10d),(10e),(10f) and applying triangle inequality, we have the result that with probability 1 $\forall\epsilon>0$, $\exists N\in\mathbf{N}$, where $N=N_1\vee N_2\vee N_3\vee N_4\vee N_5\vee N_6\vee N_7$, such that for $T>N$,

$$\Bigg|\binom{T}{2}^{-1}\sum_{1<i<j<T} d^\alpha(Z_i,Z_j)-({\gamma_1}^2\mu_{XX} \\ +(\gamma_2-\gamma_1)^2\mu_{UU}+(1-\gamma_2)^2\mu_{YY} \\ +2\gamma_1(\gamma_2-\gamma_1)\mu_{XY}+2\gamma_1(1-\gamma_2)\mu_{XY} \\ +2(1-\gamma_2)(\gamma_2-\gamma_1)\mu_{UY})\Bigg|<\epsilon.$$

Such that, we have proved $\binom{T}{2}^{-1}\sum_{1<i<j<T} d^\alpha(Z_i,Z_j)\xrightarrow{a.s.}{\gamma_1}^2\mu_{XX}+(\gamma_2-\gamma_1)^2\mu_{UU}+(1-\gamma_2)^2\mu_{YY}+2\gamma_1(\gamma_2-\gamma_1)\mu_{XY}+2\gamma_1(1-\gamma_2)\mu_{XY}+2(1-\gamma_2)(\gamma_2-\gamma_1)\mu_{UY}$, uniformly in γ_A and γ_B. □

Proof of Lemma 4. As the first part of $\eta(\gamma_1,\gamma_2)$, $2\mu_{XU}-\mu_{UU}-\mu_{XX}$ is derived to be:

$$\begin{aligned}
&2\mu_{XY}-\mu_{XX}-\mu_{YY}\\
=&2E(X-Y)^2-E(X-X^{'})^2-E(Y-Y^{'})^2 \qquad (11)\\
=&2\big((E(X-Y))^2+Var(X-Y)\big)\\
&-\big((E(X-X^{'}))^2+Var(X-X^{'})\big)-\big((E(Y-Y^{'}))^2+Var(Y-Y^{'})\big)\\
=&2\big((\mu_X-\mu_Y)^2+\sigma_X^2+\sigma_Y^2\big)-\big((\mu_X-\mu_X)^2+\sigma_X^2+\sigma_X^2\big)-\big((\mu_Y-\mu_Y)^2+\sigma_Y^2+\sigma_Y^2\big)\\
=&2(\mu_X-\mu_Y)^2.
\end{aligned}$$

Applied the same steps in the rest part of $\eta(\gamma_1,\gamma_2)$ and $\varphi(\gamma_1,\gamma_2)$, we have

$$\eta(\gamma_1,\gamma_2)=2\left((\mu_X-\mu_Y)+\gamma_1(\mu_U-\mu_X)-\gamma_2(\mu_U-\mu_Y)\right)^2,$$

$$\varphi(\gamma_1,\gamma_2)=2\left(\gamma_1(\mu_X-\mu_U)+\gamma_2(\mu_U-\mu_Y)\right)^2$$

respectively. Obviously, both $\eta(\gamma_1,\gamma_2)$ and $\varphi(\gamma_1,\gamma_2)$ are non-negative in which the equality holds when $\mu_X=\mu_Y=\mu_U$.

□

Proof of Lemma 5. We use $\mathcal{X}$ and $\mathcal{Y}$ to denote two probability distributions in Ω, such that as the first part of $\eta(\gamma_1,\gamma_2)$, $2\mu_{XU}-\mu_{UU}-\mu_{XX}$ is derived to be:

$$\begin{aligned}
&2\mu_{XY}-\mu_{XX}-\mu_{YY}\\
=&2E\int_0^1\big(\mathcal{X}^{-1}(t)-\mathcal{Y}^{-1}(t)\big)^2dt-E\int_0^1\big(\mathcal{X}^{-1}(t)-\mathcal{X}^{'-1}(t)\big)^2dt-E\int_0^1\big(\mathcal{Y}^{-1}(t)-\mathcal{Y}^{'-1}(t)\big)^2dt\\
\overset{Fubini}{=}&2\int_0^1E\big(\mathcal{X}^{-1}(t)-\mathcal{Y}^{-1}(t)\big)^2dt-\int_0^1E\big(\mathcal{X}^{-1}(t)-\mathcal{X}^{'-1}(t)\big)^2dt-\int_0^1E\big(\mathcal{Y}^{-1}(t)-\mathcal{Y}^{'-1}(t)\big)^2dt\\
=&\int_0^1 2E\big(\mathcal{X}^{-1}(t)-\mathcal{Y}^{-1}(t)\big)^2-E\big(\mathcal{X}^{-1}(t)-\mathcal{X}^{'-1}(t)\big)^2-E\big(\mathcal{Y}^{-1}(t)-\mathcal{Y}^{'-1}(t)\big)^2dt.
\end{aligned}$$

For a fixed $t\in(0,1)$, $\mathcal{X}^{-1}(t)$ can be considered as a random variable denoted as X. Similarly, $\mathcal{X}^{'-1}(t)$, $\mathcal{Y}^{-1}(t)$ and $\mathcal{Y}^{'-1}(t)$ can be considered as random variables $X^{'}$, Y and $Y^{'}$ respectively, such that the formula inside the integral becomes:

$$2E(X-Y)^2-E(X-X^{'})^2-E(Y-Y^{'})^2,$$

which is the same expression as Expression 11 we have proved in Lemma 4. Applying the same steps in the rest of $\eta(\gamma_1, \gamma_2)$ and $\varphi(\gamma_1, \gamma_2)$, the expressions of $\eta(\gamma_1, \gamma_2)$ and $\varphi(\gamma_1, \gamma_2)$ are the same as we attained in Lemma 4 for a fixed t. Thus, we can conclude $\eta(\gamma_1, \gamma_2)$ and $\varphi(\gamma_1, \gamma_2)$ are non-negative after integral t from 0 to 1.

□

Proof of Proposition. We first show how to get the expression of $f(\tilde{\gamma})$ shown in Equation 4.

If $\gamma_1 < \tilde{\gamma} \leq \gamma_2$, the random object from cluster $\mathbf{Z}_1^{\lfloor \tilde{\gamma} T \rfloor}$ follows a mixture distribution $\frac{\gamma_1}{\tilde{\gamma}} F_L + \frac{\tilde{\gamma}-\gamma_1}{\tilde{\gamma}} F_C$, while the random object from cluster $\mathbf{Z}_{\lfloor \tilde{\gamma} T \rfloor + 1}^T$ follows the other mixture distribution of $\frac{\gamma_2-\tilde{\gamma}}{1-\tilde{\gamma}} F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}} F_R$. Hence, the expected distance between two clusters is

$$\begin{aligned}
&\left(\frac{\gamma_1}{\tilde{\gamma}}\right)\left(\frac{\gamma_2-\tilde{\gamma}}{1-\tilde{\gamma}}\right)\mu_{XU} + \left(\frac{\gamma_1}{\tilde{\gamma}}\right)\left(\frac{1-\gamma_2}{1-\tilde{\gamma}}\right)\mu_{XY} \\
&+ \left(\frac{\tilde{\gamma}-\gamma_1}{\tilde{\gamma}}\right)\left(\frac{\gamma_2-\tilde{\gamma}}{1-\tilde{\gamma}}\right)\mu_{UU} + \left(\frac{\tilde{\gamma}-\gamma_1}{\tilde{\gamma}}\right)\left(\frac{1-\gamma_2}{1-\tilde{\gamma}}\right)\mu_{UY}.
\end{aligned} \tag{12}$$

the expected distance within cluster $\mathbf{Z}_1^{\lfloor \tilde{\gamma} T \rfloor}$ is

$$\left(\frac{\gamma_1}{\tilde{\gamma}}\right)^2 \mu_{XX} + \left(\frac{\tilde{\gamma}-\gamma_1}{\tilde{\gamma}}\right)^2 \mu_{UU} + 2\left(\frac{\gamma_1}{\tilde{\gamma}}\right)\left(\frac{\tilde{\gamma}-\gamma_1}{\tilde{\gamma}}\right)\mu_{XU}, \tag{13}$$

and the expected distance within cluster $\mathbf{Z}_{\lfloor \tilde{\gamma} T \rfloor + 1}^T$ is

$$\left(\frac{\gamma_2-\tilde{\gamma}}{1-\tilde{\gamma}}\right)^2 \mu_{UU} + \left(\frac{1-\gamma_2}{1-\tilde{\gamma}}\right)^2 \mu_{YY} + 2\left(\frac{1-\gamma_2}{1-\tilde{\gamma}}\right)\left(\frac{\gamma_2-\tilde{\gamma}}{1-\tilde{\gamma}}\right)\mu_{UY}. \tag{14}$$

Taking twice of expression (12) subtracted by expression (13) and (14), the energy distance yields some tedious manipulation as following

$$\begin{aligned}
\mathcal{E}\left(\frac{\gamma_1}{\tilde{\gamma}} F_L + \frac{\tilde{\gamma}-\gamma_1}{\tilde{\gamma}} F_C, \frac{\gamma_2-\tilde{\gamma}}{1-\tilde{\gamma}} F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}} F_R\right) &= \left(\frac{\gamma_1}{\tilde{\gamma}}\right)^2 [2\mu_{XU} - \mu_{XX} - \mu_{UU}] \\
&+ \left(\frac{1-\gamma_2}{1-\tilde{\gamma}}\right)^2 [2\mu_{UY} - \mu_{UU} - \mu_{YY}] \\
&+ 2\left(\frac{\gamma_1}{\tilde{\gamma}}\right)\left(\frac{1-\gamma_2}{1-\tilde{\gamma}}\right)[\mu_{XY} + \mu_{UU} - \mu_{XU} - \mu_{UY}].
\end{aligned}$$

If $\tilde{\gamma} \leq \gamma_1 < \gamma_2$, the random object from cluster $\mathbf{Z}_1^{\lfloor \tilde{\gamma} T \rfloor}$ follows distribution F_L, while the random object from cluster $\mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^{T}$ follows a mixture distribution of $\frac{\gamma_1-\tilde{\gamma}}{1-\tilde{\gamma}}F_L+\frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}F_C+\frac{1-\gamma_2}{1-\tilde{\gamma}}F_R$. Thus, the energy distance regrading of sets $\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}$ and $\mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^{T}$ is derived as

$$
\begin{aligned}
&\mathcal{E}(F_L, \frac{\gamma_1-\tilde{\gamma}}{1-\tilde{\gamma}}F_L+\frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}F_C+\frac{1-\gamma_2}{1-\tilde{\gamma}}F_R)\\
=&2\left(\frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}\right)\left(\frac{1-\gamma_1}{1-\tilde{\gamma}}\right)\mu_{XU}-\left(\frac{1-\gamma_1}{1-\tilde{\gamma}}\right)^2\mu_{XX}-\left(\frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}\right)^2\mu_{UU}\\
&+2\left(\frac{1-\gamma_2}{1-\tilde{\gamma}}\right)\left(\frac{1-\gamma_1}{1-\tilde{\gamma}}\right)\mu_{XY}-2\left(\frac{1-\gamma_2}{1-\tilde{\gamma}}\right)\left(\frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}\right)\mu_{UY}-\left(\frac{1-\gamma_2}{1-\tilde{\gamma}}\right)^2\mu_{YY}.
\end{aligned}
$$

If $\gamma_1<\gamma_2<\tilde{\gamma}$, the the random object from cluster $\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}$ follow a mixture distribution of $\frac{\gamma_1-\tilde{\gamma}}{1-\tilde{\gamma}}F_L+\frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}F_C+\frac{1-\gamma_2}{1-\tilde{\gamma}}F_R$, while the random object from cluster $\mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^{T}$ follow distribution F_R, such that the energy distance between $\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}$ and $\mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^{T}$ is calculated and simplified as following

$$
\begin{aligned}
\mathcal{E}(\frac{\gamma_1-\tilde{\gamma}}{1-\tilde{\gamma}}F_L+\frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}F_C+\frac{1-\gamma_2}{1-\tilde{\gamma}}F_R, F_R)=&2\frac{\gamma_1\gamma_2}{\tilde{\gamma}^2}\mu_{XY}+2\left(\frac{\gamma_2-\gamma_1}{\tilde{\gamma}}\right)\left(\frac{\gamma_2}{\tilde{\gamma}}\right)\mu_{UY}\\
&-2\left(\frac{\gamma_1}{\tilde{\gamma}}\right)\left(\frac{\gamma_2-\gamma_1}{\tilde{\gamma}}\right)\mu_{XU}-\left(\frac{\gamma_1}{\tilde{\gamma}}\right)^2\mu_{XX}\\
&-\left(\frac{\gamma_2}{\tilde{\gamma}}\right)^2\mu_{YY}-\left(\frac{\gamma_2-\gamma_1}{\tilde{\gamma}}\right)^2\mu_{UU}.
\end{aligned}
$$

Above all, the scaled energy distance is expressed as a piecewise function $f(\tilde{\gamma})$ in Equation 4.

When $\gamma_1<\tilde{\gamma}\leq\gamma_2$, the second derivative of $\tilde{\gamma}(1-\tilde{\gamma})\mathcal{E}(\frac{\gamma_1}{\tilde{\gamma}}F_L+\frac{\tilde{\gamma}-\gamma_1}{\tilde{\gamma}}F_C, \frac{\gamma_2-\tilde{\gamma}}{1-\tilde{\gamma}}F_C+\frac{1-\gamma_2}{1-\tilde{\gamma}}F_R)$ is shown as below:

$$
\begin{aligned}
&\frac{d^2}{d\tilde{\gamma}^2}\tilde{\gamma}(1-\tilde{\gamma})\mathcal{E}(\frac{\gamma_1}{\tilde{\gamma}}F_L+\frac{\tilde{\gamma}-\gamma_1}{\tilde{\gamma}}F_C, \frac{\gamma_2-\tilde{\gamma}}{1-\tilde{\gamma}}F_C+\frac{1-\gamma_2}{1-\tilde{\gamma}}F_R)\\
=&(\gamma_1)^2\left(\frac{2(1-\tilde{\gamma})}{\tilde{\gamma}^3}+\frac{2}{\tilde{\gamma}^2}\right)(2\mu_{XU}-\mu_{XX}-\mu_{UU})\\
&+(1-\gamma_2)^2\left(\frac{2}{(1-\tilde{\gamma})^2}+\frac{2\tilde{\gamma}}{(1-\tilde{\gamma})^3}\right)(2\mu_{UY}-\mu_{UU}-\mu_{YY}).
\end{aligned}
$$

Due to the fact of $\tilde{\gamma}\in(0,1)$, the two terms $\frac{2(1-\tilde{\gamma})}{\tilde{\gamma}^3}+\frac{2}{\tilde{\gamma}^2}$ and $\frac{2}{(1-\tilde{\gamma})^2}+\frac{2\tilde{\gamma}}{(1-\tilde{\gamma})^3}$ are positive. Also, knowing that d is a negative definite kernel, $2\mu_{XU}-\mu_{XX}-\mu_{UU}$ and $2\mu_{UY}-\mu_{UU}-\mu_{YY}$ are positive (Rachev et al. (2013)). In conclusion, $\tilde{\gamma}(1-\tilde{\gamma})\mathcal{E}(\frac{\gamma_1}{\tilde{\gamma}}F_L+\frac{\tilde{\gamma}-\gamma_1}{\tilde{\gamma}}F_C, \frac{\gamma_2-\tilde{\gamma}}{1-\tilde{\gamma}}F_C+\frac{1-\gamma_2}{1-\tilde{\gamma}}F_R)$ is a strictly convex function and maximized at the endpoint γ_1 or γ_2 because of the positiveness of the second derivative.

When $\tilde{\gamma} \leq \gamma_1 < \gamma_2$, the first derivative of the scaled energy distance is

$$\frac{d}{d\tilde{\gamma}}\tilde{\gamma}(1-\tilde{\gamma})\mathcal{E}(F_L, \frac{\gamma_1 - \tilde{\gamma}}{1-\tilde{\gamma}}F_L + \frac{\gamma_2 - \gamma_1}{1-\tilde{\gamma}}F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}}F_R) = \frac{1}{(1-\tilde{\gamma})^2}\eta(\gamma_1, \gamma_2),$$

where $\eta(\gamma_1, \gamma_2)$ is defined in Assumption 3. The second derivative of the scaled energy distance is

$$\frac{d^2}{d\tilde{\gamma}^2}\tilde{\gamma}(1-\tilde{\gamma})\mathcal{E}(F_L, \frac{\gamma_1 - \tilde{\gamma}}{1-\tilde{\gamma}}F_L + \frac{\gamma_2 - \gamma_1}{1-\tilde{\gamma}}F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}}F_R) = \frac{2}{(1-\tilde{\gamma})^3}\eta(\gamma_1, \gamma_2).$$

Under Assumption 3, $\eta(\gamma_1, \gamma_2) > 0$, therefore, both the first and the second derivative of the scaled energy distance are greater than 0. we can conclude that $\tilde{\gamma}(1-\tilde{\gamma})\mathcal{E}(F_L, \frac{\gamma_1-\tilde{\gamma}}{1-\tilde{\gamma}}F_L + \frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}}F_R)$ is increasing and strictly convex function.

When $\gamma_1 < \gamma_2 \leq \tilde{\gamma}$, the first derivative of the scaled energy distance is

$$\frac{d}{d\tilde{\gamma}}\tilde{\gamma}(1-\tilde{\gamma})\mathcal{E}(\frac{\gamma_1 - \tilde{\gamma}}{1-\tilde{\gamma}}F_L + \frac{\gamma_2 - \gamma_1}{1-\tilde{\gamma}}F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}}F_R, F_R) = \frac{-1}{\tilde{\gamma}^2}\varphi(\gamma_1, \gamma_2),$$

where $\varphi(\gamma_1, \gamma_2)$ is defined in Assumption 4. The second derivative of the scaled energy distance is

$$\frac{d^2}{d\tilde{\gamma}^2}\tilde{\gamma}(1-\tilde{\gamma})\mathcal{E}(\frac{\gamma_1 - \tilde{\gamma}}{1-\tilde{\gamma}}F_L + \frac{\gamma_2 - \gamma_1}{1-\tilde{\gamma}}F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}}F_R, F_R) = \frac{2}{\tilde{\gamma}^3}\varphi(\gamma_1, \gamma_2).$$

Under Assumption 4, $\varphi(\gamma_1, \gamma_2) > 0$, therefore, the first derivative of the scaled energy distance is less than 0 while the second derivative is greater than 0, such that $\tilde{\gamma}(1-\tilde{\gamma})\mathcal{E}(F_L, \frac{\gamma_1-\tilde{\gamma}}{1-\tilde{\gamma}}F_L + \frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}}F_R)$ is decreasing and strictly convex function.

Above all, we have showed that when $\tilde{\gamma} \leq \gamma_1 < \gamma_2$ $\tilde{\gamma}(1-\tilde{\gamma})\mathcal{E}(F_L, \frac{\gamma_1-\tilde{\gamma}}{1-\tilde{\gamma}}F_L + \frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}}F_R)$ is increasing, when $\gamma_1 < \tilde{\gamma} \leq \gamma_2$ $\tilde{\gamma}(1-\tilde{\gamma})\mathcal{E}(\frac{\gamma_1}{\tilde{\gamma}}F_L + \frac{\tilde{\gamma}-\gamma_1}{\tilde{\gamma}}F_C, \frac{\gamma_2-\tilde{\gamma}}{1-\tilde{\gamma}}F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}}F_R)$ is a strictly convex function, as well as $\tilde{\gamma}(1-\tilde{\gamma})\mathcal{E}(\frac{\gamma_1-\tilde{\gamma}}{1-\tilde{\gamma}}F_L + \frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}}F_R, F_R)$ is decreasing when $\gamma_1 < \gamma_2 \leq \tilde{\gamma}$. Moreover, it is easy to show that $f(\tilde{\gamma})$ is continuous at γ_1 and γ_2. Hence, $f(\tilde{\gamma})$ is maximized when either $\tilde{\gamma} = \gamma_1$ or $\tilde{\gamma} = \gamma_2$. □

Proof of . Due to the composition of $f(\tilde{\gamma})$, the proof is consist of three parts.

If $\gamma_1 < \tilde{\gamma} \leq \gamma_2$, we can observe that the average distance of observations in set $\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}$ is convergent uniformly to statement (13) and the average distance of observations in set $\mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^{T}$ is convergent uniformly to statement (14) according to Lemma 1 on set $\tilde{\gamma} \in [\delta_T, 1-\delta_T]$. Moreover, the average distance between observations from $\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}$ and

$\mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^{T}$ converges uniformly to statement (12) according to the note of Lemma 1. Using twice between distance subtracted by two within distances, we will have the fact that $\hat{\mathcal{E}}_2(\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}, \mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^{T})$ converges uniformly to $\mathcal{E}(\frac{\gamma_1}{\tilde{\gamma}}F_L + \frac{\tilde{\gamma}-\gamma_1}{\tilde{\gamma}}F_C, \frac{\gamma_2-\tilde{\gamma}}{1-\tilde{\gamma}}F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}}F_R)$ on set $\tilde{\gamma} \in [\delta_T, 1-\delta_T]$.

When $\tilde{\gamma} \leq \gamma_1 \leq \gamma_2$, the within distance for $\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}$ is μ_{XX} and within distance for $\mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^{T}$ is:

$$\begin{aligned}&\left(\frac{\gamma_1-\tilde{\gamma}}{1-\tilde{\gamma}}\right)^2\mu_{XX} + 2\left(\frac{\gamma_1-\tilde{\gamma}}{1-\tilde{\gamma}}\right)\left(\frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}\right)\mu_{XU} + \left(\frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}\right)^2\mu_{UU}\\ &+ 2\left(\frac{1-\gamma_2}{1-\tilde{\gamma}}\right)\left(\frac{\gamma_1-\tilde{\gamma}}{1-\tilde{\gamma}}\right)\mu_{XY} + \left(\frac{1-\gamma_2}{1-\tilde{\gamma}}\right)^2\mu_{YY} + 2\left(\frac{1-\gamma_2}{1-\tilde{\gamma}}\right)\left(\frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}\right)\mu_{UY}, \end{aligned} \tag{15}$$

between distance for $\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}$ and $\mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^{T}$ is $\frac{\gamma_1-\tilde{\gamma}}{1-\tilde{\gamma}}\mu_{XX} + \frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}\mu_{XU} + \frac{1-\gamma_2}{1-\tilde{\gamma}}\mu_{XY}$. By Lemma 3, the within distance for $\mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^{T}$ converges uniformly to statement (15), and it is easily shown the convergence of within distance for $\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}$ and the between distance. Using twice between distance subtract by two within distances, we will have $\hat{\mathcal{E}}_1(\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}, \mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^{T})$ converges uniformly to $\mathcal{E}(F_L, \frac{\gamma_1-\tilde{\gamma}}{1-\tilde{\gamma}}F_L + \frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}}F_R)$.

When $\gamma_1 \leq \gamma_2 \leq \tilde{\gamma}$, the proof is analogous to $\mathcal{E}(F_L, \frac{\gamma_1-\tilde{\gamma}}{1-\tilde{\gamma}}F_L + \frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}}F_R)$, so is omitted. Above all, combining $\mathcal{E}(F_L, \frac{\gamma_1-\tilde{\gamma}}{1-\tilde{\gamma}}F_L + \frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}}F_R)$, $\mathcal{E}(\frac{\gamma_1}{\tilde{\gamma}}F_L + \frac{\tilde{\gamma}-\gamma_1}{\tilde{\gamma}}F_C, \frac{\gamma_2-\tilde{\gamma}}{1-\tilde{\gamma}}F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}}F_R)$ and $\mathcal{E}(\frac{\gamma_1-\tilde{\gamma}}{1-\tilde{\gamma}}F_L + \frac{\gamma_2-\gamma_1}{1-\tilde{\gamma}}F_C + \frac{1-\gamma_2}{1-\tilde{\gamma}}F_R, F_R)$, we will have the conclusion that $\hat{\mathcal{E}}(\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}, \mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^{T})$ is convergent uniformly to $f(\tilde{\gamma})$ on set $\tilde{\gamma} \in [\delta_T, 1-\delta_T]$. Observed that $\frac{1}{T}\hat{Q}(\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}, \mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^{T}) = \frac{(\lfloor\tilde{\gamma}T\rfloor)(T-\lfloor\tilde{\gamma}T\rfloor)}{T^2}\hat{\mathcal{E}}(\mathbf{Z}_1^{\lfloor\tilde{\gamma}T\rfloor}, \mathbf{Z}_{\lfloor\tilde{\gamma}T\rfloor+1}^{T})$, Lemma 6 is proved. $\square$

REFERENCES

Hoeffding, W. (1961). The strong law of large numbers for u-statistics. Technical report, North Carolina State University. Dept. of Statistics.

Rachev, S. T., L. B. Klebanov, S. V. Stoyanov, and F. J. Fabozzi (2013). Positive and negative definite kernels and their properties. In *The Methods of Distances in the Theory of Probability and Statistics*, pp. 519–537. Springer.