J. Math. Sci. Univ. Tokyo
30 (2023), 21–54.

Decision Tree-Based Estimation of the Overlap of Two Probability Distributions

By Hisashi Johno and Kazunori Nakamoto

Abstract. A new nonparametric approach, based on a decision tree algorithm, is proposed to calculate the overlap between two probability distributions. The devised framework is described analytically and numerically. The convergence of the estimated overlap to the true value is proved along with some experimental results.

1. Introduction

In various scientific fields, it is important to assess the similarity between data sets or distributions. The overlap coefficient (OVL) is an interpretable measure of such similarity, defined as the common area under two probability density functions (PDFs). While a variety of parametric techniques to estimate OVL have been developed, existing nonparametric ones are wholly based on kernel density estimation (KDE) [3, 4, 6]. Although KDE is a useful and widely practiced method to estimate probability density functions, the optimal setting of its parameters (kernel function and bandwidth) is a challenging task.

Here we propose a new nonparametric method to calculate OVL based on a decision tree algorithm. We start with notation and preliminaries in Section 2. The devised framework is described analytically in Section 3 and numerically in Section 4. Experimental results are shown in Section 5, and the conclusion follows in Section 6.

2020 *Mathematics Subject Classification.* 62G05.
Key words: Probability distribution, Crossover point, Overlap coefficient, Nonparametric, Decision tree.

2. Preliminaries

Let f_1 and f_2 be two continuous PDFs on the real line $\mathbb{R}$. The OVL between f_1 and f_2 is defined as

$$\rho(f_1, f_2) = \int_{-\infty}^{\infty} \min\{f_1(x), f_2(x)\}\ dx.$$

Definition 2.1. Suppose g_1 and g_2 are real continuous functions on $\mathbb{R}$. Then we call $x \in \mathbb{R}$ a *crossover point* between g_1 and g_2 if there exist points a, b in any neighborhood of x such that $[g_1(a) - g_2(a)][g_1(b) - g_2(b)] < 0$. We also call $x \in \mathbb{R}$ a *coincidence point* between g_1 and g_2 if $g_1(x) = g_2(x)$. The set of crossover points and that of coincidence points are denoted by $C(g_1, g_2)$ and $C'(g_1, g_2)$, respectively. Note that $C(g_1, g_2) \subset C'(g_1, g_2)$.

Under the assumption that $C'(f_1, f_2)$ is finite and the cardinality of $C(f_1, f_2)$ is known in advance, we present a decision tree-based method to estimate $\rho(f_1, f_2)$. The rest of this section provides further notations and terminologies.

Definition 2.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(X, Y) : \Omega \to \mathbb{R} \times \{1, 2\}$ a random variable with distribution P, defined as $P(A) = \mathbb{P}((X, Y)^{-1}(A))$ for all Borel sets $A \subset \mathbb{R} \times \{1, 2\}$. From the viewpoint of binary classification, the measurable functions $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \{1, 2\}$ can be regarded as explanatory and response variables, respectively. Given a Borel set $B \subset \mathbb{R}$, we may simply write $P(X \in B)$ for $P(B \times \{1, 2\})$, π_j for $P(\mathbb{R} \times \{j\})$, $F_j(x)$ for $P((-\infty, x] \times \{j\})/\pi_j$, $P(X \in B, Y = j)$ for $P(B \times \{j\})$, and $P(Y = j \mid X \in B)$ for $P(X \in B, Y = j)/P(X \in B)$, provided $\pi_j \neq 0$ and $P(X \in B) \neq 0$ as necessary.

We shall consider the random variable (X, Y) with

$$F_j(x) = \int_{-\infty}^{x} f_j(t)\, dt \qquad (x \in \mathbb{R};\ j = 1, 2),$$

so that each F_j is the cumulative distribution function (CDF) corresponding to the continuous PDF f_j. We also define $F_j(-\infty) = 0$ and $F_j(\infty) = 1$ $(j = 1, 2)$.

DEFINITION 2.3. Let Δ^1 be the standard 1-simplex, which consists of all points $(a, b) \in \mathbb{R}^2$ such that $a + b = 1$, $a \geq 0$, and $b \geq 0$. An *impurity function* on Δ^1 is a function ι with the following properties:

1. ι attains its maximum only at $(1/2, 1/2)$,
2. ι attains its minimum only at $(1, 0)$ and $(0, 1)$,
3. ι is a symmetric function, i.e., $\iota(a, b) = \iota(b, a)$.

DEFINITION 2.4. For a positive integer m, let $\mathbb{R}^m_{\leq}$ be the set of all $\boldsymbol{v} = (v_1, \ldots, v_m) \in \mathbb{R}^m$ with $v_1 \leq \cdots \leq v_m$. By the $(m+1)$-*ary split* on $\mathbb{R}$ at a point $\boldsymbol{v} \in \mathbb{R}^m_{\leq}$, we mean the collection $S_{\boldsymbol{v}} = \{S_{\boldsymbol{v},1}, \ldots, S_{\boldsymbol{v},m+1}\}$ with $S_{\boldsymbol{v},1} = \{x \in \mathbb{R} \mid x \leq v_1\}$, $S_{\boldsymbol{v},m+1} = \{x \in \mathbb{R} \mid x > v_m\}$, and $S_{\boldsymbol{v},k} = \{x \in \mathbb{R} \mid v_{k-1} < x \leq v_k\}$ for $k = 2, \ldots, m$. Note that each $S_{\boldsymbol{v},k}$ is a Borel set in $\mathbb{R}$, $S_{\boldsymbol{v},k} \cap S_{\boldsymbol{v},l} = \emptyset$ if $k \neq l$, and $S_{\boldsymbol{v},1} \cup \cdots \cup S_{\boldsymbol{v},m+1} = \mathbb{R}$.

Using an impurity function ι on Δ^1, we define the *impurity* of a Borel set $B \subset \mathbb{R}$ for the binary classification by

$$I(B) = \begin{cases} \iota\left(P(Y=1 \mid X \in B), P(Y=2 \mid X \in B)\right) & \text{if} \quad P(X \in B) > 0, \\ 0 & \text{if} \quad P(X \in B) = 0, \end{cases}$$

and the *goodness* of $S_{\boldsymbol{v}}$ ($\boldsymbol{v} \in \mathbb{R}^m_{\leq}$) by

$$\Delta I(S_{\boldsymbol{v}}) = I(\mathbb{R}) - \sum_{k=1}^{m+1} P(X \in S_{\boldsymbol{v},k}) I(S_{\boldsymbol{v},k}), \tag{1}$$

according to the conventional decision tree algorithm [1]. If there exists $\boldsymbol{v}' \in \mathbb{R}^m_{\leq}$ such that $\Delta I(S_{\boldsymbol{v}'}) = \sup \Delta I(S_{\boldsymbol{v}})$, where the supremum is over all $\boldsymbol{v} \in \mathbb{R}^m_{\leq}$, then we call $S_{\boldsymbol{v}'}$ a *best* $(m+1)$-*ary split* on $\mathbb{R}$.

3. Analytical Framework

In this section, we present the theoretical foundation of our method to calculate $C(\pi_1 f_1, \pi_2 f_2)$ and $\rho(\pi_1 f_1, \pi_2 f_2)$ under the assumptions that $C'(\pi_1 f_1, \pi_2 f_2)$ is finite, the cardinality n of $C(\pi_1 f_1, \pi_2 f_2)$ is known in advance, $\pi_1 > 0$, and $\pi_2 > 0$. We can obtain $C(f_1, f_2) = C(\pi_1 f_1, \pi_2 f_2)$ and

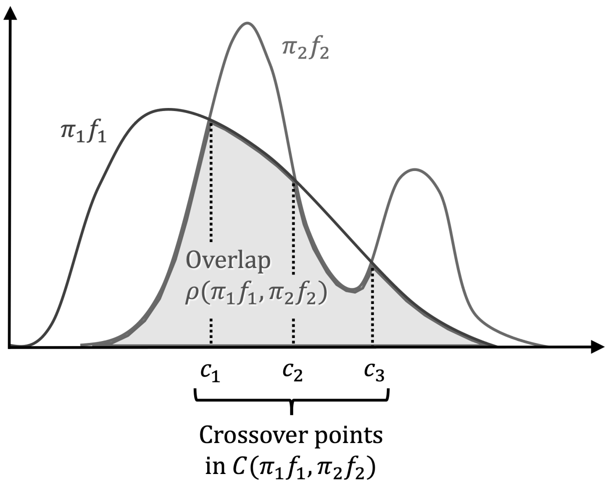

Fig. 1. A schematic example of $C(\pi_1 f_1, \pi_2 f_2)$ and $\rho(\pi_1 f_1, \pi_2 f_2)$.

$\rho(f_1, f_2) = 2\rho(\pi_1 f_1, \pi_2 f_2)$ if $\pi_1 = \pi_2 = 1/2$, which may be realized with sampling techniques, e.g., drawing the same number of samples from both the distributions corresponding to f_1 and f_2. Here we use the setting of the previous section and, in addition, adopt the misclassification-based impurity function [1], i.e.,

$$\iota(a, b) = 1 - \max\{a, b\} \qquad ((a, b) \in \Delta^1). \tag{2}$$

Suppose $C(\pi_1 f_1, \pi_2 f_2) = \emptyset$, or $n = 0$. Then either $\pi_1 f_1 \leq \pi_2 f_2$ or $\pi_1 f_1 \geq \pi_2 f_2$ holds. (Recall that $C'(\pi_1 f_1, \pi_2 f_2)$ is finite.) In the former case, we have $\rho(\pi_1 f_1, \pi_2 f_2) = \pi_1$, and in the latter, $\rho(\pi_1 f_1, \pi_2 f_2) = \pi_2$. Of note, $\pi_1 = \pi_2 = 1/2$ cannot occur here.

In the following, we assume $C(\pi_1 f_1, \pi_2 f_2) \neq \emptyset$, so that n is a positive integer. Put $C(\pi_1 f_1, \pi_2 f_2) = \{c_1, \ldots, c_n\}$ with $c_1 < \cdots < c_n$, $\boldsymbol{c} = (c_1, \ldots, c_n)$, $c_0 = -\infty$, and $c_{n+1} = \infty$. The $(n+1)$-ary split on $\mathbb{R}$ at $\boldsymbol{c}$ is defined by $S_{\boldsymbol{c}} = \{S_{\boldsymbol{c},1}, \ldots, S_{\boldsymbol{c},n+1}\}$ (see Definition 2.4). Figure 1 is a schematic example to illustrate $C(\pi_1 f_1, \pi_2 f_2)$ and $\rho(\pi_1 f_1, \pi_2 f_2)$.

PROPOSITION 3.1. *For $\boldsymbol{v} = (v_1, \ldots, v_m) \in \mathbb{R}^m_{\leq}$ with m a positive inte-*

ger, we have

$$\begin{aligned}\Delta I(S_{\boldsymbol{v}}) &= \sum_{k=1}^{m+1} \max_j \{\pi_j\,[F_j(v_k) - F_j(v_{k-1})]\} - \max_j\{\pi_j\} \\ &= \sum_{k=1}^{m+1} \max_j \left\{ \int_{v_{k-1}}^{v_k} \pi_j f_j(x)\,dx \right\} - \max_j\{\pi_j\}\,,\end{aligned}$$

where $v_0 = -\infty$ and $v_{m+1} = \infty$.

Proof. From (1) and (2), we have

$$\Delta I(S_{\boldsymbol{v}}) = \sum_k P(X \in S_{\boldsymbol{v},k}) \max_j \{P(Y = j \mid X \in S_{\boldsymbol{v},k})\} - \max_j \{P(Y = j \mid X \in \mathbb{R})\}\,, \tag{3}$$

where the sum is over all k with $P(X \in S_{\boldsymbol{v},k}) > 0$. Since $P(Y = j \mid X \in S_{\boldsymbol{v},k}) = P(S_{\boldsymbol{v},k} \times \{j\})/P(X \in S_{\boldsymbol{v},k})$ and $P(S_{\boldsymbol{v},k} \times \{j\}) = \pi_j[F_j(v_k) - F_j(v_{k-1})]$, we obtain

$$P(X \in S_{\boldsymbol{v},k}) \max_j \{P(Y = j \mid X \in S_{\boldsymbol{v},k})\} = \max_j \{\pi_j\,[F_j(v_k) - F_j(v_{k-1})]\}\,.$$

As for the last term of (3), we have $P(Y = j \mid X \in \mathbb{R}) = \pi_j$ by definition. □

The following corollary is immediate from Proposition 3.1.

Corollary 3.2. *For $\boldsymbol{v} = (v_1, \ldots, v_m) \in \mathbb{R}^m_{\le}$ with m a positive integer, let*

$$g_{\boldsymbol{v}}(x) = \pi_{j_k} f_{j_k}(x) \qquad (x \in S_{\boldsymbol{v},k};\ k = 1, \ldots, m+1)$$

where $j_k \in \arg\max_j \{\pi_j[F_j(v_k) - F_j(v_{k-1})]\}$. Let $g = \max\{\pi_1 f_1, \pi_2 f_2\}$. Then

$$\Delta I(S_{\boldsymbol{v}}) = \int_{-\infty}^{\infty} g_{\boldsymbol{v}}(x)\,dx - \max_j\{\pi_j\}\,, \quad \Delta I(S_{\boldsymbol{c}}) = \int_{-\infty}^{\infty} g(x)\,dx - \max_j\{\pi_j\}\,.$$

Furthermore, $g_{\boldsymbol{v}} \le g$ and $\Delta I(S_{\boldsymbol{v}}) \le \Delta I(S_{\boldsymbol{c}})$.

LEMMA 3.3. *Suppose m is a positive integer, $\boldsymbol{v} = (v_1, \dots, v_m) \in \mathbb{R}^m_{\leq}$, $v_0 = -\infty$, and $v_{m+1} = \infty$. If $v_{k-1} < c_p < v_k$ for some $k \in \{1, \dots, m+1\}$ and $p \in \{1, \dots, n\}$, then $\Delta I(S_{\boldsymbol{v}}) < \Delta I(S_{\boldsymbol{c}})$.*

PROOF. Since $C'(\pi_1 f_1, \pi_2 f_2)$ is finite, there exists a neighborhood U of c_p such that $U \subset (v_{k-1}, v_k)$ and $U \cap C'(\pi_1 f_1, \pi_2 f_2) = \{c_p\}$. Then, $[\pi_1 f_1(a) - \pi_2 f_2(a)][\pi_1 f_1(b) - \pi_2 f_2(b)] < 0$ for all $a, b \in U$ with $a < c_p < b$. Without loss of generality, we assume that $\pi_1 f_1(a) < \pi_2 f_2(a)$ and $\pi_2 f_2(b) < \pi_1 f_1(b)$. If $g_{\boldsymbol{v}} = \pi_1 f_1$ on $S_{\boldsymbol{v},k}$, then $g_{\boldsymbol{v}} < g$ on the open interval (a, c_p), so that

$$\Delta I(S_{\boldsymbol{c}}) - \Delta I(S_{\boldsymbol{v}}) \geq \int_a^{c_p} [g(x) - g_{\boldsymbol{v}}(x)]\ dx > 0.$$

The proof for the case $g_{\boldsymbol{v}} = \pi_2 f_2$ on $S_{\boldsymbol{v},k}$ is similar. □

PROPOSITION 3.4. *The supremum of $\Delta I(S_{\boldsymbol{v}})$ over $\boldsymbol{v} \in \mathbb{R}^n_{\leq}$ is uniquely attained at $\boldsymbol{v} = \boldsymbol{c}$.*

In other words, $S_{\boldsymbol{c}}$ is the unique best $(n+1)$-ary split on $\mathbb{R}$.

PROOF. If $\boldsymbol{v} \neq \boldsymbol{c}$, then $c_p \notin \{v_1, \dots, v_n\}$ for some p. Hence $v_{k-1} < c_p < v_k$ for some k as in the assumption of Lemma 3.3, so that $\Delta I(S_{\boldsymbol{v}}) < \Delta I(S_{\boldsymbol{c}})$. □

PROPOSITION 3.5. *Suppose m is a positive integer with $m < n$. Then for every $\boldsymbol{v} \in \mathbb{R}^m_{\leq}$, $\Delta I(S_{\boldsymbol{v}}) < \Delta I(S_{\boldsymbol{c}})$.*

PROOF. Since $m < n$, $c_p \notin \{v_1, \dots, v_m\}$ for some p. The proof is similar as above. □

Now we see that $C(\pi_1 f_1, \pi_2 f_2)$ can be obtained by finding $\boldsymbol{v} \in \mathbb{R}^n_{\leq}$ that yields the maximum of $\Delta I(S_{\boldsymbol{v}})$. Given $C(\pi_1 f_1, \pi_2 f_2)$, we have

$$\begin{aligned} \rho(\pi_1 f_1, \pi_2 f_2) &= \sum_{k=1}^{n+1} \int_{c_{k-1}}^{c_k} \min_j \{\pi_j f_j(x)\}\ dx \\ &= \sum_{k=1}^{n+1} \min_j \{P(X \in S_{\boldsymbol{c},k}, Y = j)\}. \end{aligned} \tag{4}$$

4. Numerical Framework

Here we show how to estimate $C(\pi_1 f_1, \pi_2 f_2)$ and $\rho(\pi_1 f_1, \pi_2 f_2)$, given independent and identically distributed (i.i.d.) random variables $(X_1, Y_1), \dots,$ (X_N, Y_N) with the distribution P on $\mathbb{R} \times \{1, 2\}$. Let us keep the setting of the previous section.

Definition 4.1. For a Borel set $B \subset \mathbb{R}$ and $j \in \{1, 2\}$, put

$$\begin{aligned} N_X(B) = \#\{i \mid X_i \in B\}, \quad N_Y(j) = \#\{i \mid Y_i = j\} \\ N_{XY}(B, j) = \#\{i \mid X_i \in B, Y_i = j\}, \quad \hat{\pi}_{j,N} = N_Y(j)/N \\ \widehat{P}_N(X \in B) = N_X(B)/N, \quad \widehat{P}_N(X \in B, Y = j) = N_{XY}(B, j)/N, \\ \widehat{P}_N(Y = j \mid X \in B) = N_{XY}(B, j)/N_X(B) \quad \text{if} \quad \widehat{P}_N(X \in B) > 0, \end{aligned}$$

where $\#$ denotes the cardinality of a set. Define

$$\widehat{I}_N(B) = \begin{cases} \iota\Big(\widehat{P}_N(Y = 1 \mid X \in B), \widehat{P}_N(Y = 2 \mid X \in B)\Big) & \\ & \text{if} \quad \widehat{P}_N(X \in B) > 0, \\ 0 & \text{if} \quad \widehat{P}_N(X \in B) = 0 \end{cases}$$

and

$$\Delta\widehat{I}_N(S_{\boldsymbol{v}}) = \widehat{I}_N(\mathbb{R}) - \sum_{k=1}^{m+1} \widehat{P}_N(X \in S_{\boldsymbol{v},k})\widehat{I}_N(S_{\boldsymbol{v},k}) \qquad (\boldsymbol{v} \in \mathbb{R}^m_{\leq};\ m = 1, 2, \dots) \tag{5}$$

as the estimators of $I(B)$ and $\Delta I(S_{\boldsymbol{v}})$, respectively.

Definition 4.2. Let $X_{N:1} \leq \cdots \leq X_{N:N}$ be the order statistics of $X_1, \dots, X_N$,

$$\begin{aligned} & Z_i = (X_{N:i} + X_{N:i+1})/2 \qquad (i = 1, \dots, N-1), \\ & \widehat{\mathbb{R}}^m_N = \{(Z_{i_1}, \dots, Z_{i_m}) \mid 1 \leq i_1 \leq \cdots \leq i_m \leq N-1\} \qquad (m = 1, 2, \dots). \end{aligned}$$

To avoid trivialities, we set $Z_1 = X_1$ if $N = 1$. Note that $\widehat{\mathbb{R}}^m_N \subset \mathbb{R}^m_{\leq}$ $(m = 1, 2, \dots)$ and recall that $n = \#C(\pi_1 f_1, \pi_2 f_2)$. Define $\widehat{\boldsymbol{v}}_N$

$\in \arg\max_{\boldsymbol{v}\in\widehat{\mathbb{R}}_N^n}\{\varDelta\widehat{I}_N(S_{\boldsymbol{v}})\}$ and

$$(6)\quad \widehat{\rho}_{\boldsymbol{v},N}=\sum_{k=1}^{m+1}\min_j\left\{\widehat{P}_N(X\in S_{\boldsymbol{v},k},Y=j)\right\}\qquad(\boldsymbol{v}\in\mathbb{R}^m_{\leq};\ m=1,2,\dots).$$

We propose $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$ as an estimator of $\rho(\pi_1 f_1,\pi_2 f_2)$.

DEFINITION 4.3. Let ξ be a random variable and $\{\xi_i\}$ a sequence of random variables on $(\Omega,\mathcal{F},\mathbb{P})$ taking values in a separable metric space (A,d). We say that $\{\xi_i\}$ *converges almost surely* to ξ if

$$\mathbb{P}\left(\left\{\omega\in\Omega\,\middle|\,\lim_{i\to\infty}\xi_i(\omega)=\xi(\omega)\right\}\right)=1.$$

We also say that $\{\xi_i\}$ *converges completely* to ξ if

$$\sum_{i=1}^{\infty}\mathbb{P}\left(\{\omega\in\Omega\mid d\left(\xi_i(\omega),\xi(\omega)\right)>\epsilon\}\right)<\infty$$

for any $\epsilon>0$.

REMARK 4.4 (See [2] for reference). In Definition 4.3, $\{\xi_i\}$ converges almost surely to ξ if and only if

$$\lim_{l\to\infty}\mathbb{P}\left(\bigcup_{i=l}^{\infty}\{\omega\in\Omega\mid d\left(\xi_i(\omega),\xi(\omega)\right)>\epsilon\}\right)=0$$

for any $\epsilon>0$. If $\{\xi_i\}$ converges completely to ξ, then $\{\xi_i\}$ converges almost surely to ξ.

THEOREM 4.5. *As N tends to ∞, $\widehat{\boldsymbol{v}}_N$ converges completely to $\boldsymbol{c}$.*

THEOREM 4.6. *As N tends to ∞, $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$ converges completely to $\rho(\pi_1 f_1,\pi_2 f_2)$.*

The proofs of Theorems 4.5 and 4.6 are given in Appendix A. While $\widehat{\boldsymbol{v}}_N$ and $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$ are treated as random variables, their measurability is in fact nontrivial and will be discussed in Appendix B.

REMARK 4.7. For each $N = 1, 2, \ldots,$ let $(X_1^{(N)}, Y_1^{(N)}), \ldots,$ $(X_N^{(N)}, Y_N^{(N)})$ be i.i.d. random variables with the distribution P on $\mathbb{R} \times \{1, 2\}$ to calculate $\widehat{\boldsymbol{v}}_N^{(N)} \in \widehat{\mathbb{R}}_N^n$ and $\widehat{\rho}^{(N)}_{\widehat{\boldsymbol{v}}_N^{(N)},N}$ in the same way as $\widehat{\boldsymbol{v}}_N$ and $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$ in Definition 4.2, respectively. By Theorems 4.5 and 4.6, we have

$$\sum_{N=1}^{\infty} \mathbb{P}\left(\left\{\omega \in \Omega \,\middle|\, \left\|\widehat{\boldsymbol{v}}_N^{(N)} - \boldsymbol{c}\right\| > \epsilon\right\}\right) = \sum_{N=1}^{\infty} \mathbb{P}\left(\{\omega \in \Omega \mid \|\widehat{\boldsymbol{v}}_N - \boldsymbol{c}\| > \epsilon\}\right) < \infty$$

and

$$\begin{aligned}
&\sum_{N=1}^{\infty} \mathbb{P}\left(\left\{\omega \in \Omega \,\middle|\, \left|\widehat{\rho}^{(N)}_{\widehat{\boldsymbol{v}}_N^{(N)},N} - \rho(\pi_1 f_1, \pi_2 f_2)\right| > \epsilon\right\}\right) \\
&= \sum_{N=1}^{\infty} \mathbb{P}\left(\left\{\omega \in \Omega \mid \left|\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N} - \rho(\pi_1 f_1, \pi_2 f_2)\right| > \epsilon\right\}\right) < \infty
\end{aligned}$$

for any $\epsilon > 0$, where $\|\cdot\|$ denotes the Euclidean norm. Hence $\widehat{\boldsymbol{v}}_N^{(N)}$ and $\widehat{\rho}^{(N)}_{\widehat{\boldsymbol{v}}_N^{(N)},N}$, as well as $\widehat{\boldsymbol{v}}_N$ and $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$, converge completely to $\boldsymbol{c}$ and $\rho(\pi_1 f_1, \pi_2 f_2)$, respectively.

5. Numerical Experiments

Here we perform numerical simulations to illustrate the results in Section 4. A set of random samples $\{(X_i, Y_i) \mid 1 \leq i \leq N\}$ was simulated under the following two conditions: first,

$$\pi_1 = 2/3, \qquad \pi_2 = 1/3, \qquad f_1 = \nu_{-1,1}, \qquad f_2 = \nu_{1,1},$$

and second,

$$\pi_1 = \pi_2 = 0.5, \qquad f_1 = 0.5\nu_{-1,1} + 0.5\nu_{1,1}, \qquad f_2 = 0.8\nu_{0,1} + 0.2\tau_{0,0.5},$$

where $\nu_{\mu,\sigma}$ represents the Gaussian PDF defined as

$$\nu_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad (x \in \mathbb{R}) \tag{7}$$

and $\tau_{a,b}$ is the triangular PDF defined as

(8)
$$\tau_{a,b}(x) = \begin{cases} 4(x-a)/(b-a)^2 & \text{if} \quad a \le x \le (a+b)/2, \\ 4(b-x)/(b-a)^2 & \text{if} \quad (a+b)/2 < x \le b, \\ 0 & \text{otherwise.} \end{cases} \qquad (x \in \mathbb{R};\ a < b),$$

Then, we can analytically calculate

$$C(\pi_1 f_1, \pi_2 f_2) = \{c_1\} = \{(\log 2)/2\} \simeq \{0.347\}, \tag{9}$$
$$\rho(\pi_1 f_1, \pi_2 f_2) = [2 - 2\Phi(c_1+1) + \Phi(c_1 - 1)]/3 \simeq 0.145 \tag{10}$$

for the first case, and

$$C(\pi_1 f_1, \pi_2 f_2) = \{c_1, c_2\} = \cosh^{-1}\left(0.8\sqrt{\mathrm{e}}\right) \simeq \{-0.779, 0.779\}, \tag{11}$$
$$\rho(\pi_1 f_1, \pi_2 f_2) = 0.8 - 0.5\Phi(c_1+1) + 0.5\Phi(c_2+1) - 0.8\Phi(c_2) \simeq 0.362 \tag{12}$$

for the second case, where Φ denotes the cumulative distribution function of the standard normal distribution given by

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) dt \qquad (x \in \mathbb{R}). \tag{13}$$

See Appendix C for the proof of (9)–(12). With the knowledge that $n = 1$ and $n = 2$ for the first and second cases, respectively, we numerically calculated $\widehat{\boldsymbol{v}}_N$ and $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$ for each case with $N = 10000$. The subsets $\{(X_i, Y_i) \mid 1 \le i \le 100\}$ and $\{(X_i, Y_i) \mid 1 \le i \le 1000\}$ were also applied to calculate $\widehat{\boldsymbol{v}}_N$ and $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$. This trial (from the generation of 10000 random samples) was repeated independently for 30 times, and the convergence of $\widehat{\boldsymbol{v}}_N$ and $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$ was visually assessed.

To begin with, we exhibit a representative sample distribution ($N = 10000$) for each case with the calculated values of $\widehat{\boldsymbol{v}}_N$ and $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$ (Figures 2 and 3). As a result of the 30 trials for each case, $\widehat{\boldsymbol{v}}_N$ and $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$ appear to converge to $\boldsymbol{c}$ and $\rho(\pi_1 f_1, \pi_2 f_2)$, respectively, as N increases (Figures 4 and 5).

Similarly, we next performed 30 independent trials for each case to simulate three independent sets of random samples, of the forms $\{(X_i, Y_i) \mid 1 \le i \le 100\}$, $\{(X_i', Y_i') \mid 1 \le i \le 1000\}$, and $\{(X_i'', Y_i'') \mid 1 \le i \le 10000\}$.

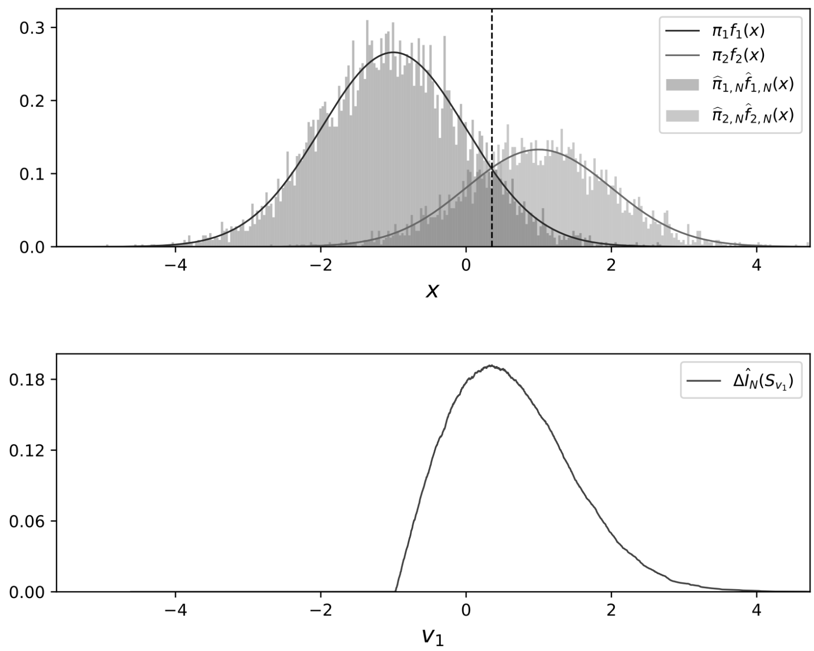

Fig. 2. In the upper row, $\pi_1 f_1$ and $\pi_2 f_2$ for the first case are plotted. Normalized histograms corresponding to $\pi_1 f_1$ and $\pi_2 f_2$ (denoted by $\widehat{\pi}_{1,N}\widehat{f}_{1,N}$ and $\widehat{\pi}_{2,N}\widehat{f}_{2,N}$, respectively) were generated using a representative set of $N = 10000$ random samples, $\{(X_i, Y_i) \mid 1 \le i \le 10000\}$. The vertical dotted line indicates the estimated crossover point $\widehat{v}_{1,N} \approx 0.355$, where its theoretical counterpart is $c_1 \simeq 0.347$. In the lower row, $\Delta\widehat{I}_N(S_{v_1})$ for all $v_1 \in \widehat{\mathbb{R}}^1_N$ are plotted. The overlap $\rho(\pi_1 f_1, \pi_2 f_2) \simeq 0.145$ was estimated as $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N, N} \approx 0.140$.

Each set was used to calculate $\widehat{\boldsymbol{v}}_N^{(N)}$ and $\widehat{\rho}^{(N)}_{\widehat{\boldsymbol{v}}_N^{(N)}, N}$ (see Remark 4.7). Then, in both the cases, $\widehat{\boldsymbol{v}}_N^{(N)}$ and $\widehat{\rho}^{(N)}_{\widehat{\boldsymbol{v}}_N^{(N)}, N}$ appear to converge to $\boldsymbol{c}$ and $\rho(\pi_1 f_1, \pi_2 f_2)$, respectively, as N increases (Figures 6 and 7).

6. Conclusion

In this paper, we propose a new nonparametric framework to calculate OVL based on a decision tree algorithm. The estimators of crossover points and overlaps for continuous PDFs were shown to converge to the expected

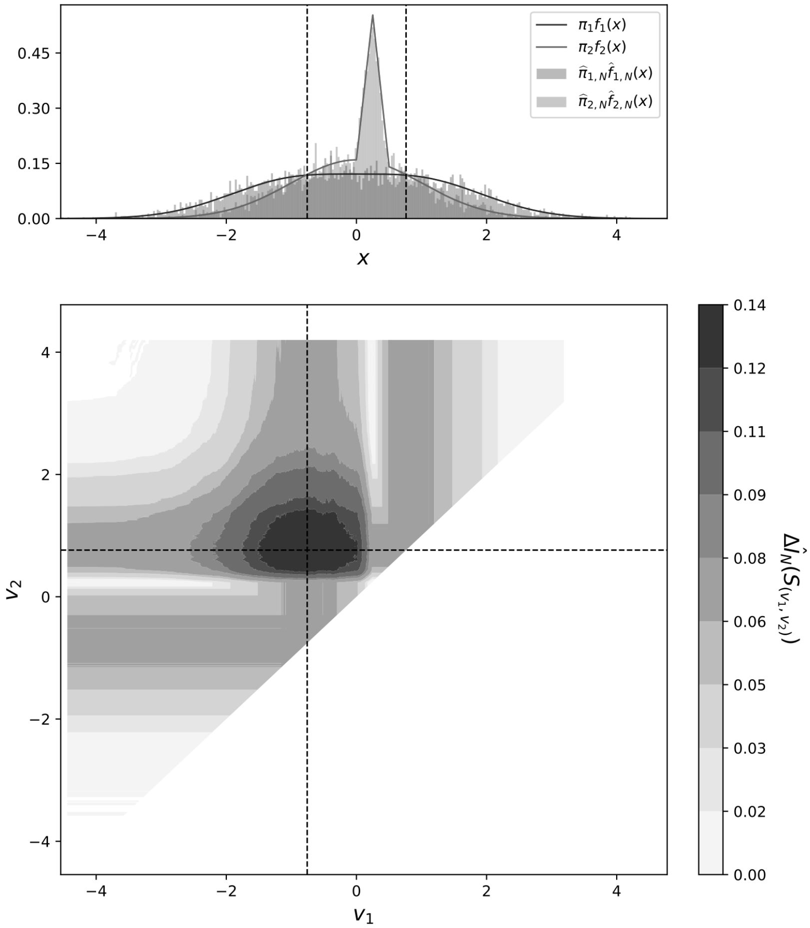

Fig. 3. In the upper row, $\pi_1 f_1$ and $\pi_2 f_2$ for the second case are plotted. Normalized histograms corresponding to $\pi_1 f_1$ and $\pi_2 f_2$ (denoted by $\widehat{\pi}_{1,N}\widehat{f}_{1,N}$ and $\widehat{\pi}_{2,N}\widehat{f}_{2,N}$, respectively) were generated using a representative set of $N = 10000$ random samples, $\{(X_i, Y_i) \mid 1 \leq i \leq 10000\}$. The dotted lines indicate the estimated crossover points $\widehat{v}_{1,N} \approx -0.757$ and $\widehat{v}_{2,N} \approx 0.763$, where their theoretical counterparts are $c_1 \simeq -0.779$ and $c_2 \simeq 0.779$, respectively. In the lower row, $\Delta\widehat{I}_N(S_{(v_1,v_2)})$ for all $(v_1, v_2) \in \widehat{\mathbb{R}}^2_N$ are visualized in a heatmap. The overlap $\rho(\pi_1 f_1, \pi_2 f_2) \simeq 0.362$ was estimated as $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N} \approx 0.361$.

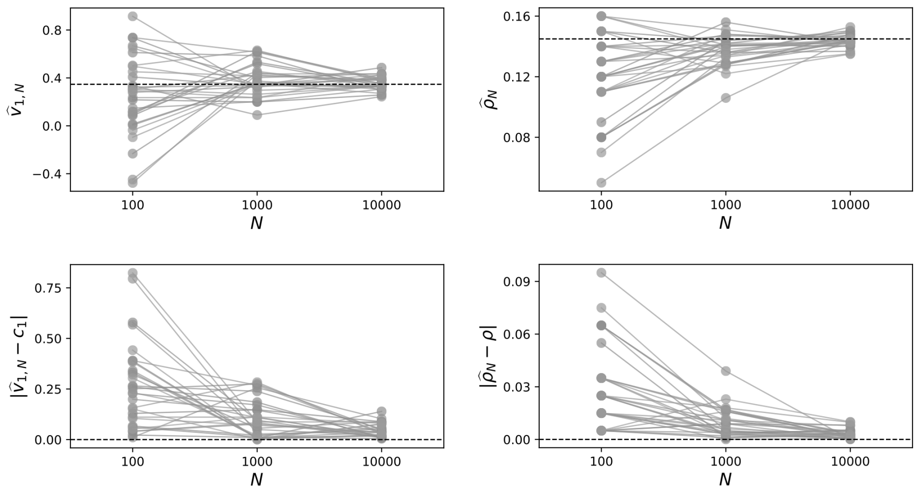

Fig. 4. In the first case, 30 independent trials were performed to simulate 10000 random samples: $(X_1, Y_1), \ldots, (X_{10000}, Y_{10000})$. For each trial, $\{(X_i, Y_i) \mid 1 \leq i \leq 100\}$, $\{(X_i, Y_i) \mid 1 \leq i \leq 1000\}$, and $\{(X_i, Y_i) \mid 1 \leq i \leq 10000\}$ were used to calculate $\widehat{v}_{1,N}$, $|\widehat{v}_{1,N} - c_1|$, $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$, and $|\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N} - \rho(\pi_1 f_1, \pi_2 f_2)|$. Each dotted line indicates the expected value: $c_1 \simeq 0.347$ for $\widehat{v}_{1,N}$, 0 for $|\widehat{v}_{1,N} - c_1|$, $\rho(\pi_1 f_1, \pi_2 f_2) \simeq 0.145$ for $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$, and 0 for $|\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N} - \rho(\pi_1 f_1, \pi_2 f_2)|$. In this figure, $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$ and $\rho(\pi_1 f_1, \pi_2 f_2)$ are abbreviated as $\widehat{\rho}_N$ and ρ, respectively.

values (both analytically and numerically). However, there remain several issues to be addressed:

1. We have not established a general way to know the number n of crossover points (which is required to be known in advance), though we may estimate it beforehand by obtaining partial information about the distributions (e.g., there exist precisely two crossover points between any two normal distributions with different variances) or by using some numerical tools like histograms.

2. Our method has not been applied to real data or compared numerically with other nonparametric methods, though the following arguments seem to exemplify the theoretical advantages of ours over the previous ones (described in detail in [6]): (i) our OVL estimator depends only on the rank statistics of $X_1, ..., X_N$ (labeled by $Y_1, \ldots, Y_N$, respectively),

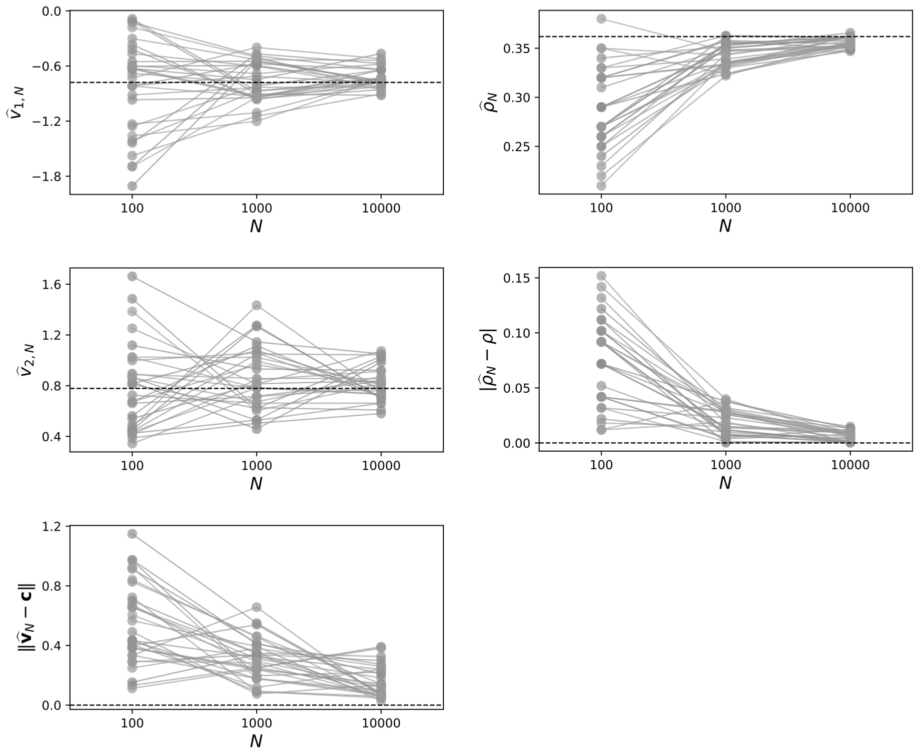

Fig. 5. In the second case, 30 independent trials were performed to simulate 10000 random samples: $(X_1, Y_1), \dots, (X_{10000}, Y_{10000})$. For each trial, $\{(X_i, Y_i) \mid 1 \le i \le 100\}$, $\{(X_i, Y_i) \mid 1 \le i \le 1000\}$, and $\{(X_i, Y_i) \mid 1 \le i \le 10000\}$ were used to calculate $\widehat{v}_{1,N}$, $\widehat{v}_{2,N}$, $\|\widehat{\boldsymbol{v}}_N - \boldsymbol{c}\|$, $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$, and $|\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N} - \rho(\pi_1 f_1, \pi_2 f_2)|$. Each dotted line indicates the expected value: $c_1 \simeq -0.779$ for $\widehat{v}_{1,N}$, $c_2 \simeq 0.779$ for $\widehat{v}_{2,N}$, 0 for $\|\widehat{\boldsymbol{v}}_N - \boldsymbol{c}\|$, $\rho(\pi_1 f_1, \pi_2 f_2) \simeq 0.362$ for $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$, and 0 for $|\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N} - \rho(\pi_1 f_1, \pi_2 f_2)|$. In this figure, $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$ and $\rho(\pi_1 f_1, \pi_2 f_2)$ are abbreviated as $\widehat{\rho}_N$ and ρ, respectively.

as is consistent with the nature of OVL, while the OVL estimators in [6] depend not only on the rank statistics ([6, pp. 1588–1589]); (ii) our OVL estimator converges completely to the true value (Theorem 4.6).

Further studies on these problems are needed for the practical use of our method.

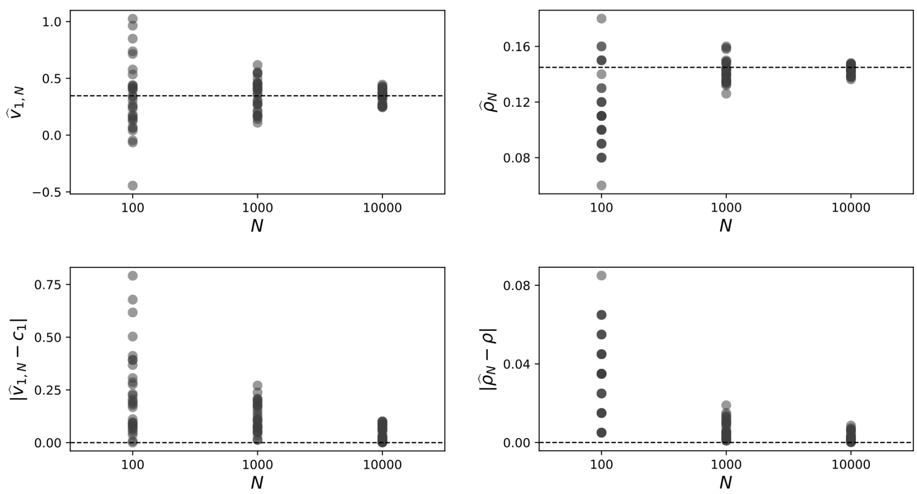

Fig. 6. In the first case, 30 independent trials were performed to simulate three independent sets of random samples, of the forms $\{(X_i, Y_i) \mid 1 \leq i \leq 100\}$, $\{(X_i', Y_i') \mid 1 \leq i \leq 1000\}$, and $\{(X_i'', Y_i'') \mid 1 \leq i \leq 10000\}$. Each set was used to calculate $\widehat{v}_{1,N}$, $|\widehat{v}_{1,N} - c_1|$, $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$, and $|\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N} - \rho(\pi_1 f_1, \pi_2 f_2)|$. Note that the superscript (N) in Remark 4.7 is omitted here. The dotted lines indicate the expected values: $c_1 \simeq 0.347$ for $\widehat{v}_{1,N}$, 0 for $|\widehat{v}_{1,N} - c_1|$, $\rho(\pi_1 f_1, \pi_2 f_2) \simeq 0.145$ for $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$, and 0 for $|\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N} - \rho(\pi_1 f_1, \pi_2 f_2)|$. In this figure, $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$ and $\rho(\pi_1 f_1, \pi_2 f_2)$ are abbreviated as $\widehat{\rho}_N$ and ρ, respectively.

Appendix A. Additional Proofs

Theorems 4.5 and 4.6 will be proved in this section. We shall take over the notations in Section 4 and, in addition, write $h(\boldsymbol{v})$ and $\widehat{h}_N(\boldsymbol{v})$ in place of $\Delta I(S_{\boldsymbol{v}})$ and $\Delta \widehat{I}_N(S_{\boldsymbol{v}})$, respectively.

Definition A.1. For $j \in \{1, 2\}$ and $x \in \mathbb{R}$, define

$$\widehat{F}_{j,N}(x) = \begin{cases} N_{XY}((-\infty, x], j)/N_Y(j) & \text{if} \quad N_Y(j) > 0, \\ 0 & \text{if} \quad N_Y(j) = 0. \end{cases}$$

We also define $\widehat{F}_{j,N}(-\infty) = 0$ and $\widehat{F}_{j,N}(\infty) = 1$.

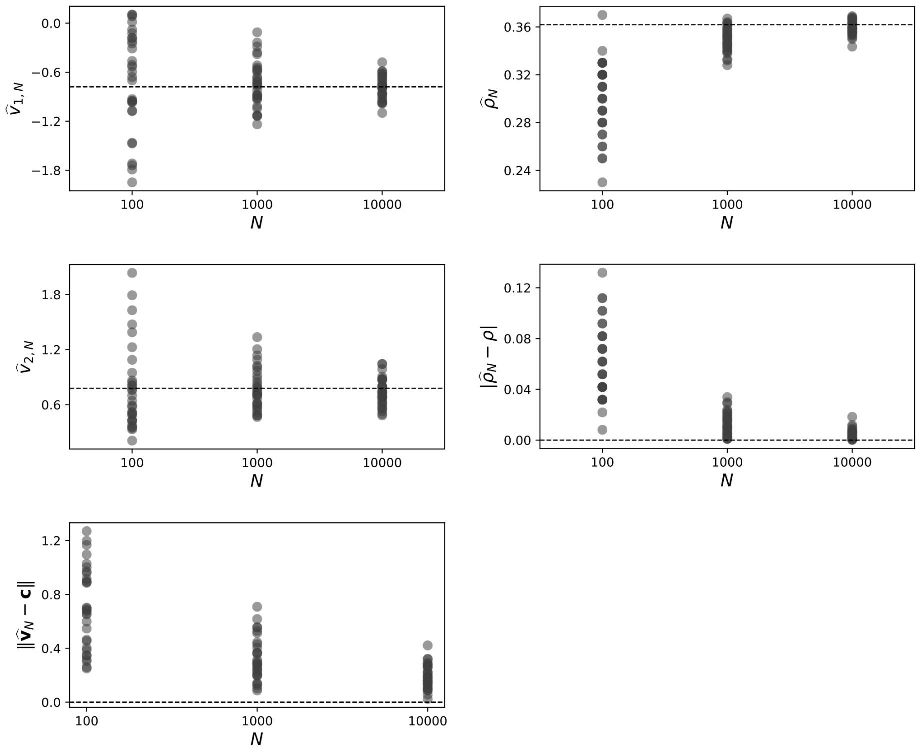

Fig. 7. In the second case, 30 independent trials were performed to simulate three independent sets of random samples, of the forms $\{(X_i, Y_i) \mid 1 \leq i \leq 100\}$, $\{(X'_i, Y'_i) \mid 1 \leq i \leq 1000\}$, and $\{(X''_i, Y''_i) \mid 1 \leq i \leq 10000\}$. Each set was used to calculate $\widehat{v}_{1,N}$, $\widehat{v}_{2,N}$, $\|\widehat{\boldsymbol{v}}_N - \boldsymbol{c}\|$, $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$, and $|\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N} - \rho(\pi_1 f_1, \pi_2 f_2)|$. Note that the superscript (N) in Remark 4.7 is omitted here. The dotted lines indicate the expected values: $c_1 \simeq -0.779$ for $\widehat{v}_{1,N}$, $c_2 \simeq 0.779$ for $\widehat{v}_{2,N}$, 0 for $\|\widehat{\boldsymbol{v}}_N - \boldsymbol{c}\|$, $\rho(\pi_1 f_1, \pi_2 f_2) \simeq 0.362$ for $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$, and 0 for $|\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N} - \rho(\pi_1 f_1, \pi_2 f_2)|$. In this figure, $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$ and $\rho(\pi_1 f_1, \pi_2 f_2)$ are abbreviated as $\widehat{\rho}_N$ and ρ, respectively.

Proposition A.2. *For $\boldsymbol{v} = (v_1, \ldots, v_m) \in \mathbb{R}^m_{\leq}$ with m a positive integer,*

$$\widehat{h}_N(\boldsymbol{v}) = \sum_{k=1}^{m+1} \max_j \left\{ \widehat{\pi}_{j,N} \left[\widehat{F}_{j,N}(v_k) - \widehat{F}_{j,N}(v_{k-1}) \right] \right\} - \max_j \left\{ \widehat{\pi}_{j,N} \right\},$$

where $v_0 = -\infty$, $v_{m+1} = \infty$, $\widehat{F}_{j,N}(v_0) = 0$, and $\widehat{F}_{j,N}(v_{m+1}) = 1$.

PROOF. From (5), we have

$$\begin{aligned}\widehat{h}_N(\boldsymbol{v}) = &\sum_k \widehat{P}_N(X \in S_{\boldsymbol{v},k}) \max_j \left\{\widehat{P}_N(Y = j \mid X \in S_{\boldsymbol{v},k})\right\} \\ &- \max_j \left\{\widehat{P}_N(Y = j \mid X \in \mathbb{R})\right\},\end{aligned} \tag{14}$$

where the sum is over all k with $N_X(S_{\boldsymbol{v},k}) > 0$. Since $\widehat{P}_N(X \in S_{\boldsymbol{v},k}) = N_X(S_{\boldsymbol{v},k})/N$, $\widehat{P}_N(Y = j \mid X \in S_{\boldsymbol{v},k}) = N_{XY}(S_{\boldsymbol{v},k}, j)/N_X(S_{\boldsymbol{v},k})$, and $N_{XY}(S_{\boldsymbol{v},k}, j) = N\widehat{\pi}_{j,N}[\widehat{F}_{j,N}(v_k) - \widehat{F}_{j,N}(v_{k-1})]$, we obtain

$$\begin{aligned}&\widehat{P}_N(X \in S_{\boldsymbol{v},k}) \max_j \left\{\widehat{P}_N(Y = j \mid X \in S_{\boldsymbol{v},k})\right\} \\ &= \max_j \left\{\widehat{\pi}_{j,N} \left[\widehat{F}_{j,N}(v_k) - \widehat{F}_{j,N}(v_{k-1})\right]\right\}.\end{aligned}$$

As for the last term of (14), we have $\widehat{P}_N(Y = j \mid X \in \mathbb{R}) = \widehat{\pi}_{j,N}$ by definition. □

COROLLARY A.3. *For $\boldsymbol{v} \in \mathbb{R}^m_{\leq}$ with m a positive integer, $\widehat{h}_N(\boldsymbol{v}) \geq 0$ and $h(\boldsymbol{v}) \geq 0$.*

PROOF. Let $\widehat{\pi}_{p,N} = \max\{\widehat{\pi}_{1,N}, \widehat{\pi}_{2,N}\}$. By Proposition A.2, we have

$$\begin{aligned}\widehat{h}_N(\boldsymbol{v}) &= \sum_{k=1}^{m+1} \max_j \left\{\widehat{\pi}_{j,N} \left[\widehat{F}_{j,N}(v_k) - \widehat{F}_{j,N}(v_{k-1})\right]\right\} - \max_j \{\widehat{\pi}_{j,N}\} \\ &\geq \sum_{k=1}^{m+1} \widehat{\pi}_{p,N} \left[\widehat{F}_{p,N}(v_k) - \widehat{F}_{p,N}(v_{k-1})\right] - \widehat{\pi}_{p,N} = 0.\end{aligned}$$

We can similarly prove that $h(\boldsymbol{v}) \geq 0$ from Proposition 3.1. □

For simplicity, we may write $\varphi_j(v, v')$ and $\widehat{\varphi}_{j,N}(v, v')$ in place of $\pi_j[F_j(v) - F_j(v')]$ and $\widehat{\pi}_{j,N}[\widehat{F}_{j,N}(v) - \widehat{F}_{j,N}(v')]$, respectively, so that

$$h(\boldsymbol{v}) = \sum_{k=1}^{m+1} \max_j \{\varphi_j(v_k, v_{k-1})\} - \max_j \{\pi_j\}, \tag{15}$$

$$\widehat{h}_N(\boldsymbol{v}) = \sum_{k=1}^{m+1} \max_j \{\widehat{\varphi}_{j,N}(v_k, v_{k-1})\} - \max_j \{\widehat{\pi}_{j,N}\} \tag{16}$$

by Propositions 3.1 and A.2.

DEFINITION A.4. For $m = 1, \ldots, n$, define

$$\begin{aligned}
\mathcal{V}_m &= \mathop{\arg\max}_{\boldsymbol{v} \in \mathbb{R}^m_{\leq}} \{h(\boldsymbol{v})\}, \\
\widehat{\mathcal{V}}_{m,N} &= \mathop{\arg\max}_{\boldsymbol{v} \in \widehat{\mathbb{R}}^m_N} \left\{\widehat{h}_N(\boldsymbol{v})\right\}, \\
\mathcal{C}_m &= \{(c_{i_1}, \ldots, c_{i_m}) \mid 1 \leq i_1 < \cdots < i_m \leq n\}.
\end{aligned}$$

REMARK A.5. We will see that $\mathcal{V}_m \neq \emptyset$ $(m \leq n)$ by Corollary A.7 and Proposition 3.4. Since $\widehat{\mathbb{R}}^m_N$ is a nonempty finite set (see Definition 4.2), $\widehat{\mathcal{V}}_{m,N} \neq \emptyset$.

PROPOSITION A.6. *Let m be a positive integer with $m < n$. Then for any $\boldsymbol{v} = (v_1, \ldots, v_m) \in \mathbb{R}^m_{\leq}$, there exists $\boldsymbol{w} = (c_{i_1}, \ldots, c_{i_m})$ with $1 \leq i_1 \leq \cdots \leq i_m \leq n$ such that $h(\boldsymbol{w}) \geq h(\boldsymbol{v})$.*

PROOF. Let $\boldsymbol{v} = (v_1, \ldots, v_m) \in \mathbb{R}^m_{\leq}$ be given. Set $v_0 = -\infty$, $v_{m+1} = \infty$, and $r(\boldsymbol{v}) = \#\{k \in \{1, \ldots, m\} \mid v_k \notin C(\pi_1 f_1, \pi_2 f_2)\}$. The statement obviously holds when $r(\boldsymbol{v}) = 0$.

Let $r(\boldsymbol{v}) > 0$. Then we can choose $v_p \notin C(\pi_1 f_1, \pi_2 f_2)$ $(1 \leq p \leq m)$ and $c_q \in C(\pi_1 f_1, \pi_2 f_2)$ $(1 \leq q \leq n)$ satisfying $c_{q-1} < v_p < c_q \leq v_{p+1}$ or $v_{p-1} \leq c_q < v_p < c_{q+1}$. We will only show the case $c_{q-1} < v_p < c_q \leq v_{p+1}$, as the other is similar. Without loss of generality, we may assume that $\pi_1 f_1 \geq \pi_2 f_2$ on (v_p, c_q), so that $\varphi_1(c_q, v_p) > \varphi_2(c_q, v_p)$ and $\varphi_1(v_p, c_{q-1}) > \varphi_2(v_p, c_{q-1})$, since $C'(\pi_1 f_1, \pi_2 f_2)$ is finite. In the following, we consider the cases (I) $\varphi_1(v_p, v_{p-1}) \geq \varphi_2(v_p, v_{p-1})$ and (II) $\varphi_1(v_p, v_{p-1}) < \varphi_2(v_p, v_{p-1})$.

(I) Suppose $\varphi_1(v_p, v_{p-1}) \geq \varphi_2(v_p, v_{p-1})$. Then

$$\begin{aligned}
&\varphi_1(c_q, v_{p-1}) > \varphi_2(c_q, v_{p-1}), \\
&\varphi_j(v_{p+1}, c_q) = \varphi_j(v_{p+1}, v_p) - \varphi_j(c_q, v_p) \qquad (j = 1, 2), \\
&\varphi_j(c_q, v_{p-1}) = \varphi_j(v_p, v_{p-1}) + \varphi_j(c_q, v_p) \qquad (j = 1, 2),
\end{aligned}$$

hence

$$\begin{aligned}
&\max_j \{\varphi_j(c_q, v_{p-1})\} + \max_j \{\varphi_j(v_{p+1}, c_q)\} \\
&= \varphi_1(c_q, v_{p-1}) + \max_j \{\varphi_j(v_{p+1}, v_p) - \varphi_j(c_q, v_p)\}
\end{aligned}$$

$$
\begin{aligned}
&= \varphi_1(v_p, v_{p-1}) + \varphi_1(c_q, v_p) + \max_j \{\varphi_j(v_{p+1}, v_p) - \varphi_j(c_q, v_p)\} \\
&\geq \varphi_1(v_p, v_{p-1}) + \varphi_1(c_q, v_p) + \max_j \{\varphi_j(v_{p+1}, v_p)\} - \varphi_1(c_q, v_p) \\
&= \varphi_1(v_p, v_{p-1}) + \max_j \{\varphi_j(v_{p+1}, v_p)\} \\
&= \max_j \{\varphi_j(v_p, v_{p-1})\} + \max_j \{\varphi_j(v_{p+1}, v_p)\},
\end{aligned}
$$

and setting $\boldsymbol{v}' = (v_1, \dots, v_{p-1}, c_q, v_{p+1}, \dots, v_m) \in \mathbb{R}^m_{\leq}$ gives $r(\boldsymbol{v}') < r(\boldsymbol{v})$ and $h(\boldsymbol{v}') \geq h(\boldsymbol{v})$.

(II) Suppose $\varphi_1(v_p, v_{p-1}) < \varphi_2(v_p, v_{p-1})$. Since $\pi_1 f_1 \geq \pi_2 f_2$ on (v_p, c_q), we can see that $v_{p-1} < c_{q-1} < v_p$ and $\varphi_1(c_{q-1}, v_{p-1}) < \varphi_2(c_{q-1}, v_{p-1})$. First consider the case (II-a) $\varphi_1(v_{p+1}, v_p) \geq \varphi_2(v_{p+1}, v_p)$. Then $\varphi_1(v_{p+1}, c_{q-1}) > \varphi_2(v_{p+1}, c_{q-1})$, hence

$$
\begin{aligned}
&\max_j \{\varphi_j(c_{q-1}, v_{p-1})\} + \max_j \{\varphi_j(v_{p+1}, c_{q-1})\} \\
&= \varphi_2(c_{q-1}, v_{p-1}) + \varphi_1(v_{p+1}, c_{q-1}) \\
&= \varphi_2(c_{q-1}, v_{p-1}) + \varphi_1(v_{p+1}, v_p) + \varphi_1(v_p, c_{q-1}) \\
&> \varphi_2(c_{q-1}, v_{p-1}) + \varphi_1(v_{p+1}, v_p) + \varphi_2(v_p, c_{q-1}) \\
&= \varphi_2(v_p, v_{p-1}) + \varphi_1(v_{p+1}, v_p) \\
&= \max_j \{\varphi_j(v_p, v_{p-1})\} + \max_j \{\varphi_j(v_{p+1}, v_p)\},
\end{aligned}
$$

and setting $\boldsymbol{v}' = (v_1, \dots, v_{p-1}, c_{q-1}, v_{p+1}, \dots, v_m) \in \mathbb{R}^m_{\leq}$ gives $r(\boldsymbol{v}') < r(\boldsymbol{v})$ and $h(\boldsymbol{v}') > h(\boldsymbol{v})$. Next consider the case (II-b) $\varphi_1(v_{p+1}, v_p) < \varphi_2(v_{p+1}, v_p)$. If there exists $x \in (c_{q-1}, v_p)$ such that $\varphi_1(v_{p+1}, x) \geq \varphi_2(v_{p+1}, x)$, then $\varphi_1(x, v_{p-1}) < \varphi_2(x, v_{p-1})$, hence the case (II-a) applies to $\boldsymbol{v}'' = (v_1, \dots, v_{p-1}, x, v_{p+1}, \dots, v_m) \in \mathbb{R}^m_{\leq}$, where $r(\boldsymbol{v}'') = r(\boldsymbol{v})$ and

$$
\begin{aligned}
&h(\boldsymbol{v}'') - h(\boldsymbol{v}) \\
&= \max_j \{\varphi_j(x, v_{p-1})\} + \max_j \{\varphi_j(v_{p+1}, x)\} \\
&\quad - \max_j \{\varphi_j(v_p, v_{p-1})\} - \max_j \{\varphi_j(v_{p+1}, v_p)\} \\
&= \varphi_2(x, v_{p-1}) + \varphi_1(v_{p+1}, x) - \varphi_2(v_p, v_{p-1}) - \varphi_2(v_{p+1}, v_p) \\
&\geq \varphi_2(x, v_{p-1}) + \varphi_2(v_{p+1}, x) - \varphi_2(v_p, v_{p-1}) - \varphi_2(v_{p+1}, v_p) \\
&= \varphi_2(v_{p+1}, v_{p-1}) - \varphi_2(v_{p+1}, v_{p-1}) = 0.
\end{aligned}
$$

If $\varphi_1(v_{p+1}, x) < \varphi_2(v_{p+1}, x)$ for any $x \in (c_{q-1}, v_p)$, then $\varphi_1(v_{p+1}, c_{q-1}) \leq \varphi_2(v_{p+1}, c_{q-1})$, and setting $\boldsymbol{v}' = (v_1, \ldots, v_{p-1}, c_{q-1}, v_{p+1}, \ldots, v_m) \in \mathbb{R}^m_{\leq}$ gives $r(\boldsymbol{v}') < r(\boldsymbol{v})$ and

$$
\begin{aligned}
&h(\boldsymbol{v}') - h(\boldsymbol{v}) \\
&= \max_j \{\varphi_j(c_{q-1}, v_{p-1})\} + \max_j \{\varphi_j(v_{p+1}, c_{q-1})\} \\
&\quad - \max_j \{\varphi_j(v_p, v_{p-1})\} - \max_j \{\varphi_j(v_{p+1}, v_p)\} \\
&= \varphi_2(c_{q-1}, v_{p-1}) + \varphi_2(v_{p+1}, c_{q-1}) - \varphi_2(v_p, v_{p-1}) - \varphi_2(v_{p+1}, v_p) \\
&= \varphi_2(v_{p+1}, v_{p-1}) - \varphi_2(v_{p+1}, v_{p-1}) = 0.
\end{aligned}
$$

Taken together, for any $\boldsymbol{v} \in \mathbb{R}^m_{\leq}$ with $r(\boldsymbol{v}) > 0$, there exists $\boldsymbol{v}' \in \mathbb{R}^m_{\leq}$ such that $r(\boldsymbol{v}') < r(\boldsymbol{v})$ and $h(\boldsymbol{v}') \geq h(\boldsymbol{v})$. The statement follows by induction. □

COROLLARY A.7. *If m is a positive integer with $m < n$, then there exists $\boldsymbol{c}' \in \mathcal{C}_m$ such that $h(\boldsymbol{c}') = \sup \{h(\boldsymbol{v}) \mid \boldsymbol{v} \in \mathbb{R}^m_{\leq}\}$. Furthermore, $h(\boldsymbol{c}') < h(\boldsymbol{c})$.*

PROOF. Since there are only finitely many choices for $\boldsymbol{w} \in \mathbb{R}^m_{\leq}$ in Proposition A.6, we can choose $\boldsymbol{w}' = (c_{i_1}, \ldots, c_{i_m}) \in \arg\max_{\boldsymbol{w}} h(\boldsymbol{w})$, where $\boldsymbol{w}$ ranges over the choices. Then $h(\boldsymbol{w}') \geq h(\boldsymbol{v})$ for all $\boldsymbol{v} \in \mathbb{R}^m_{\leq}$. Let $A = \{c_{i_1}, \ldots, c_{i_m}\}$ and assume that $\boldsymbol{w}' \notin \mathcal{C}_m$. Then $\#A < m$, and there exists $A' = \{c_{j_1}, \ldots, c_{j_m}\}$ such that $A \subset A'$ and $1 \leq j_1 < \cdots < j_m \leq n$. Put $\boldsymbol{c}' = (c_{j_1}, \ldots, c_{j_m})$. Then $\boldsymbol{c}' \in \mathcal{C}_m$, and we can see that $h(\boldsymbol{c}') \geq h(\boldsymbol{w}')$ by definition. Furthermore, $h(\boldsymbol{c}') < h(\boldsymbol{c})$ by Proposition 3.5. □

REMARK A.8. Note that $\boldsymbol{v} \in \mathcal{V}_m$ does not necessarily imply $\boldsymbol{v} \in \mathcal{C}_m$. Here we give an example for the case where $(n, m) = (2, 1)$ and $\mathcal{V}_1 \not\subset \mathcal{C}_1$. Assume that $\pi_1 = 0.9$, $\pi_2 = 0.1$, $f_1 = \nu_{0,1}$, and $f_2 = \tau_{-0.1,0.1}$ (see (7) and (8) for the definitions of ν and τ). Then $\pi_1 f_1(0) < \pi_2 f_2(0)$, $n = 2$, and $\mathcal{C}_2 = \{c_1, c_2\}$ where $-0.1 < c_1 < 0 < c_2 < 0.1$. Since $\varphi_1(\infty, 0.1) = \varphi_1(-0.1, -\infty) = \pi_1 \Phi(-0.1) \simeq 0.4142 > \pi_2$ (see (13) for the definition of Φ), $\varphi_1(v, -\infty) > \varphi_2(v, -\infty)$ and $\varphi_1(\infty, v) > \varphi_2(\infty, v)$ hold for all $v \in \mathbb{R}$. Hence $h(v) = \pi_1$ for all $v \in \mathbb{R}$, and therefore $\mathcal{V}_1 = \mathbb{R} \not\subset \{c_1, c_2\} = \mathcal{C}_1$.

For a real random variable ξ on $(\Omega, \mathcal{F}, \mathbb{P})$, we denote its expectation and variance by

$$\mathbb{E}[\xi] = \int_\Omega \xi \, d\mathbb{P}, \qquad \operatorname{Var}[\xi] = \int_\Omega (\xi - \mathbb{E}[\xi])^2 \, d\mathbb{P},$$

respectively. We also denote by $\mathbb{1}_A$ the indicator function of a set A, i.e.,

$$\mathbb{1}_A(t) = \begin{cases} 1 & \text{if} \quad t \in A, \\ 0 & \text{if} \quad t \notin A. \end{cases}$$

THEOREM A.9 (Kolmogorov's strong law of large numbers. See [2] for the proof). *Let $\{\xi_i\}$ be a sequence of i.i.d. real random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}[|\xi_1|] < \infty$ and $\operatorname{Var}[\xi_1] < \infty$. Let $\mu = \mathbb{E}[\xi_1]$ and $s_k = \xi_1 + \cdots + \xi_k$ (k=1,2, ...). Then s_k/k converges completely to μ.*

THEOREM A.10 (The Glivenko-Cantelli theorem. See [7, Theorem A, Section 2.1.4] for the proof). *For each $j \in \{1,2\}$, $\sup_{x\in\mathbb{R}} |\widehat{F}_{j,N}(x) - F_j(x)|$ converges completely to 0 as $N \to \infty$.*

PROPOSITION A.11. *For each $j \in \{1,2\}$, $\widehat{\pi}_{j,N}$ converges completely to π_j as $N \to \infty$.*

PROOF. We can see $\mathbb{1}_{\{j\}}(Y_1), \ldots, \mathbb{1}_{\{j\}}(Y_N)$ as i.i.d. random variables with $\mathbb{E}[\mathbb{1}_{\{j\}}(Y_1)] = \pi_j < \infty$ and $\operatorname{Var}[\mathbb{1}_{\{j\}}(Y_1)] = \pi_j(1-\pi_j) < \infty$. Since $N_Y(j) = \mathbb{1}_{\{j\}}(Y_1) + \cdots + \mathbb{1}_{\{j\}}(Y_N)$, $\widehat{\pi}_{j,N} = N_Y(j)/N$ converges completely to π_j by Theorem A.9. □

LEMMA A.12. *If $x, y, z, w \in \mathbb{R}$, then*

(a) $|\max\{x,y\} - \max\{z,w\}| \le |x-z| + |y-w|$,

(b) $|\min\{x,y\} - \min\{z,w\}| \le |x-z| + |y-w|$.

PROOF. For (a), suppose $\max\{x,y\} \ge \max\{z,w\}$ and $x \ge y$ without loss of generality. If $z \ge w$, then $|\max\{x,y\} - \max\{z,w\}| = |x-z| \le |x-z| + |y-w|$. If $z < w$, then $|\max\{x,y\} - \max\{z,w\}| = |x-w| < |x-z| \le |x-z| + |y-w|$.

For (b), suppose $\min\{x,y\} \geq \min\{z,w\}$ and $x \geq y$ without loss of generality. If $z \geq w$, then $|\min\{x,y\} - \min\{z,w\}| = |y-w| \leq |x-z| + |y-w|$. If $z < w$, then $|\min\{x,y\} - \min\{z,w\}| = |y-z| \leq |x-z| \leq |x-z| + |y-w|$. $\square$

THEOREM A.13. *For any positive integer m, $\sup_{\boldsymbol{v}\in\mathbb{R}^m_{\leq}} |\widehat{h}_N(\boldsymbol{v}) - h(\boldsymbol{v})|$ converges completely to 0 as $N \to \infty$.*

PROOF. For all $\boldsymbol{v} \in \mathbb{R}^m_{\leq}$, we have

$$\begin{aligned}
&\left|\widehat{h}_N(\boldsymbol{v}) - h(\boldsymbol{v})\right| \\
&\leq \sum_{k=1}^{m+1} \left| \max_j \{\widehat{\varphi}_{j,N}(v_k, v_{k-1})\} - \max_j \{\varphi_j(v_k, v_{k-1})\} \right| \\
&\quad + \left| \max_j \{\widehat{\pi}_{j,N}\} - \max_j \{\pi_j\} \right| \\
&\leq \sum_{k=1}^{m+1} \sum_{j=1}^{2} |\widehat{\varphi}_{j,N}(v_k, v_{k-1}) - \varphi_j(v_k, v_{k-1})| + \sum_{j=1}^{2} |\widehat{\pi}_{j,N} - \pi_j|
\end{aligned}$$

by (15), (16), and Lemma A.12. Since

$$\begin{aligned}
&|\widehat{\varphi}_{j,N}(v_k, v_{k-1}) - \varphi_j(v_k, v_{k-1})| \\
&= \Big| (\widehat{\pi}_{j,N} - \pi_j) \left[\widehat{F}_{j,N}(v_k) - \widehat{F}_{j,N}(v_{k-1})\right] \\
&\qquad + \pi_j \left[\widehat{F}_{j,N}(v_k) - F_j(v_k)\right] - \pi_j \left[\widehat{F}_{j,N}(v_{k-1}) - F_j(v_{k-1})\right] \Big| \\
&\leq |\widehat{\pi}_{j,N} - \pi_j| \left|\widehat{F}_{j,N}(v_k) - \widehat{F}_{j,N}(v_{k-1})\right| \\
&\qquad + \pi_j \left|\widehat{F}_{j,N}(v_k) - F_j(v_k)\right| + \pi_j \left|\widehat{F}_{j,N}(v_{k-1}) - F_j(v_{k-1})\right| \\
&\leq |\widehat{\pi}_{j,N} - \pi_j| + 2\pi_j \sup_{x\in\mathbb{R}} \left|\widehat{F}_{j,N}(x) - F_j(x)\right|,
\end{aligned}$$

we obtain

$$\begin{aligned}
&\sup_{\boldsymbol{v}\in\mathbb{R}^m_{\leq}} \left|\widehat{h}_N(\boldsymbol{v}) - h(\boldsymbol{v})\right| \\
&\leq (m+2) \sum_{j=1}^{2} |\widehat{\pi}_{j,N} - \pi_j| + 2(m+1) \sum_{j=1}^{2} \pi_j \sup_{x\in\mathbb{R}} \left|\widehat{F}_{j,N}(x) - F_j(x)\right|.
\end{aligned}$$

Hence

$$\left\{\omega \in \Omega \;\middle|\; \sup_{\boldsymbol{v} \in \mathbb{R}^m_{\leq}} \left|\widehat{h}_N(\boldsymbol{v}) - h(\boldsymbol{v})\right| > \epsilon\right\}$$

is contained in

$$\begin{aligned}
&\bigcup_{j=1}^{2} \left\{\omega \in \Omega \;\middle|\; |\widehat{\pi}_{j,N} - \pi_j| > \frac{\epsilon}{4(m+2)}\right\} \\
&\cup \bigcup_{j=1}^{2} \left\{\omega \in \Omega \;\middle|\; \sup_{x \in \mathbb{R}} \left|\widehat{F}_{j,N}(x) - F_j(x)\right| > \frac{\epsilon}{8(m+1)}\right\},
\end{aligned}$$

and therefore

$$\begin{aligned}
&\sum_{N=1}^{\infty} \mathbb{P}\left(\left\{\omega \in \Omega \;\middle|\; \sup_{\boldsymbol{v} \in \mathbb{R}^m_{\leq}} \left|\widehat{h}_N(\boldsymbol{v}) - h(\boldsymbol{v})\right| > \epsilon\right\}\right) \\
&\leq \sum_{j=1}^{2} \sum_{N=1}^{\infty} \mathbb{P}\left(\left\{\omega \in \Omega \;\middle|\; |\widehat{\pi}_{j,N} - \pi_j| > \frac{\epsilon}{4(m+2)}\right\}\right) \\
&\qquad + \sum_{j=1}^{2} \sum_{N=1}^{\infty} \mathbb{P}\left(\left\{\omega \in \Omega \;\middle|\; \sup_{x \in \mathbb{R}} \left|\widehat{F}_{j,N}(x) - F_j(x)\right| > \frac{\epsilon}{8(m+1)}\right\}\right) \\
&< \infty
\end{aligned}$$

by Theorem A.10 and Proposition A.11. □

Definition A.14. Let (A, d) be a metric space. We define a discrepancy of $A_1 \subset A$ from $A_2 \subset A$ by

$$D(A_1, A_2) = \sup_{a_1 \in A_1} \left\{\inf_{a_2 \in A_2} d(a_1, a_2)\right\}.$$

If d is a Euclidean metric, we may write D_{E} in place of D.

Lemma A.15. *Let (A, d) be a metric space. Let g and g_i (i=1,2, ...) be real functions on A such that* $\max\{g(t) \mid t \in A\}$ *and* $\max\{g_i(t) \mid t \in A\}$ *exist. Put* $T = \arg\max_{t \in A}\{g(t)\}$ *and* $T_i = \arg\max_{t \in A}\{g_i(t)\}$. *Suppose g*

is continuous on A, $\sup_{t\in A}|g_i(t)-g(t)|\to 0$ as $i\to\infty$, and there exists a compact set $K\subset A$ such that

$$\sup\{g(t)\mid t\in A\setminus K\}<\max\{g(t)\mid t\in A\}.$$

Then $D(T_i,T)\to 0$ as $i\to\infty$.

Proof. Put $w_1=\max\{g(t)\mid t\in A\}$, $w_0=\sup\{g(t)\mid t\in A\setminus K\}$, and $w=(w_1-w_0)/3$. (Note that $w_0<w_0+w<w_0+2w=w_1-w<w_1$.) For any $\epsilon>0$, there exists $\delta>0$ such that $\delta<\epsilon$ and $g(t)>w_1-w$ for all $t\in T_\delta=\cup_{t\in T}\{x\in K\mid d(x,t)<\delta\}$, since g is uniformly continuous on K (see [5, Theorem 4.19]). (Note that $T\subset K$.) Put $w_0'=\max\{g(t)\mid t\in K\setminus T_\delta\}$ (this exists because $K\setminus T_\delta$ is compact) and $w'>0$ such that $w'<w$ and $w_0'<w_0'+2w'<w_1$. (Note that $\{t\in A\mid g(t)>w_1-2w'\}\subset T_\delta$ since $w_1-2w'>w_1-2w>w_0$ and $w_1-2w'>w_0'$.) Since $\sup_{t\in A}|g_i(t)-g(t)|\to 0$ as $i\to\infty$, there is an integer M such that $i\geq M$ implies $\sup_{t\in A}|g_i(t)-g(t)|<w'$. Hence, for any $i\geq M$ and for all $t_1\in T_i$, we have $g(t_1)>w_1-2w'$ (because $g(t_1)+w'>g_i(t_1)\geq g_i(t_2)>w_1-w'$ where $t_2\in T$), and thus $t_1\in T_\delta$. Therefore, $\sup_{t_1\in T_i}\{\inf_{t_2\in T}d(t_1,t_2)\}\leq\delta<\epsilon$ for any $i\geq M$. Since ϵ was arbitrary, the claim follows. □

Lemma A.16. *There exists a compact set $K\subset\mathbb{R}^n_{\leq}$ such that*

$$\sup\left\{h(\boldsymbol{v})\mid\boldsymbol{v}\in\mathbb{R}^n_{\leq}\setminus K\right\}<\max\left\{h(\boldsymbol{v})\mid\boldsymbol{v}\in\mathbb{R}^n_{\leq}\right\}.$$

Proof. By Propositions 3.4 and 3.5 and Corollary A.7, there exist

$$M_m=\max\left\{h(\boldsymbol{v})\mid\boldsymbol{v}\in\mathbb{R}^m_{\leq}\right\}\qquad(m=1\dots,n)$$

and $M=\max\{M_1,\dots,M_{n-1}\}<M_n$. Take $\epsilon>0$ such that $\epsilon<(M_n-M)/3$. We can take $\alpha,\beta\in\mathbb{R}$ such that $F_j(\alpha)<\epsilon$ and $1-F_j(\beta)<\epsilon$ $(j=1,2)$, since F_j are non-decreasing functions with $\lim_{x\to-\infty}F_j(x)=0$ and $\lim_{x\to\infty}F_j(x)=1$. Let $K=[\alpha,\beta]^n\cap\mathbb{R}^n_{\leq}$ and $\boldsymbol{v}=(v_1,\dots,v_n)\in\mathbb{R}^n_{\leq}\setminus K$. Then $v_1<\alpha$ or $v_n>\beta$ holds.

Suppose $v_1 < \alpha$. Put $\boldsymbol{v}' = (v_2, \dots, v_n)$ and recall that $\varphi_j(v, v') = \pi_j[F_j(v) - F_j(v')]$. Using Lemma A.12, we obtain

$$\begin{aligned}
&\left|h(\boldsymbol{v}) - h(\boldsymbol{v}')\right| \\
&= \left|\max_j \{\varphi_j(v_1, -\infty)\} + \max_j \{\varphi_j(v_2, v_1)\} - \max_j \{\varphi_j(v_2, -\infty)\}\right| \\
&\leq \left|\max_j \{\varphi_j(v_1, -\infty)\}\right| + \left|\max_j \{\varphi_j(v_2, v_1)\} - \max_j \{\varphi_j(v_2, -\infty)\}\right| \\
&< \epsilon + |\varphi_1(v_2, v_1) - \varphi_1(v_2, -\infty)| + |\varphi_2(v_2, v_1) - \varphi_2(v_2, -\infty)| \\
&= \epsilon + |\varphi_1(v_1, -\infty)| + |\varphi_2(v_1, -\infty)| \\
&< 3\epsilon.
\end{aligned}$$

Hence $|h(\boldsymbol{v})| \leq |h(\boldsymbol{v}) - h(\boldsymbol{v}')| + |h(\boldsymbol{v}')| < 3\epsilon + M$. We can similarly prove that $|h(\boldsymbol{v})| < 3\epsilon + M$ for the case $v_n > \beta$. Therefore, $\sup\{h(\boldsymbol{v}) \mid \boldsymbol{v} \in \mathbb{R}^n_{\leq} \setminus K\} \leq 3\epsilon + M < (M_n - M) + M = M_n$. This completes the proof. $\square$

THEOREM A.17. *The discrepancy $D_{\mathrm{E}}(\widehat{\mathcal{V}}_{n,N}, \mathcal{V}_n)$ converges completely to 0 as $N \to \infty$.*

PROOF. In Lemma A.15, let (A, d) be the subspace $\mathbb{R}^n_{\leq}$ of the Euclidean metric space $\mathbb{R}^n$, $g = h$ (which is continuous on $\mathbb{R}^n_{\leq}$), and $g_i = \widehat{h}_i$. It follows from Remark A.5 and Lemma A.16 that for any $\epsilon > 0$, we can take $w' > 0$ as in the proof of Lemma A.15, and observe that $D_{\mathrm{E}}(\widehat{\mathcal{V}}_{n,N}, \mathcal{V}_n) < \epsilon$ if $\sup_{\boldsymbol{v} \in \mathbb{R}^n_{\leq}} |\widehat{h}_N(\boldsymbol{v}) - h(\boldsymbol{v})| < w'$. (Note that $\max\{\widehat{h}_N(\boldsymbol{v}) \mid \boldsymbol{v} \in \widehat{\mathbb{R}}^n_N\} = \max\{\widehat{h}_N(\boldsymbol{v}) \mid \boldsymbol{v} \in \mathbb{R}^n_{\leq}\}$, hence $\widehat{\mathcal{V}}_{n,N} \subset \arg\max_{\boldsymbol{v} \in \mathbb{R}^n_{\leq}} \{\widehat{h}_N(\boldsymbol{v})\}$.) This means that

$$\left\{\omega \in \Omega \;\middle|\; \sup_{\boldsymbol{v} \in \mathbb{R}^n_{\leq}} \left|\widehat{h}_N(\boldsymbol{v}) - h(\boldsymbol{v})\right| < w'\right\} \subset \left\{\omega \in \Omega \;\middle|\; D_{\mathrm{E}}\left(\widehat{\mathcal{V}}_{n,N}, \mathcal{V}_n\right) < \epsilon\right\},$$

hence

$$\left\{\omega \in \Omega \;\middle|\; D_{\mathrm{E}}\left(\widehat{\mathcal{V}}_{n,N}, \mathcal{V}_n\right) > \epsilon\right\} \subset \left\{\omega \in \Omega \;\middle|\; \sup_{\boldsymbol{v} \in \mathbb{R}^n_{\leq}} \left|\widehat{h}_N(\boldsymbol{v}) - h(\boldsymbol{v})\right| > \frac{w'}{2}\right\},$$

and therefore

$$\begin{aligned}
&\sum_{N=1}^{\infty} \mathbb{P}\left(\left\{\omega \in \Omega \;\middle|\; D_{\mathrm{E}}\left(\widehat{\mathcal{V}}_{n,N}, \mathcal{V}_n\right) > \epsilon\right\}\right) \\
&\leq \sum_{N=1}^{\infty} \mathbb{P}\left(\left\{\omega \in \Omega \;\middle|\; \sup_{\boldsymbol{v} \in \mathbb{R}^n_{\leq}} \left|\widehat{h}_N(\boldsymbol{v}) - h(\boldsymbol{v})\right| > \frac{w'}{2}\right\}\right) \\
&< \infty
\end{aligned}$$

by Theorem A.13. □

COROLLARY A.18. *The estimator $\widehat{\boldsymbol{v}}_N \in \widehat{\mathcal{V}}_{n,N}$ converges completely to $\boldsymbol{c}$ as $N \to \infty$.*

PROOF. Since $\mathcal{V}_n = \{\boldsymbol{c}\}$ by Proposition 3.4, we have $D_{\mathrm{E}}(\widehat{\mathcal{V}}_{n,N}, \mathcal{V}_n) = \sup_{\boldsymbol{v} \in \widehat{\mathcal{V}}_{n,N}} \|\boldsymbol{v} - \boldsymbol{c}\| \geq \|\widehat{\boldsymbol{v}}_N - \boldsymbol{c}\|$. Hence the claim follows from Theorem A.17. □

THEOREM A.19. *The estimator $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N, N}$ converges completely to $\rho(\pi_1 f_1, \pi_2 f_2)$ as $N \to \infty$.*

PROOF. From (4) and (6), we have

$$\rho(\pi_1 f_1, \pi_2 f_2) = \sum_{k=1}^{n+1} \min_j \left\{\pi_j \left[F_j(c_k) - F_j(c_{k-1})\right]\right\}, \tag{17}$$

$$\widehat{\rho}_{\widehat{\boldsymbol{v}}_N, N} = \sum_{k=1}^{n+1} \min_j \left\{\widehat{\pi}_{j,N} \left[\widehat{F}_{j,N}(\widehat{v}_k) - \widehat{F}_{j,N}(\widehat{v}_{k-1})\right]\right\}, \tag{18}$$

where $\widehat{\boldsymbol{v}}_N = (\widehat{v}_1, \ldots, \widehat{v}_n) \in \widehat{\mathcal{V}}_{n,N}$, $\widehat{v}_0 = -\infty$, and $\widehat{v}_{n+1} = \infty$. By Lemma A.12,

$$\begin{aligned}
&\left|\widehat{\rho}_{\widehat{\boldsymbol{v}}_N, N} - \rho(\pi_1 f_1, \pi_2 f_2)\right| \\
&\leq \sum_{k=1}^{n+1} \sum_{j=1}^{2} \left|\widehat{\pi}_{j,N} \left[\widehat{F}_{j,N}(\widehat{v}_k) - \widehat{F}_{j,N}(\widehat{v}_{k-1})\right] - \pi_j \left[F_j(c_k) - F_j(c_{k-1})\right]\right|,
\end{aligned}$$

where

$$
\begin{aligned}
&\left|\widehat{\pi}_{j,N}\left[\widehat{F}_{j,N}(\widehat{v}_k)-\widehat{F}_{j,N}(\widehat{v}_{k-1})\right]-\pi_j\left[F_j(c_k)-F_j(c_{k-1})\right]\right|\\
&\le\left|\widehat{\pi}_{j,N}\left[\widehat{F}_{j,N}(\widehat{v}_k)-\widehat{F}_{j,N}(\widehat{v}_{k-1})\right]-\pi_j\left[F_j(\widehat{v}_k)-F_j(\widehat{v}_{k-1})\right]\right|\\
&\qquad+\left|\pi_j\left[F_j(\widehat{v}_k)-F_j(\widehat{v}_{k-1})\right]-\pi_j\left[F_j(c_k)-F_j(c_{k-1})\right]\right|\\
&\le\left|\widehat{\pi}_{j,N}\widehat{F}_{j,N}(\widehat{v}_k)-\pi_jF_j(\widehat{v}_k)\right|+\left|\widehat{\pi}_{j,N}\widehat{F}_{j,N}(\widehat{v}_{k-1})-\pi_jF_j(\widehat{v}_{k-1})\right|\\
&\qquad+\pi_j\left|F_j(\widehat{v}_k)-F_j(c_k)\right|+\pi_j\left|F_j(\widehat{v}_{k-1})-F_j(c_{k-1})\right|\\
&\le\left|\widehat{\pi}_{j,N}\widehat{F}_{j,N}(\widehat{v}_k)-\pi_j\widehat{F}_{j,N}(\widehat{v}_k)\right|+\left|\pi_j\widehat{F}_{j,N}(\widehat{v}_k)-\pi_jF_j(\widehat{v}_k)\right|\\
&\qquad+\left|\widehat{\pi}_{j,N}\widehat{F}_{j,N}(\widehat{v}_{k-1})-\pi_j\widehat{F}_{j,N}(\widehat{v}_{k-1})\right|+\left|\pi_j\widehat{F}_{j,N}(\widehat{v}_{k-1})-\pi_jF_j(\widehat{v}_{k-1})\right|\\
&\qquad+\pi_j\left|F_j(\widehat{v}_k)-F_j(c_k)\right|+\pi_j\left|F_j(\widehat{v}_{k-1})-F_j(c_{k-1})\right|\\
&\le 2\left|\widehat{\pi}_{j,N}-\pi_j\right|+\pi_j\left|\widehat{F}_{j,N}(\widehat{v}_k)-F_j(\widehat{v}_k)\right|+\pi_j\left|\widehat{F}_{j,N}(\widehat{v}_{k-1})-F_j(\widehat{v}_{k-1})\right|\\
&\qquad+\pi_j\left|F_j(\widehat{v}_k)-F_j(c_k)\right|+\pi_j\left|F_j(\widehat{v}_{k-1})-F_j(c_{k-1})\right|.
\end{aligned}
$$

Hence

$$
\begin{aligned}
&\left|\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}-\rho(\pi_1f_1,\pi_2f_2)\right|\\
&\le 2(n+1)\sum_{j=1}^{2}\left|\widehat{\pi}_{j,N}-\pi_j\right|\\
&\quad+\sum_{k=1}^{n+1}\sum_{j=1}^{2}\pi_j\left|\widehat{F}_{j,N}(\widehat{v}_k)-F_j(\widehat{v}_k)\right|\\
&\quad+\sum_{k=1}^{n+1}\sum_{j=1}^{2}\pi_j\left|\widehat{F}_{j,N}(\widehat{v}_{k-1})-F_j(\widehat{v}_{k-1})\right|\\
&\quad+\sum_{k=1}^{n+1}\sum_{j=1}^{2}\pi_j\left|F_j(\widehat{v}_k)-F_j(c_k)\right|+\sum_{k=1}^{n+1}\sum_{j=1}^{2}\pi_j\left|F_j(\widehat{v}_{k-1})-F_j(c_{k-1})\right|\\
&=2(n+1)\sum_{j=1}^{2}\left|\widehat{\pi}_{j,N}-\pi_j\right|\\
&\quad+2\sum_{k=1}^{n}\sum_{j=1}^{2}\pi_j\left|\widehat{F}_{j,N}(\widehat{v}_k)-F_j(\widehat{v}_k)\right|+2\sum_{k=1}^{n}\sum_{j=1}^{2}\pi_j\left|F_j(\widehat{v}_k)-F_j(c_k)\right|.
\end{aligned}\tag{19}
$$

For any $\epsilon > 0$, there exists $\delta > 0$ such that $|F_j(x) - F_j(c_k)| < \epsilon/(6n)$ for all $x \in \mathbb{R}$ with $|x - c_k| < \delta$ $(j = 1, 2;\ k = 1, \dots, n)$. If

$$|\widehat{\pi}_{j,N} - \pi_j| < \frac{\epsilon}{12(n+1)}, \qquad \sup_{x\in\mathbb{R}} \left|\widehat{F}_{j,N}(x) - F_j(x)\right| < \frac{\epsilon}{6n}, \qquad |\widehat{v}_k - c_k| < \delta,$$

for $j = 1, 2$ and $k = 1, \dots, n$, then

$$\begin{aligned}
&\left|\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N} - \rho(\pi_1 f_1, \pi_2 f_2)\right| \\
&< 2(n+1)\frac{2\epsilon}{12(n+1)} + 2n(\pi_1+\pi_2)\frac{\epsilon}{6n} + 2n(\pi_1+\pi_2)\frac{\epsilon}{6n} = \epsilon
\end{aligned}$$

by (19). Hence $\{\omega \in \Omega \mid |\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N} - \rho(\pi_1 f_1, \pi_2 f_2)| > \epsilon\}$ is contained in

$$\begin{aligned}
&\bigcup_{j=1}^{2} \left\{\omega \in \Omega \;\middle|\; |\widehat{\pi}_{j,N} - \pi_j| > \frac{\epsilon}{24(n+1)}\right\} \\
&\cup \bigcup_{j=1}^{2} \left\{\omega \in \Omega \;\middle|\; \sup_{x\in\mathbb{R}} \left|\widehat{F}_{j,N}(x) - F_j(x)\right| > \frac{\epsilon}{12n}\right\} \\
&\cup \left\{\omega \in \Omega \;\middle|\; \|\widehat{\boldsymbol{v}}_N - \boldsymbol{c}\| > \frac{\delta}{2}\right\},
\end{aligned}$$

and therefore

$$\begin{aligned}
&\sum_{N=1}^{\infty} \mathbb{P}\left(\left\{\omega \in \Omega \mid \left|\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N} - \rho(\pi_1 f_1, \pi_2 f_2)\right| > \epsilon\right\}\right) \\
&\leq \sum_{j=1}^{2} \sum_{N=1}^{\infty} \mathbb{P}\left(\left\{\omega \in \Omega \;\middle|\; |\widehat{\pi}_{j,N} - \pi_j| > \frac{\epsilon}{24(n+1)}\right\}\right) \\
&\quad + \sum_{j=1}^{2} \sum_{N=1}^{\infty} \mathbb{P}\left(\left\{\omega \in \Omega \;\middle|\; \sup_{x\in\mathbb{R}} \left|\widehat{F}_{j,N}(x) - F_j(x)\right| > \frac{\epsilon}{12n}\right\}\right) \\
&\quad + \sum_{N=1}^{\infty} \mathbb{P}\left(\left\{\omega \in \Omega \;\middle|\; \|\widehat{\boldsymbol{v}}_N - \boldsymbol{c}\| > \frac{\delta}{2}\right\}\right) \\
&< \infty
\end{aligned}$$

by Theorem A.10, Proposition A.11, and Corollary A.18. □

Note that Corollary A.18 and Theorem A.19 are exactly Theorems 4.5 and 4.6, respectively.

As stated above, we have estimated $\boldsymbol{c}$ as $\widehat{\boldsymbol{v}}_N \in \widehat{\mathcal{V}}_{n,N}$. In fact, it is possible to estimate $\boldsymbol{c}$ in another way. For $\boldsymbol{v} = (v_1, \ldots, v_m) \in \mathbb{R}^m_{\leq}$ with m a positive integer, let us define

$$\rho_{\boldsymbol{v}} = \sum_{k=1}^{m+1} \min_j \{\varphi_j(v_k, v_{k-1})\}, \tag{20}$$

where $v_0 = -\infty$ and $v_{m+1} = \infty$. Note that we have

$$\widehat{\rho}_{\boldsymbol{v},N} = \sum_{k=1}^{m+1} \min_j \{\widehat{\varphi}_j(v_k, v_{k-1})\} \tag{21}$$

by (6). Here recall that

$$\begin{aligned} \varphi_j(v_k, v_{k-1}) &= \pi_j[F_j(v_k) - F_j(v_{k-1})], \\ \widehat{\varphi}_{j,N}(v_k, v_{k-1}) &= \widehat{\pi}_{j,N}[\widehat{F}_{j,N}(v_k) - \widehat{F}_{j,N}(v_{k-1})]. \end{aligned}$$

LEMMA A.20. *For $\boldsymbol{v} = (v_1, \ldots, v_m) \in \mathbb{R}^m_{\leq}$ with m a positive integer, we have*

$$h(\boldsymbol{v}) + \rho_{\boldsymbol{v}} = 1 - \max_j \{\pi_j\}, \tag{22}$$

$$\widehat{h}_N(\boldsymbol{v}) + \widehat{\rho}_{\boldsymbol{v},N} = 1 - \max_j \{\widehat{\pi}_{j,N}\}. \tag{23}$$

PROOF. For $k = 1, \ldots, m+1$, choose

$$\begin{aligned} j_k &\in \arg\max_j \{\varphi_j(v_k, v_{k-1})\}, \\ l_k &\in \arg\min_j \{\varphi_j(v_k, v_{k-1})\} \end{aligned}$$

such that $\{j_k, l_k\} = \{1, 2\}$, where $v_0 = -\infty$ and $v_{m+1} = \infty$. By (15) and (20), we have

$$h(\boldsymbol{v}) + \rho_{\boldsymbol{v}} = \sum_{k=1}^{m+1} \left[\max_j \{\varphi_j(v_k, v_{k-1})\} + \min_j \{\varphi_j(v_k, v_{k-1})\} \right] - \max_j \{\pi_j\}$$

$$
\begin{aligned}
&= \sum_{k=1}^{m+1} \left[\varphi_{j_k}(v_k, v_{k-1}) + \varphi_{l_k}(v_k, v_{k-1})\right] - \max_j \{\pi_j\} \\
&= \sum_{k=1}^{m+1} \left[\varphi_1(v_k, v_{k-1}) + \varphi_2(v_k, v_{k-1})\right] - \max_j \{\pi_j\} \\
&= \sum_{j=1}^{2} \sum_{k=1}^{m+1} \pi_j \left[F_j(v_k) - F_j(v_{k-1})\right] - \max_j \{\pi_j\} \\
&= \sum_{j=1}^{2} \pi_j \left[F_j(\infty) - F_j(-\infty)\right] - \max_j \{\pi_j\} \\
&= 1 - \max_j \{\pi_j\},
\end{aligned}
$$

which implies (22).

We can prove (23) in a similar way. For $k = 1, \ldots, m+1$, redefine

$$
\begin{aligned}
j_k &\in \arg\max_j \{\widehat{\varphi}_j(v_k, v_{k-1})\}, \\
l_k &\in \arg\min_j \{\widehat{\varphi}_j(v_k, v_{k-1})\}
\end{aligned}
$$

such that $\{j_k, l_k\} = \{1, 2\}$. By (16) and (21), we have

$$
\begin{aligned}
&\widehat{h}_N(\boldsymbol{v}) + \widehat{\rho}_{\boldsymbol{v},N} \\
&= \sum_{k=1}^{m+1} \left[\max_j \{\widehat{\varphi}_{j,N}(v_k, v_{k-1})\} + \min_j \{\widehat{\varphi}_{j,N}(v_k, v_{k-1})\}\right] - \max_j \{\widehat{\pi}_{j,N}\} \\
&= \sum_{k=1}^{m+1} \left[\widehat{\varphi}_{j_k,N}(v_k, v_{k-1}) + \widehat{\varphi}_{l_k,N}(v_k, v_{k-1})\right] - \max_j \{\widehat{\pi}_{j,N}\} \\
&= \sum_{k=1}^{m+1} \left[\widehat{\varphi}_{1,N}(v_k, v_{k-1}) + \widehat{\varphi}_{2,N}(v_k, v_{k-1})\right] - \max_j \{\widehat{\pi}_{j,N}\} \\
&= \sum_{j=1}^{2} \sum_{k=1}^{m+1} \widehat{\pi}_{j,N} \left[\widehat{F}_{j,N}(v_k) - \widehat{F}_{j,N}(v_{k-1})\right] - \max_j \{\widehat{\pi}_{j,N}\} \\
&= \sum_{j=1}^{2} \widehat{\pi}_{j,N} \left[\widehat{F}_{j,N}(\infty) - \widehat{F}_{j,N}(-\infty)\right] - \max_j \{\widehat{\pi}_{j,N}\}
\end{aligned}
$$

$$= 1 - \max_j \{\widehat{\pi}_{j,N}\},$$

which implies (23). □

It is immediate from Lemma A.20 that

$$\underset{\boldsymbol{v}\in\mathbb{R}^m_{\leq}}{\arg\min}\{\rho_{\boldsymbol{v}}\} = \mathcal{V}_m, \tag{24}$$

$$\underset{\boldsymbol{v}\in\widehat{\mathbb{R}}^m_N}{\arg\min}\{\widehat{\rho}_{\boldsymbol{v},N}\} = \widehat{\mathcal{V}}_{m,N} \tag{25}$$

for $m = 1, \ldots, n$.

Theorem A.21. *For $\boldsymbol{v} \in \mathbb{R}^n_{\leq}$, $\rho_{\boldsymbol{v}}$ attains its unique minimum $\rho(\pi_1 f_1, \pi_2 f_2)$ at $\boldsymbol{v} = \boldsymbol{c}$.*

Proof. This follows from Proposition 3.4, (4), and (24). □

Theorem A.22. *Let $\widehat{\boldsymbol{v}}'_N \in \arg\min_{\boldsymbol{v}\in\widehat{\mathbb{R}}^n_N}\{\widehat{\rho}_{\boldsymbol{v},N}\}$. Then $\widehat{\boldsymbol{v}}'_N$ converges completely to $\boldsymbol{c}$ as $N \to \infty$. Furthermore, $\widehat{\rho}_{\widehat{\boldsymbol{v}}'_N,N}$ converges completely to $\rho(\pi_1 f_1, \pi_2 f_2)$ as $N \to \infty$.*

Proof. Since $\widehat{\boldsymbol{v}}'_N \in \widehat{\mathcal{V}}_{n,N}$ by (25), the claim follows from Corollary A.18 and Theorem A.19. □

Appendix B. Measurability of Some Functions

B.1 The measurability of $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$ (associated with Theorems 4.6 and A.19)

It follows from (25) that $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N} = \min_{\boldsymbol{v}\in\widehat{\mathbb{R}}^n_N}\{\widehat{\rho}_{\boldsymbol{v},N}\}$, which depends only on the rank statistics of $X_1, ..., X_N$ (labeled by $Y_1, \ldots, Y_N$, respectively). We then see that $\{\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}(\omega) \mid \omega \in \Omega\}$ is a finite set and that $\widehat{\rho}_{\widehat{\boldsymbol{v}}_N,N}$ is a measurable simple function on Ω.

B.2 The measurability of $\sup_{\boldsymbol{v}\in\mathbb{R}^m_{\leq}} |\widehat{h}_N(\boldsymbol{v}) - h(\boldsymbol{v})|$ (associated with Theorem A.13)

By the right continuity of $\widehat{F}_{j,N}$ (Definition A.1), we see that $\sup_{\boldsymbol{v}\in\mathbb{R}^m_{\leq}} |\widehat{h}_N(\boldsymbol{v}) - h(\boldsymbol{v})| = \sup_{\boldsymbol{v}\in\mathbb{Q}^m_{\leq}} |\widehat{h}_N(\boldsymbol{v}) - h(\boldsymbol{v})|$ for any positive integer

m, where $\mathbb{Q}$ is the set of rational numbers and $\mathbb{Q}^m_{\leq} = \{(v_1, \ldots, v_m) \in \mathbb{Q}^m \mid v_1 \leq \cdots \leq v_m\}$. Since $\mathbb{Q}^m_{\leq}$ is countable and $\widehat{h}_N(\boldsymbol{v})$ is obviously measurable on Ω, $\sup_{\boldsymbol{v}\in\mathbb{R}^m_{\leq}} |\widehat{h}_N(\boldsymbol{v}) - h(\boldsymbol{v})|$ is also measurable on Ω.

B.3 The measurability of $D_{\mathrm{E}}(\widehat{\mathcal{V}}_{n,N}, \mathcal{V}_n)$ (associated with Theorem A.17)

Let $\{K_1, \ldots, K_m\}$ be the collection of all nonempty subsets of $\{(i_1, \ldots, i_n) \mid 1 \leq i_1 \leq \cdots \leq i_n \leq N-1\}$ with $K_j \neq K_l$ if $(j \neq l)$, Ω_{K_j} be the set of all $\omega \in \Omega$ such that $\widehat{\mathcal{V}}_{n,N} = \{(Z_{i_1}, \ldots, Z_{i_n}) \mid (i_1, \ldots, i_n) \in K_j\}$. Then $\Omega_{K_j} \in \mathcal{F}$ for all $j \in \{1, \ldots, m\}$, $\Omega = \cup_{j=1}^m \Omega_{K_j}$, and $\Omega_{K_j} \cap \Omega_{K_l} = \emptyset$ if $j \neq l$. Since the restriction of $D_{\mathrm{E}}(\widehat{\mathcal{V}}_{n,N}, \mathcal{V}_n)$ to each Ω_{K_j} coincides with $\max\{\|(Z_{i_1}, \ldots, Z_{i_n}) - \boldsymbol{c}\| \mid (i_1, \ldots, i_n) \in K_j\}$, which is measurable on Ω_{K_j}, we see that $D_{\mathrm{E}}(\widehat{\mathcal{V}}_{n,N}, \mathcal{V}_n)$ is measurable on Ω.

B.4 The measurability of $\widehat{\boldsymbol{v}}_N$ (associated with Theorem 4.5 and Corollary A.18)

We can choose $\widehat{\boldsymbol{v}}_N \in \widehat{\mathcal{V}}_{n,N}$ such that $\widehat{\boldsymbol{v}}_N : \Omega \to \mathbb{R}^n$ is measurable. Indeed, let $\mathcal{I} = \{(i_1, \ldots, i_n) \mid 1 \leq i_1 \leq \cdots \leq i_n \leq N-1\}$ and $\mathbb{Z}_{(i_1,\ldots,i_n)} = (Z_{i_1}, \ldots, Z_{i_n}) \in \widehat{\mathbb{R}}^n_N$ for $(i_1, \ldots, i_n) \in \mathcal{I}$. Note that Ω equals the disjoint union of measurable sets

$$\Omega_{\mathcal{J}} = \left\{\omega \in \Omega \;\middle|\; \widehat{h}_N(\mathbb{Z}_{\boldsymbol{j}}) = \max_{\boldsymbol{i}\in\mathcal{I}} \widehat{h}_N(\mathbb{Z}_{\boldsymbol{i}}) \text{ if and only if } \boldsymbol{j} \in \mathcal{J}\right\}$$

over all nonempty subsets $\mathcal{J}$ of $\mathcal{I}$. For such $\mathcal{J}$, we can define $\max \mathcal{J}$ and $\min \mathcal{J}$ in lexicographic order. If we put $\widehat{\boldsymbol{v}}_N = \mathbb{Z}_{\max \mathcal{J}}$ (or $\widehat{\boldsymbol{v}}_N = \mathbb{Z}_{\min \mathcal{J}}$) on each $\Omega_{\mathcal{J}}$, then $\widehat{\boldsymbol{v}}_N$ is measurable.

If we choose $\widehat{\boldsymbol{v}}_N \in \widehat{\mathcal{V}}_{n,N}$ at random independently of $(\Omega, \mathcal{F}, \mathbb{P})$, we cannot guarantee that $\widehat{\boldsymbol{v}}_N$ is measurable. In such a case, we mean by "$\widehat{\boldsymbol{v}}_N$ converges completely to $\boldsymbol{c}$ as $N \to \infty$" that for any $\epsilon > 0$, there exists a collection $\{A_1, A_2, \ldots\}$ of measurable sets such that $\sum_{N=1}^{\infty} \mathbb{P}(A_N) < \infty$ and $A_N \supset \{\omega \in \Omega \mid \|\widehat{\boldsymbol{v}}_N - \boldsymbol{c}\| > \epsilon\}$ for all N, which also implies that $\widehat{\boldsymbol{v}}_N$ converges almost surely to $\boldsymbol{c}$ (in the sense that $\mathbb{P}(\{\omega \in \Omega \mid \lim_{N\to\infty} \widehat{\boldsymbol{v}}_N = \boldsymbol{c}\}) = 1$) if $(\Omega, \mathcal{F}, \mathbb{P})$ is complete (see Remark 4.4). In fact, we can take $A_N = \{\omega \in \Omega \mid D_{\mathrm{E}}(\widehat{\mathcal{V}}_{n,N}, \mathcal{V}_n) > \epsilon\}$.

Appendix C. Additional Proofs

In this section, (9)–(12) will be proved. We shall take over the notations in Section 5.

PROPOSITION C.1. *In the first case,*

$$\begin{aligned} C(\pi_1 f_1, \pi_2 f_2) &= \{c_1\} = \{(\log 2)/2\}, \\ \rho(\pi_1 f_1, \pi_2 f_2) &= [2 - 2\Phi(c_1 + 1) + \Phi(c_1 - 1)]/3. \end{aligned}$$

PROOF. The equation $\pi_1 f_1(x) = \pi_2 f_2(x)$ gives $x = (\log 2)/2$, which is a crossover point. Hence $C(\pi_1 f_1, \pi_2 f_2) = \{c_1\} = \{(\log 2)/2\}$. Next,

$$\begin{aligned} \rho(\pi_1 f_1, \pi_2 f_2) &= \pi_2 F_2(c_1) + \pi_1[1 - F_1(c_1)] \\ &= \frac{1}{3}\Phi(c_1 - 1) + \frac{2}{3}[1 - \Phi(c_1 + 1)]. \ \square \end{aligned}$$

PROPOSITION C.2. *In the second case,*

$$\begin{aligned} C(\pi_1 f_1, \pi_2 f_2) &= \{c_1, c_2\} = \cosh^{-1}\left(0.8\sqrt{\mathrm{e}}\right), \\ \rho(\pi_1 f_1, \pi_2 f_2) &= 0.8 - 0.5\Phi(c_1 + 1) + 0.5\Phi(c_2 + 1) - 0.8\Phi(c_2). \end{aligned}$$

PROOF. If $x < 0$ or $x > 0.5$, then $f_2(x) = 0.8\nu_{0,1}(x)$, and $\pi_1 f_1(x) = \pi_2 f_2(x)$ gives $\cosh(x) = 0.8\sqrt{\mathrm{e}}$. There is a unique $c > 0$ such that $\cosh(c) = 0.8\sqrt{\mathrm{e}}$. Since $c > 0.5$ and $\pi_1 f_1 < \pi_2 f_2$ on $[0, 0.5]$, we have $C(\pi_1 f_1, \pi_2 f_2) = \{-c, c\} = \cosh^{-1}(0.8\sqrt{\mathrm{e}})$. Next,

$$\begin{aligned} \rho(\pi_1 f_1, \pi_2 f_2) &= \pi_2 F_2(-c) + \pi_1[F_1(c) - F_1(-c)] + \pi_2[1 - F_2(c)] \\ &= 0.8 - 0.5\Phi(-c + 1) + 0.5\Phi(c + 1) - 0.8\Phi(c). \ \square \end{aligned}$$

Acknowledgements. This study was partially supported by JSPS KAKENHI Grant Numbers JP15K04814, JP20K03509, and JP21K15762. We thank Atsushi Komaba (University of Yamanashi) for the validation of our numerical results.

References

[1] Breiman, L., Friedman, J. H., Olshen, R. A. and C. J. Stone, Classification and regression trees, Wadsworth Advanced Books and Software, Belmont, California, (1984).

[2] Hsu P. L. and H. Robbins, Complete convergence and the law of large numbers, Proceedings of the National Academy of Sciences of the United States of America **33** (1947), 25–31.

[3] Montoya, J. A., Figueroa P., G. and D. González-Sánchez, Statistical inference for the Weitzman overlapping coefficient in a family of distributions, Applied Mathematical Modelling **71** (2019), 558–568.

[4] Pastore, M. and A. Calcagnì, Measuring distribution similarities between samples: A distribution-free overlapping index, Frontiers in Psychology **10** (2019), 1089.

[5] Rudin, W., Principles of mathematical analysis (third edition), *International series in pure and applied mathematics*, McGraw-Hill, (1976).

[6] Schmid, F. and A. Schmidt, Nonparametric estimation of the coefficient of overlapping—theory and empirical application, Computational Statistics & Data Analysis **50** (2006), 1583–1596.

[7] Serfling, R. J., Approximation theorems of mathematical statistics, John Wiley & Sons, New York, (1980).

(Received December 17, 2021)
(Revised September 14, 2022)

Hisashi Johno
Department of Radiology
Faculty of Medicine
University of Yamanashi
Japan
E-mail: johnoh@yamanashi.ac.jp

Kazunori Nakamoto
Center for Medical Education and Sciences
Faculty of Medicine
University of Yamanashi
Japan
E-mail: nakamoto@yamanashi.ac.jp