Uniform confidence band for optimal transport map on one-dimensional data

Donlapark Ponnoprat

Chiang Mai University
e-mail: donlapark.p@cmu.ac.th

Ryo Okano

The University of Tokyo
e-mail: okano-ryo1134@g.ecc.u-tokyo.ac.jp

Masaaki Imaizumi

The University of Tokyo / RIKEN Center for Advanced Intelligence Project
e-mail: imaizumi@g.ecc.u-tokyo.ac.jp

Abstract: We develop a statistical inference method for an optimal transport map between distributions on real numbers with uniform confidence bands. The concept of optimal transport (OT) is used to measure distances between distributions, and OT maps are used to construct the distance. OT has been applied in many fields in recent years, and its statistical properties have attracted much interest. In particular, since the OT map is a function, a uniform norm-based statistical inference is significant for visualization and interpretation. In this study, we derive a limit distribution of a uniform norm of an estimation error for the OT map, and then develop a uniform confidence band based on it. In addition to our limit theorem, we develop a bootstrap method with kernel smoothing, then also derive its validation and guarantee on an asymptotic coverage probability of the confidence band. Our proof is based on the functional delta method and the representation of OT maps on the reals.

1. Introduction

We consider the framework of optimal transport (OT) and develop a statistical inference method on an OT map used in the concept. Specifically, we consider the Monge problem; for probability measure P and Q on $\mathbb{R}$, we consider the following optimization problem

$$\min_{T \in \mathcal{T}(P,Q)} \int_{\mathbb{R}} h(|x - T(x)|) dP(x), \tag{1}$$

where $h : \mathbb{R} \to \mathbb{R}$ is a nonnegative convex function (where $h(r) = r$ or $h(r) = r^2$ in most applications) and $\mathcal{T}(P, Q)$ is a set of measurable maps $T : \mathbb{R} \to \mathbb{R}$ such that $P(T^{-1}(A)) = Q(A)$ for any measurable $A \subset \mathbb{R}$. Let $T_0 \in \mathcal{T}(P, Q)$ be an OT map which is the minimizer of the problem (1). Our goal is the statistical inference on the OT map from samples: we construct a confidence band for T_0 based on samples independently generated from P and Q, respectively. To this end, we develop estimators for the OT map estimator and derive a limiting distribution of their estimation error.

OT is a general framework to measure the distance between probability measures, and has attracted attention in a wide range of fields related to data analysis, such as social science, statistics, machine learning, and image analysis [ACB17, Gal18, TGR21]. For a general textbook, see [V$^+$09, PC$^+$19, Vil21]. The formulation of OT is given by the Monge problem (1) or the Kantorovich problem related to the Wasserstein distance [Kan60], in which the OT map plays an important role. OT plays a major role in modern data science due to its flexible extensibility and adaptability to high-dimensional data.

Statistical analysis for OT has played an important role in the usage of OT with finite samples, which are useful in assessing the uncertainty of estimated elements of the OT framework. The most representative analysis examines the sample complexity of estimation methods for OT. One of the typical interests is the problem of estimating the sum of optimized transport costs, which corresponds to the Wasserstein distance or its variants, and numerous studies have proposed optimal estimation methods [WB19, GCB$^+$19, NWR22]. In recent years, several studies [HR21, DGS21] have developed estimation methods for OT maps and also investigated their theoretical accuracy. On the other hand, statistical inference for OT, i.e., statistical tests and confidence sets, has also attracted attention. In addition to inference methods on OT costs and distances [dBL19, SM18, OI23], [GKRS22a, SGK23, MBNWW23] propose inference methods on OT maps in several settings.

Despite the importance, the statistical inference for OT maps is still a developing issue, even when the sample is one-dimensional. This is because that OT maps are functions and hence it is nontrivial to handle its uncertainty properly in a function space. In particular, when investigating a global shape of an OT map (e.g., monotonicity) in some application fields, it is necessary to examine errors of the OT map estimator in the uniform norm sense, which is a nontrivial challenge.

In this paper, we develop a framework of confidence bands for an OT map between two unknown one-dimensional distributions. Specifically, we develop a *uniform* confidence band that measures the OT map considering all input points in its support. A uniform confidence band has the following advantages. First, it allows us to study the global shape of the OT map to be estimated. This advantage can not be achieved by arranging pointwise confidence intervals, which generally results in highly discontinuous curves. Second, it can be visualized on a plot and hence is easy to interpret, in contrast to confidence intervals based on the L^2-norm, which cannot be easily visualized. Our approach is to use kernel smoothing on the distribution functions, and then perform a bootstrap scheme

with kernel smoothing on the estimated distributions in order to construct a confidence band. As a theoretical contribution, we derive the asymptotic distribution of the estimation error in order to validate our confidence band.

Our contributions can be summarized as follows:

- We develop the statistical inference on OT maps by proposing the uniform confidence bands that are computationally tractable. Coverage probabilities of the proposed bands are theoretically validated, and our simulations support the validity.
- The uniform confidence bands allow us to study a global property of functions, which can investigate complex hypotheses on OT maps. Our real-data experiments with population data demonstrate that our method rigorously tests a hypothesis whether the population pyramid changes faster, by examining a difference in shape between the OT map and a linear function.

As a technical contribution, we derive a simple asymptotic representation of the OT map estimation error as a linear sum of estimation errors for cumulative distribution functions. From this, we can apply the functional delta method to the representation, thereby obtaining the asymptotic distribution of the OT map estimation error.

1.1. Related studies

Optimal transport has been actively studied in recent years. For a comprehensive review or textbook, see [V⁺09, PC⁺19, Vil21]. It is known that there are several variations of the OT problem, such as the entropic regularization [Cut13] or a low-dimensional projection named slicing [BRPP15]. Since there are numerous papers on OT, we refer to only a few that are relevant to our study.

Many studies have examined the sample complexity of the problem of estimating elements of OT from observed samples.

A typical estimation target is a sum of optimized transport costs. In the setting of one-dimensional data, [BL19] gives a comprehensive analysis for the statistical properties of OT costs. [Dud69, WB19, MNW21] studied the estimation of the cost and reported that an optimal convergence rate of the estimation error is slow depending on the dimension of data, i.e., there is the curse of dimensionality. Related to this curse, [NWR22, GCB⁺19, LFH⁺20, Lei20] reported that the curse can be reduced by introducing the low-dimensional projection or the entropic regularization to the OT problem. The estimation error of an OT map is also a target of intense study. [HR21] has shown that an estimation error of OT maps also suffers from the curse of dimensionality of data, as is the case for the sum of costs. Similarly, entropic regularization and other techniques have been shown to mitigate this rate [DGS21, MBNWW21, PNW21, DNWP22].

For statistical inference, it is also a concern to derive limiting distributions of an error of estimators. Importantly, when data are multi-dimensional, it is not easy to obtain the limiting distributions relevant to the OT problem. Therefore, statistical inference is performed for the OT problem when the data are

one-dimensional or discrete, or when some types of regularization are introduced. For a limiting distribution of estimating a sum of transportation costs, [MC98, RGTC17, DBCAMRR99, DBGU05] studied the one-dimensional case, and [SM18, BCP19, TSM19, KTM20, OI23] studied the discrete data case. [MNW19, dBGSL24, dBSLNW23, GKRS22b, MBW22, IOH22] investigated a case with the regularization. [DR23] constructs a confidence interval based on the OT distance in the problem of estimating density functions. This situation is similar to the limiting distribution of estimating OT maps. [GKRS22a] developed a statistical inference method on an OT map for the regularized OT problem case, and [SGK23] derived a limiting distribution of an estimator for an OT map with a setting of semi-discrete data. [MBNWW23] investigated the availability of limiting distributions in the multi-dimensional case and derived a pointwise convergence to a limiting distribution under certain conditions.

1.2. Notation

A *cadlag function* is a right continuous function whose limits from the left exist everywhere. $D[a, b]$ denotes the space of all cadlag functions $z : [a, b] \to [-1, 1]$ equipped with the uniform norm. For a topological space Ω, $BL_1(\Omega)$ denotes the space of Lipschitz functions $h : \Omega \to [-1, 1]$ with Lipschitz constants bounded by one. $\| \cdot \|$ denotes the Euclidean norm. $\|f\|_\infty = \sup_t |f(t)|$ denotes the sup-norm. $\ell^\infty(\mathcal{F})$ denotes the set of all uniformly bounded real functions on $\mathbb{R}$. δ_x is the Dirac measure at x. With an event E, $\mathbf{1}\{E\}$ is an indicator function.

For a distribution function $F : \Omega \to [0, 1]$ with a support $\Omega \subset \mathbb{R}$ and $p \in [0, 1]$, $F^{-1}(p) := \inf\{x \in \Omega : p \leq F(x)\}$ denotes a quantile function of F. Given probability measures P and Q on a support $\Omega \subseteq \mathbb{R}^d$, we define a set of transport maps $\mathcal{T}(P, Q)$, which is a set of maps such that $T_\sharp P := P(T^{-1}(\cdot)) = Q$. For random sequences X_n and Y_m, the notation $(X_n, Y_m) = o_{P \otimes Q}(1)$ means (X_n, Y_m) converges to $(0, 0)$ in $P \otimes Q$-probability as $n, m \to \infty$ and possibly other conditions on n and m (see Section 2.1 below). $\xrightarrow{d}$ means convergence in distribution. For non-random sequences of reals $\{x_n\}_{n \in \mathbb{N}}$ and $\{y_n\}_{n \in \mathbb{N}}$, $x_n \asymp y_n$ denotes $\lim_{n \to \infty} x_n / y_n \to c$ with some $c \in (0, \infty)$.

For any non-negative integer, β and open set $U \subseteq \mathbb{R}$, $C^\beta(U)$ is the space of all bounded continuous real-valued functions that are s-times differentiable on U. This can be extended to the notion of Hölder space: $C^\beta(U)$ is the space of functions $f \in C^{\lfloor \beta \rfloor}(U)$ ($\lfloor \beta \rfloor$ is the integer part of β) whose $\lfloor \beta \rfloor$-th derivative is Hölder continuous with exponent $\beta - \lfloor \beta \rfloor$, that is, there is some constant $C > 0$ such that for all $x, y \in U$, $|D^{\lfloor \beta \rfloor}(x) - D^{\lfloor \beta \rfloor}(y)| \leq C|x - y|^{\beta - \lfloor \beta \rfloor}$ holds. We say that a function f is (continuously) differentiable on $[a, b]$ if there is a small $\varepsilon > 0$ such that f is (continuously) differentiable on $(a - \varepsilon, b + \varepsilon)$.

1.3. Paper organization

Section 2 gives the optimal transport problem and its associated statistical inference problem. Section 3 develops a uniform confidence band of the OT map

as our proposed methodology. Section 4 shows theoretical guarantees of the developed confidence band with an outline of the proof. Section 5 conducts a numerical simulation to show the experimental validity of the proposed confidence band. Section 6 handles a real data analysis to demonstrate the usefulness of our method. Section 7 finally concludes our study.

2. Problem setting

2.1. Optimal transport problem with samples

We consider the setup with $d = 1$. Let F_P and F_Q be distribution functions of P and Q, respectively. We shall make an assumption that F_P is continuous, which guarantees a solution to the OT map problem (1) of the form (see e.g. [Vil21, Remarks 2.19]):

$$T_0(x) = F_Q^{-1} \circ F_P(x). \tag{2}$$

Throughout the paper, we will only consider the values of the OT map T_0 on a closed interval $[a, b]$. For simplicity, we make an abuse of notation and treat T_0 and members of $\mathcal{T}(P, Q)$ as functions on $[a, b]$.

Suppose that we observe n i.i.d. samples $X_1, \ldots, X_n \sim P$ and m i.i.d. samples $Y_1, \ldots, Y_m \sim Q$. For simplicity, we assume that n and m follow the following asymptotic that often appears in nonparametric two-sample tests: with some $\kappa \in (0, 1)$:

$$n/(n + m) \to \kappa. \tag{3}$$

Our goal with this problem is to infer the true OT map T_0 using the samples. Specifically, we construct an estimator of T_0 based on the sample, as well as the following statistical inference.

2.2. Statistical inference via confidence band

We aim to develop an estimator for the OT map T_0 and conduct statistical inference for T_0 from n observations. Specifically, with pre-specified $\alpha \in (0, 1)$, we will develop a set $\mathcal{C}^{(\alpha)} = \{[\underline{c}(x), \overline{c}(x)] \mid x \in [a, b]\}$ with some functions $\underline{c}(x), \overline{c}(x)$ depending on the n samples and m samples and shows that

$$\mathbb{P}(T_0 \in \mathcal{C}^{(\alpha)}) = 1 - \alpha + o(1),$$

as $n, m \to \infty$ with the limit ratio (3).

Importantly, our interest is a confidence interval using the sup-norm, rigorously, our theory will show that $T_0(x) \in [\underline{c}(x), \overline{c}(x)]$ holds for every $x \in [a, b]$ with the asymptotic probability $1 - \alpha$. The sup-norm is useful for intuitive analysis because it provides visualization in a functional space, whereas several common norms, e.g., the L^2-norm, cannot be visualized on a plot.

3. Methodology

In this section, we describe our methodology for constructing a confidence band for the OT map. First, we propose a kernel estimator for the OT map (Section 3.1). We then construct a confidence band based on this estimator and a bootstrap scheme (Section 3.2).

3.1. Kernel estimator for OT map

Suppose that we observe the samples $X_1, \ldots, X_n \sim P$ and $Y_1, \ldots, Y_m \sim Q$. We estimate density and distribution functions from the samples, then use them to construct an estimator for the OT map T_0.

First, we define a kernel density estimator. As a preparation, we define a kernel function $K : \mathbb{R} \to \mathbb{R}$, which should be a positive function and satisfy $\int K(t)dt = 1$. There are several common choices: the Gaussian kernel, the Epanechinikov kernel, and so on (for an overview, see [Sil86]). Then, we define density estimators as

$$\widehat{f}_P(x) := \frac{1}{nr_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{r_n}\right) \quad \text{and} \quad \widehat{f}_Q(y) := \frac{1}{mr_m} \sum_{j=1}^{m} K\left(\frac{y - Y_j}{r_m}\right),$$

where $r_n, r_m > 0$ are bandwidth parameters dependent on n and m, respectively. Second, we define estimators for distribution functions F_P and F_Q with $\overline{K}(x) := \int_{-\infty}^{x} K(u)du$ as follows:

$$\widehat{F}_P(x) := \int_{-\infty}^{x} \widehat{f}_P(u)du = \frac{1}{n} \sum_{i=1}^{n} \overline{K}\left(\frac{x - X_i}{r_n}\right)$$

and

$$\widehat{F}_Q(y) := \int_{-\infty}^{y} \widehat{f}_Q(u)du = \frac{1}{m} \sum_{j=1}^{m} \overline{K}\left(\frac{y - Y_j}{r_m}\right).$$

Since $\widehat{f}_P$ and $\widehat{f}_Q$ are strictly positive by the positive property of the kernel K, $\widehat{F}_P$ and $\widehat{F}_Q$ are strictly increasing function and hence invertible.

We define an estimator for the OT map T_0 using the kernel estimators. Using the form (2), we define the kernel smoothed estimator for the OT map T_0:

$$\widehat{T}_{n,m}(x) = \widehat{F}_Q^{-1} \circ \widehat{F}_P(x).$$

This estimator has several advantages. First, we can achieve a smooth estimation, and hence a smooth confidence band, for any input x. Second, we can guarantee the asymptotic validity of the uniform confidence band using smoothness. Third, as will be shown later, we can develop a bootstrap method with kernel smoothing for practical use and show its convergence. We will compare this kernel-based approach with a pointwise estimator using empirical distributions in Remark 2.

3.1.1. Assumption

For the kernel estimation, we give the following assumptions on the distribution functions F_P, F_Q, and the kernel K:

Assumption 1. *F_P and F_Q are differentiable on $\mathbb{R}$. For some $\beta \geq 0$, the densities $f_P := F_P'$ and $f_Q := F_Q'$ satisfy $f_P, f_Q \in C^\beta(\mathbb{R})$. Further, the kernel K is of order $> \beta + 1/2$, that is, $\int K(u) = 1$ and $\int u^j K(u)du = 0$ holds for $j = 1, 2, \ldots, s$ for $s > \beta + 1/2$. Also, the bandwidth parameter is $r_n \asymp n^{-1/(2\beta+1)}$.*

These conditions are widespread in density estimation using kernel methods. Regarding the order of the kernel, for example, [Tsy08] describes a method to construct kernels of arbitrary order using Legendre polynomials. The setup of the bandwidth parameter is commonly used as well, which is designed to balance bias and variance in density function estimation. See [Tsy08] for details.

Remark 1 (Choice of kernels and bandwidth)**.** The choice of the kernel K and bandwidth r_n in Assumption 1 is one of the methods used to implement *undersmoothing*, which is common in constructions of confidence bands as summarized in Section 5.7 of [Was06]. In our design, we increase the order of the kernel K from β to $\beta + 1/2$ while keeping the bandwidth as $n^{-1/(2\beta+1)}$, the usual choice that leads to optimal rates of convergence in kernel density estimation. This approach with the larger order kernels follows Corollary 2 of [GN07], which in turn allows us to utilize the functional delta method in the proof of our main theorem below.

We further put the following assumption on a density function.

Assumption 2. *f_Q is positive on $[F_Q^{-1}(F_P(a)), F_Q^{-1}(F_P(b))]$ and K has bounded variation.*

We remark that Assumption 2 is more flexible than assuming that f_Q is positive on $\mathbb{R}$ as it allows distributions that are supported on a proper subset of $\mathbb{R}$, such as gamma or chi-squared distributions.

3.2. Construction of confidence band

We develop our methodology to construct a confidence band for T_0 based on the estimator $\widehat{T}_{n,m}$.

3.2.1. Bahadur representation of estimation error

Our basic strategy is a Gaussian approximation of the estimation error $\widehat{T}_{n,m}(x) - T_0(x)$. Specifically, we investigate the following scaled estimation error

$$\sqrt{n+m} \sup_x \frac{|\widehat{T}_{n,m}(x) - T_0(x)|}{s_\kappa(x)}, \tag{4}$$

then show that it converges in distribution to a limiting Gaussian process as $n, m \to \infty$, where $s_\kappa(x)$ is a standard deviation of $\widehat{T}_{n,m}(x) - T_0(x)$. To achieve the Gaussian limit of (4), the following asymptotic linear form, named the *Bahadur representation*, plays an important role:

Proposition 1 (Bahadur Representation)**.** *Suppose that Assumption 1 and 2 hold. Then, for any $x \in [a, b]$, we have*

$$\sqrt{n+m} \left(\widehat{T}_{n,m}(x) - T_0(x) \right)$$
$$= \frac{\sqrt{n+m}}{n} \sum_{i=1}^{n} \widehat{\psi}(X_i, x) + \frac{\sqrt{n+m}}{m} \sum_{j=1}^{m} \widehat{\zeta}(Y_j, x) + o_{P \otimes Q}(1), \qquad (5)$$

where

$$\widehat{\psi}(X_i, x) := \frac{1}{f_Q(T_0(x))} \left[\overline{K} \left(\frac{x - X_i}{r_n} \right) - F_P(x) \right],$$

and

$$\widehat{\zeta}(Y_j, x) := \frac{1}{f_Q(T_0(x))} \left[F_P(x) - \overline{K} \left(\frac{T_0(x) - Y_j}{r_m} \right) \right].$$

Since we assume that Q has a density f_Q that is nowhere zero, the representation holds for all $x \in [a, b]$. With this asymptotic linear representation, we can guarantee the existence of Gaussian processes in the limit. See Theorem 3 below for a formal result.

3.2.2. Plugin-estimator for standard deviation

To use the limiting Gaussian process in practice, we should derive several values. First, we consider the standard deviation term $s_\kappa(x)$ of $\widehat{T}_{n,m}(x) - T_0(x)$ appearing in (4). In view of (5), we consider the following form of the standard deviation as

$$s_\kappa(x) = \frac{1}{f_Q(T_0(x))} \sqrt{\left(\frac{1}{\kappa} + \frac{1}{1-\kappa} \right) F_P(x)(1 - F_P(x))},$$

in which $F_Q(T_0(x)) = F_Q(F_Q^{-1}(F_P(x))) = F_P(x)$ plays an important role using the continuity of F_Q. In practice, we do not know F_P, F_Q, and f_Q, so it is impossible to compute $s_\kappa(x)$. Instead, we estimate $s_\kappa(x)$ by the plug-in estimator:

$$\widehat{s}_{n,m}(x) = \frac{1}{\widehat{f}_Q(\widehat{T}_{n,m}(x))} \sqrt{\left(\frac{n+m}{n} + \frac{n+m}{m} \right) \widehat{F}_P(x)(1 - \widehat{F}_P(x))}.$$

Note that the plug-in estimator always exists by the positivity of $\widehat{f}_Q$. The following result shows its consistency:

Lemma 2. *Suppose that Assumption 1 and 2 hold. Then, we have the following convergence:*

$$\sup_{x \in [a,b]} \left| \frac{\widehat{s}_{n,m}(x)}{s_\kappa(x)} - 1 \right| \xrightarrow{P \otimes Q} 0. \tag{6}$$

3.2.3. Bootstrap approach with kernel smoothing and confidence band

We approximate the Gaussian process for which (4) converges by a distribution generated by a bootstrap method. Specifically, we develop a bootstrap method with kernel smoothing which newly generates samples from the estimated distribution functions $\widehat{F}_P$ and $\widehat{F}_Q$ by the smooth kernels. In the bootstrap scheme, we sample $X_1^*, \ldots, X_n^* \sim \widehat{F}_P$ and $Y_1^*, \ldots, Y_m^* \sim \widehat{F}_Q$. Define bootstrap distribution functions $\widehat{F}_P^*(x) := n^{-1} \sum_{i=1}^n \mathbf{1}\{X_i^* \leq x\}$ and $\widehat{F}_Q^*(y) := m^{-1} \sum_{j=1}^m \mathbf{1}\{Y_j^* \leq y\}$. Then, we consider the bootstrap estimator for the OT map T_0 as

$$\widehat{T}_{n,m}^* := \widehat{F}_Q^{*-1}(\widehat{F}_P^*(x))$$

Note that X_i^* and Y_j^* are not subsamples of the dataset, but are generated from $\widehat{F}_P$ and $\widehat{F}_Q$. This approach is more suitable when we apply the functional delta method to validate a confidence band in our proof.

Using the distribution of the bootstrap estimator $\widehat{T}_{n,m}^*$, we derive quantiles of the distribution of the bootstrap version of the estimation error. Let $\widehat{P}_n$ and $\widehat{Q}_m$ denote the conditional probability given $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$, respectively. For any $\alpha \in (0,1)$ define

$$\widehat{q}_{n,m}(1-\alpha) := \text{the } (1-\alpha)\text{-quantile of } \sqrt{n+m} \sup_x \frac{\left| \widehat{T}_{n,m}^*(x) - \widehat{T}_{n,m}(x) \right|}{\widehat{s}_{n,m}(x)},$$

under $\widehat{\mathbb{P}}_{n,m} := \widehat{P}_n \otimes \widehat{Q}_m$. Then, we propose the bootstrap confidence band

$$\mathcal{C}_{n,m}^{(\alpha)}(x) := \left[\widehat{T}_{n,m}(x) \pm \frac{\widehat{s}_{n,m}(x)\widehat{q}_{n,m}(1-\alpha)}{\sqrt{n+m}} \right], \ x \in [a,b] \tag{7}$$

Note that except for the bandwidth r_n, this confidence band is computed in a data-driven way. Also, we will later propose a method to select r_n based on the observed samples.

Remark 2 (Comparison with a pointwise confidence interval)**.** Another natural approach is to construct a pointwise confidence interval for the OT map by plugging in the empirical distributions. To complement the main study, we present in Section A.4 a methodology for constructing a pointwise confidence interval, along with proof of its asymptotic validity and its empirical evaluations. Compared to this pointwise approach, our main proposed kernel-based method has several advantages. First, it produces smooth confidence bands, which benefit interpretability. In contrast, pointwise confidence intervals based

on the empirical distributions are riddled with discontinuities and can be difficult to interpret. Second, the uniform confidence band can be used to infer the OT map across the whole domain. For instance, a 95%-level uniform confidence band has asymptotically 95% chance to cover the true OT map over the domain. In contrast, a 95%-level pointwise confidence interval never contains all, but asymptotically only 95% of the true OT map.

Remark 3 (Relation to ROC curves)**.** We discuss a relation of the confidence band for OT maps to that for ROC (receiver operating characteristic) curves. ROC curves have a similar form $F_P \circ F_Q^{-1}$ and its confidence analysis has been developed by [HHZ08]. As a point of distinction between our study and the existing work, we develop the confidence band whose width differs for each input $x \in [a, b]$. Rigorously, we have introduced the standard deviation $s_\kappa(x)$ and its estimator, then our confidence band achieves the adaptive widths for each input x. This result is in contrast to the confidence band by [HHZ08] for ROC curves, which has a constant width independent of x.

4. Theoretical result

In this section, we present the main theoretical contributions of this paper, namely the bootstrap consistency (Theorem 3) and the asymptotic validity of the confidence band $\mathcal{C}_{n,m}^{(\alpha)}$ (Corollary 4).

We first state the theorems in Section 4.1, and then provide an outline of the proof of Theorem 3 in Section 4.2. The full proofs of the theorems can be found in Appendix C and D.

4.1. Validity of confidence band

We start with a consistency result of the bootstrap estimator. Notice the inclusions of the supremums on the left-hand sides of (8) and (9), which are essential for obtaining a *uniform* confidence band of T_0.

Theorem 3 (Bootstrap consistency)**.** *Suppose that the distribution functions F_P and F_Q satisfy Assumption 1 and 2. Then, there are independent Brownian bridges $\mathbb{G}_1$ and $\mathbb{G}_2$ such that for any $x_0 > 0$, the followings holds as $n, m \to \infty$ and $n/(n + m) \to \kappa$:*

$$\sqrt{n + m} \sup_x \frac{\left|\widehat{T}_{n,m}(x) - T_0(x)\right|}{\widehat{s}_{n,m}(x)} \xrightarrow{d} \sup_x |\mathsf{Z}_\kappa(x)|, \tag{8}$$

and

$$\widehat{\mathbb{P}}_{n,m} \left(\sqrt{n + m} \sup_x \frac{\left|\widehat{T}_{n,m}^*(x) - \widehat{T}_{n,m}(x)\right|}{\widehat{s}_{n,m}(x)} \leq x_0 \right) \xrightarrow{P \otimes Q} \mathbb{P}\Big(\sup_x |\mathsf{Z}_\kappa(x)| \leq x_0 \Big), \tag{9}$$

where

$$Z_\kappa(x) = \frac{1}{f_Q(T_0(x))s_\kappa(x)} \left[\sqrt{1/\kappa}\mathbb{G}_1(F_P(x)) - \sqrt{1/(1-\kappa)}\mathbb{G}_2(F_Q(x)) \right]$$

$$= \frac{\sqrt{1/\kappa}\mathbb{G}_1(F_P(x)) - \sqrt{1/(1-\kappa)}\mathbb{G}_2(F_P(x))}{\sqrt{(\kappa^{-1} + (1-\kappa)^{-1})F_P(x)(1-F_P(x))}}, \tag{10}$$

is a Gaussian process with unit variance for any $x \in \mathbb{R}$.

This theorem implies that the supremum values of both the scaled estimation error by the kernel estimator and the estimation error by the bootstrap estimator converge in distribution to the supremum of the same Gaussian process. In essence, the convergence (8) is the intermediate result on the estimation error of the kernel method, and the convergence (9) additionally provides the convergence of the bootstrap method.

Based on this result, we show the asymptotic validity of the proposed confidence band:

Corollary 4 (Asymptotic Validity of Bootstrap Confidence Band)**.** *Consider the proposed confidence band $\mathcal{C}_{n,m}^{(\alpha)}$ in (7). Then, under Assumption 1 and 2, the bootstrap confidence band $\mathcal{C}_{n,m}^{(\alpha)}$ is asymptotically uniformly consistent at level $1 - \alpha$, that is, it holds that*

$$\mathbb{P}\left(T_0(x) \in \mathcal{C}_{n,m}^{(\alpha)}(x), \forall x \in [a,b]\right) \to 1 - \alpha,$$

as $n, m \to \infty$ and $n/(n+m) \to \kappa$.

This result shows that our confidence bands are asymptotically valid in a uniform sense, that is, the OT map T_0 is included in our confidence band for every input x simultaneously with the probability.

4.2. Proof outline of Theorem 3

Our proof consists of two parts: the first is the convergence of the estimated distributions by the kernel of the developed bootstrap method, and the second is the convergence of an application of the functional delta method. The details are described below.

4.2.1. Convergence of estimated distributions

As a preparation, we first derive limiting Gaussian processes of the distributions $\widehat{F}_P$, $\widehat{F}_Q$, $\widehat{F}_P^*$, and $\widehat{F}_Q^*$.

We first describe the analysis of the distributions $\widehat{F}_P^*$ and $\widehat{F}_Q^*$ by the bootstrap method with kernel smoothing. Rigorously, the central limit theorem in

[KMT76] shows that there exists a Browninan bridge $\mathbb{G}_1$ such that the following holds for every $\delta > 0$:

$$\widehat{P}_n \left(\sup_x \left| \sqrt{n+m}(\widehat{F}_P^*(x) - \widehat{F}_P(x)) - \sqrt{1/\kappa}\mathbb{G}_1(F_P(x)) \right| > \delta \right) \xrightarrow{P\text{-a.s.}} 0,$$

as $n \to \infty$. This shows that the error $\widehat{F}_P^*(x) - \widehat{F}_P(x)$ by the bootstrap with kernel smoothing converges to the Brownian bridge with the proper scaling in the uniform norm sense. For technical reasons in further proof, we also derive the convergence in the sense of bounded Lipschitz metrics, that is, we obtain

$$\sup_{h \in BL_1(\ell^\infty(\mathcal{F}))} \left| \mathbb{E}_n h(\sqrt{n+m}(\widehat{F}_P^* - \widehat{F}_P)) - \mathbb{E}h(\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P) \right| \xrightarrow{P\text{-a.s.}} 0,$$

as $n \to \infty$. Here, we denote by $\mathbb{E}_n$ the expectation with respect to $\widehat{P}_n$. Similarly, we can obtain a similar limiting statement for the error $\widehat{F}_Q^*(x) - \widehat{F}_Q(x)$ by the other distribution by the bootstrap.

Next, we also analyze the error by the estimated distributions $\widehat{F}_P$ and $\widehat{F}_Q$ by the kernel method. We apply the seminal analysis on the convergence of kernel convolutions [GN07] and obtain the following joint convergence:

$$\sqrt{n+m}(\widehat{F}_P - F_P, \widehat{F}_Q - F_Q) \xrightarrow{d} (\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P, \sqrt{1/(1-\kappa)}\mathbb{G}_2 \circ F_Q),$$

as $n \to \infty$. Here, $\mathbb{G}_1$ and $\mathbb{G}_2$ are some independent Brownian bridges.

4.2.2. Functional delta method

We study the convergence of the estimator $\widehat{T}_{n,m}^*$ by using the above convergence results of the distributions and the representation (2) of the OT map. To the aim, we use the functional delta method (see Appendix B for a brief exposition).

Formally, we define a functional $\phi : D[a,b] \times D[a,b] \to D[a,b]$ as

$$\phi(u,v) = (v^{-1} \circ u(\cdot)),$$

which implies that $\widehat{T}_{n,m}^* = \phi(\widehat{F}_P^*, \widehat{F}_Q^*)$, $\widehat{T}_{n,m} = \phi(\widehat{F}_P, \widehat{F}_Q)$, and $T_0 = \phi(F_P, F_Q)$. Then, we shall make a first-order approximation $\phi(\widehat{F}_P, \widehat{F}_Q) - \phi(F_P, F_Q) \approx \phi'(\widehat{F}_P - F_P, \widehat{F}_Q - F_Q)$. Thus it is important to first derive the *Hadamard* derivative ϕ'.

Lemma 5. *Let $[a,b]$ satisfy the conditions in Proposition 6. Define the functional $\phi : D[a,b] \times D[a,b] \to D[a,b]$ by $\phi(u,v) = v^{-1} \circ u$. Then, ϕ is Hadamard differentiable at (F_P, F_Q). Denoting $T_0 = F_Q^{-1} \circ F_P$, the Hadamard derivative of ϕ at (F_P, F_Q) is given by*

$$\phi'(u,v) = \frac{1}{f_Q \circ T_0} \left[u - v \circ T_0 \right]. \tag{11}$$

We slightly extend this derivative for the design of confidence bands. Define a functional $\Psi : D[a,b] \times D[a,b] \to D[a,b]$ as

$$\Psi(u,v) = (v^{-1} \circ u(\cdot))/s_\kappa(\cdot).$$

Using Lemma 5, we derive its derivative as

$$\Psi'(u,v) = (u(\cdot) - v \circ T_0(\cdot))/((f_Q \circ T)(\cdot)s_\kappa(\cdot)).$$

Note that we have added the term $s_\kappa(\cdot)$, which determines the scale of the confidence band.

Finally, we apply the functional delta method (Lemma 10 in Appendix) and study the limit of the estimation error of the estimator $\widehat{F}_P^*$ as

$$
\begin{aligned}
\sqrt{n+m}(\widehat{T}_{n,m}^* - \widehat{T}_{n,m})(\cdot)/s_\kappa(\cdot) &= \sqrt{n+m}(\Psi(\widehat{F}_P^*, \widehat{F}_Q^*) - \Psi(\widehat{F}_P, \widehat{F}_Q)) \\
&\xrightarrow{d} \Psi'(\sqrt{1/\kappa}\,\mathbb{G}_1 \circ F_P, \sqrt{1/(1-\kappa)}\,\mathbb{G}_2 \circ F_Q) \\
&= \mathsf{Z}_\kappa,
\end{aligned}
$$

where Z_κ is defined in (10). By a similar discussion, we also prove that the estimation error $\widehat{T}_{n,m} - T_0$ of the kernel estimator also converges to the same Gaussian process. In addition, we give evaluations of several plug-in estimators such as $\widehat{s}_{n,m}(\cdot)$, then obtain the statement of Theorem 3.

5. Simulation

5.1. Simulation design

To support our asymptotic validity results, we perform a Monte Carlo simulation to evaluate the coverage probabilities of the confidence bands. For 1,000 iterations, we sample from two different probability distributions: $X_1, \ldots, X_n \sim N(0,1)$ and $Y_1, \ldots, Y_m \sim \text{Gamma}(5, 0.5)$, where $n \in \{100, 200, \ldots, 2000\}$ and $m = n/4$ (that is, $\kappa = 0.2$). In each iteration, we use 2,500 bootstrap samples to construct $(1 - \alpha)$-level uniform confidence bands on the interval $[-2.5, 2.5]$, where $\alpha \in \{0.01, 0.05, 0.10\}$. The true OT map can be directly computed as $T_0 = F_Q^{-1} \circ F_P$ where F_P and F_Q are the distribution function of $N(0,1)$ and $\text{Gamma}(5, 0.5)$, respectively. The Gaussian kernel is used for the uniform confidence bands with various smoothness parameters $\beta \in \{0.5, 1.0, 1.4\}$ and $r_n = \frac{1}{2}n^{-1/(2\beta+1)}$; so if we assume that the density functions are in $C^\beta(\mathbb{R})$, then Assumptions 1 and 2 are satisfied.

To evaluate our confidence bands, we estimate the coverage probability as the proportion of 1000 runs in which $T_0(x)$ is contained in the confidence band *for all* $x \in [-2.5, 2.5]$. Additionally, we assess the confidence bands by calculating the median of the bands' average widths.

Evaluations of our $(1-\alpha)$-level uniform confidence bands of the optimal transport map from $N(0,1)$ to Gamma$(5,0.5)$ *based on 1,000 pairs of samples from each distribution. The table displays the median of average widths and the coverage probabilities of the confidence bands on $[-2.5, 2.5]$.*

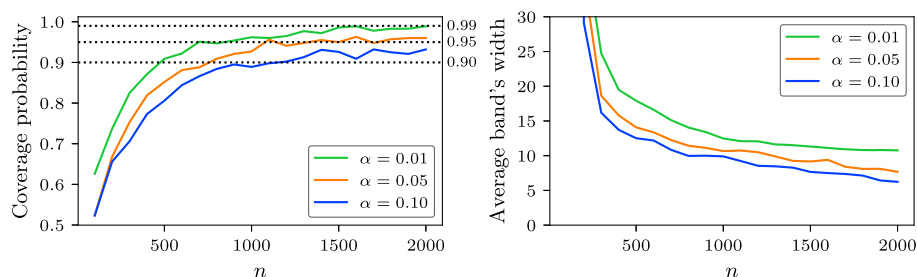| $1-\alpha$	n	m	Average width	Coverage probability
	200	50	29.13	0.656
0.90	700	175	10.85	0.866
	2000	500	6.23	**0.932**
	200	50	35.67	0.668
0.95	700	175	12.26	0.888
	2000	500	7.68	**0.960**
	200	50	40.94	0.737
0.99	700	175	15.13	0.951
	2000	500	10.77	**0.989**

FIG 1. *Plots of the coverage probabilities (left) and the median of average widths (right) of the simulated uniform confidence bands on $[-2.5, 2.5]$ as functions of sample size n.*

5.2. Result and discussion

The median of per-iteration average widths and the coverage probabilities over $x \in [-2.5, 2.5]$ for $\beta = 0.5$ and $n = 200, 700$ and 2000 are shown in Table 1. From the table, we can see that the coverage probabilities approach the nominal probabilities $(1-\alpha)$, and the widths become smaller as n increases. In particular, when $n = 2000$ and $\alpha = 0.90$ or 0.95, the coverage probabilities are slightly larger than the nominal probabilities.

The plots in Figure 1 illustrate the coverage probability and median of average width as functions of n. These plots lead us to the same conclusion: as n increases, the average coverage probabilities approach the nominal probabilities, and the width of the band decreases.

We now examine the uniform confidence bands of a specific sample with $\beta = 0.5, 1.0$ and 1.4 (recall Assumption 1 that $\beta + 0.5$ must be less than the order of the Gaussian kernel, which is 2). The plots of the true optimal transport maps, their kernel estimates, and the uniform confidence bands are shown in Figure 2. We observe that for $x > 1.8$, the estimated transport map (the red curve) remains significantly distant from the actual transport map (the black curve). This disparity arises due to the heavier right tail of Gamma$(5,0.5)$ compared to
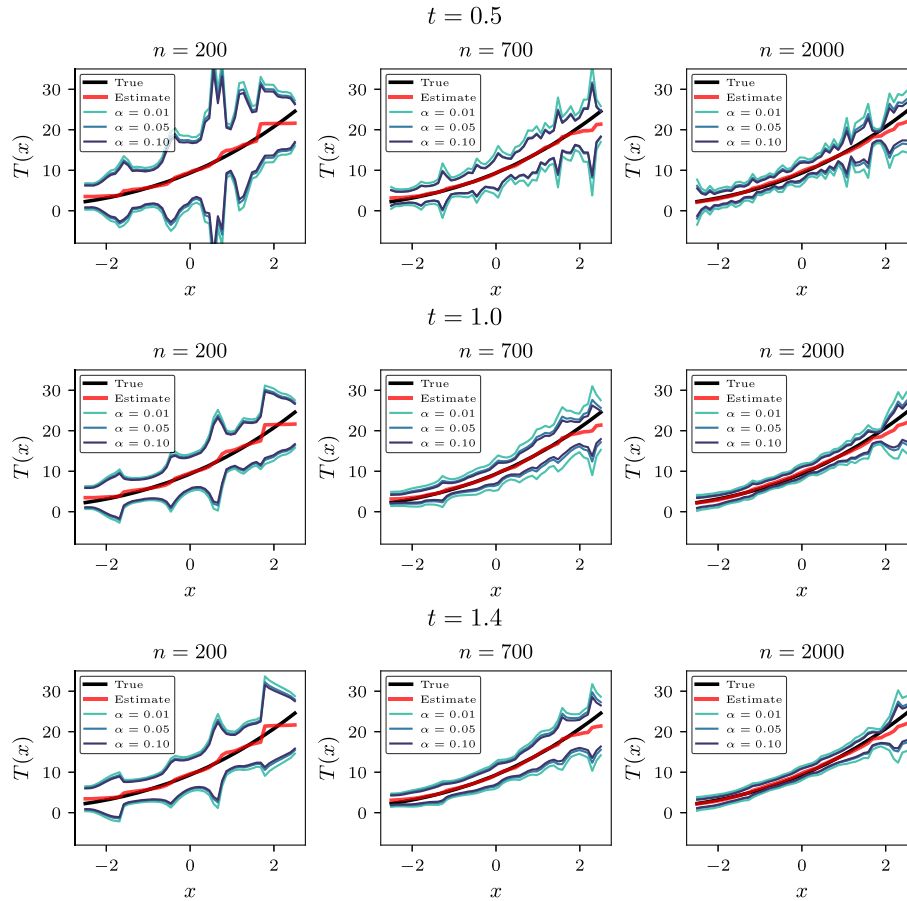
$$t = 0.5$$

$$t = 1.0$$

$$t = 1.4$$

FIG 2. *Examples of* $(1 - \alpha)$*-level uniform confidence bands on* $[-2.5, 2.5]$*, for three different values of* α *and three different values of the bandwidth parameter* β*, based on specific samples of size* $200, 700$ *and* 2000.

that of $N(0, 1)$. Consequently, there is an inadequate number of sample points on the right tail of $N(0, 1)$ to estimate the transport map between the two distributions.

As β increases, the kernel bandwidths increase and the confidence bands become smoother. Note that if the actual density functions are rougher than $C^\beta(\mathbb{R})$, the kernel estimate and the confidence band might be too smooth.

6. Real data analysis

As an application, we use our confidence bands to assess the uncertainty of our estimate of the transport map of the distribution of ages of death in 2001 to those in 2021. The data of the age of deaths from 12 countries were taken from
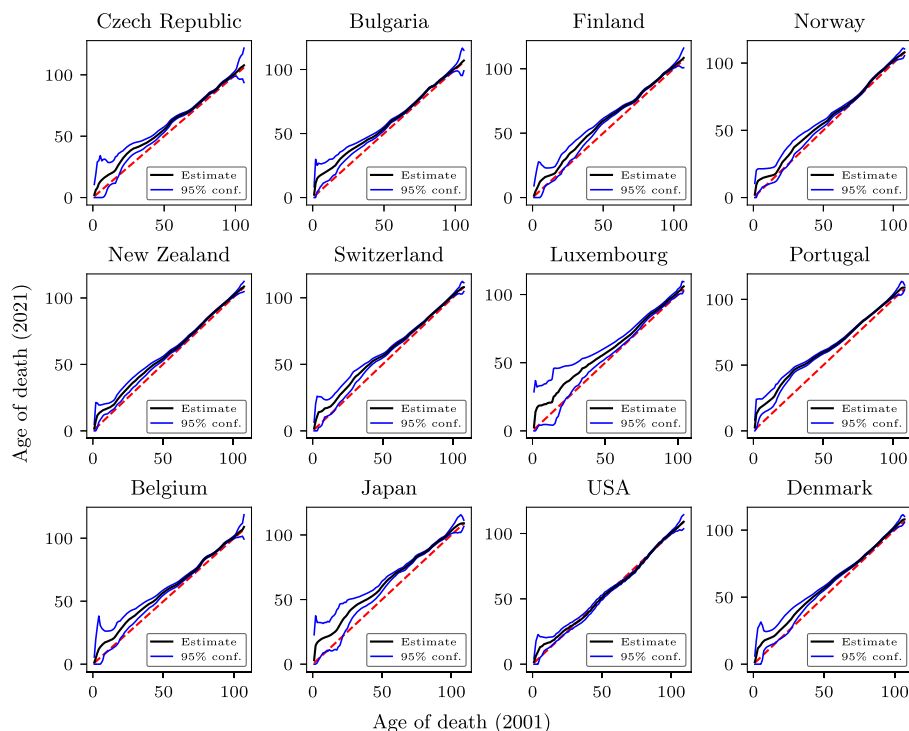
FIG 3. *Analysis of the distribution shifts in ages of death from the year 2001 to 2021 using our uniform confidence bands.*

the Human Mortality Database [Max23]. For each country, let us simply denote the dataset from the year 2001 by X, and those from the year 2021 by Y. Let $|X| = n$ and $|Y| = m$. Assume that the observed age of deaths in 2001 and 2021 are sampled from two separate continuous probability distributions. Then there is some uncertainty in our estimate due to randomness in the sampling.

To construct the estimators and confidence bands at level $1 - \alpha$, we use the Gaussian kernel. Our choices of bandwidths are guided by our theory in Section 3. Recall from Assumption 1 that $r_n \approx n^{-1/(2\beta+1)}$ where $\beta + 0.5$ must be less than the order of the kernel. Since the Gaussian kernel is of second order, any $\beta < 1.5$ is permissible. In particular, we choose $\beta = 1.25$ so that $2\beta + 1 = 3.5$; this leads to our choices of kernel bandwidths $r_n = 2n^{-1/3.5}$ for X, and $r_m = 2m^{-1/3.5}$ for Y. With these bandwidths, we use the method in Section 3 to construct a kernel estimate of the optimal transport map and a 95% uniform confidence band for each country.

The plots of our estimates and confidence bands for 12 countries are shown in Figure 3. The kernel estimate of the optimal transport map is the black curve, showing the correspondence between the age of deaths in 2001 and 2021. The identity function is the red dashed line. The estimate lying above the identity

function indicates a shift in mortality towards higher ages. From these estimates and the confidence bands, we observe the most significant age shifts in Portugal (from 30 to 45) and Japan (from 30 to 40). Other countries also show minor age shifts, except for Norway, New Zealand, Luxembourg, and the USA, where we cannot confidently assert an upward shift in ages.

7. Conclusion

In this paper, we develop a method to construct uniform confidence bands for the optimal transport maps based on two samples from two unknown continuous distributions. First, we use a kernel to estimate the densities, and then we use the empirical bootstrap to construct the confidence bands. We show that our confidence bands are asymptotically valid, meaning that they contain the actual OT map on an interval with a probability that approaches the nominal coverage probability. We perform simulations to verify the validity of our confidence bands. As an application, we apply our confidence bands to analyze the shift in life expectancy across 12 countries from the year 2001 to 2021.

There are a couple of directions for future research. Firstly, our delta method and bootstrap procedure rely on first-order approximations. Exploring higher-order approximations would be an improvement worth considering. Secondly, our choice of kernel bandwidth directly follows from the theory of kernel density estimation, which, in practice, may not be sample-efficient in achieving consistency. This raises another research problem of finding a bandwidth search procedure that can achieve consistency more efficiently than the one presented in this paper.

Appendix A: Pointwise confidence intervals via the empirical distributions

In certain situations, one might wish to construct a confidence *interval* for the value of the optimal transport at a specific point. Our approach to constructing such an interval follows closely to that of confidence bands, as the interval can be seen as a specific instance of confidence bands that covers only a single point. One key distinction is that, with only a single point, there is no need to estimate the standard deviation, and consequently removing the necessity for kernel density estimation.

A.1. Methodology

We first develop an empirical estimator for T_0. We denote $\widetilde{P}_n := n^{-1} \sum_{i=1}^{n} \delta_{X_i}$ and $\widetilde{Q}_m := m^{-1} \sum_{j=1}^{m} \delta_{Y_i}$ as empirical measures with the observations. With the empirical distribution functions $\widetilde{F}_P(x) = n^{-1} \sum_{i=1}^{n} \mathbf{1}\{X_i \leq x\}$ and $\widetilde{F}_Q(y) = m^{-1} \sum_{j=1}^{m} \mathbf{1}\{Y_j \leq y\}$, we define a plug-in estimator as the OT map from $\widetilde{F}_P$

to $\widetilde{F}_Q$:

$$\widetilde{T}_{n,m}(x) = \widetilde{F}_Q^{-1} \circ \widetilde{F}_P(x).$$

A.2. Bahadur representation

Proposition 6 (Bahadur Representation of 1D Transport Map)**.** *Suppose that there exists an interval* $[a,b]$ *such that: (i)* F_Q *is continuously differentiable on the interval* $[F_Q^{-1}(F_P(a)), F_Q^{-1}(F_P(b))]$, *(ii)* $f_Q = F_Q'$ *is nonzero on* $[F_Q^{-1}(F_P(a)), F_Q^{-1}(F_P(b))]$. *Then, for any* $x \in [a,b]$, *we have*

$$\sqrt{n+m}\left(\widetilde{T}_{n,m}(x) - T_0(x)\right)$$

$$= \frac{\sqrt{n+m}}{f_Q(T_0(x))}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{X_i \le x\} - \frac{1}{m}\sum_{j=1}^{m}\mathbf{1}\{Y_j \le T_0(x)\}\right] + o_{P\otimes Q}(1)$$

$$= \frac{\sqrt{n+m}}{n}\sum_{i=1}^{n}\psi(X_i, x) - \frac{\sqrt{n+m}}{m}\sum_{j=1}^{m}\zeta(Y_j, x) + o_{P\otimes Q}(1), \qquad (12)$$

where

$$\psi(X_i, x) := \frac{1}{f_Q(T_0(x))}\left[\mathbf{1}\{X_i \le x\} - F_P(x)\right],$$

and

$$\zeta(Y_j, x) := \frac{1}{f_Q(T_0(x))}\left[\mathbf{1}\{Y_j \le T_0(x)\} - F_P(x)\right].$$

Moreover, the $o_{P\otimes Q}(1)$ *term does not depend on the choice of* x.

The conditions on F_Q imply that F_Q is continuous on $[F_Q^{-1}(F_P(a)), F_Q^{-1}(F_P(b))]$, which implies $F_Q \circ T_0 = F_Q \circ F_Q^{-1} \circ F_P = F_P$ on $[a,b]$.

We derive an asymptotic representation of the estimation error $\widetilde{T}_{n,m}(x) - T_0(x)$.

Proof of Lemma 5. For $t \ge 0$, let $u_t \to u$ and $v_t \to v$ be functions in $D[a,b]$ such that $F_{P,t} = F_P + tu_t$ and $F_{Q,t} = F_Q + tv_t$ are in $D[a,b]$ for each t. To compute the Hadamard derivative, we consider the difference quotient:

$$\frac{F_{Q,t}^{-1} \circ F_{P,t} - F_Q^{-1} \circ F_P}{t} = \frac{F_{Q,t}^{-1} \circ F_t - G^{-1} \circ F_t}{t} + \frac{F_Q^{-1} \circ F_{P,t} - F_Q^{-1} \circ F_P}{t} \quad (13)$$

We now apply the Taylor approximation for F_Q^{-1}. Since F_Q' is continuous and nonzero on $[F_Q^{-1}(F_P(a)), F_Q^{-1}(F_P(b))]$, it follows from the inverse function theorem that F_Q^{-1} is differentiable on $[F_Q(F_Q^{-1}(F_P(a))), F_Q(F_Q^{-1}(F_P(b)))] = [F_P(a), F_P(b)]$ and we have

$$F_Q^{-1}(F_{P,t}(x)) = F_Q^{-1}(F_P(x) + tu_t(x))$$

$$= F_Q^{-1}(F(x)) + \frac{1}{f_Q(F_Q^{-1}(F_P(x)))} t u_t(x) + o(t u_t(x)).$$

As $\|u_t\|_\infty \leq 1$, we have $o(t u_t(x)) = o(t)$ independent of x. From this, we approximate the second term on the right in (13).

$$\frac{F_Q^{-1} \circ F_{P,t} - F_Q^{-1} \circ F_P}{t} = \frac{1}{f_Q \circ F_Q^{-1} \circ F_P} u_t + o(1), \tag{14}$$

as $t \to 0$. From [vdV98, Lemma 21.3], the Hadamard derivative of the quantile function $Q \mapsto F_Q^{-1}$ at Q is $v \mapsto -\frac{v \circ F_Q^{-1}}{f_Q \circ F_Q^{-1}}$. From this, we use Taylor's formula again to obtain:

$$F_{Q,t}^{-1} \circ F_{P,t} = F_Q^{-1} \circ F_{P,t} - \frac{v \circ F_Q^{-1} \circ F_{P,t}}{f_Q \circ F_Q^{-1} \circ F_{P,t}} t + o(t).$$

Therefore, the first term on the right of (13) is

$$\frac{F_{Q,t}^{-1} \circ F_{P,t} - F_Q^{-1} \circ F_{P,t}}{t} = F_Q^{-1} \circ F_{P,t} - \frac{v \circ F_Q^{-1} \circ F_{P,t}}{f_Q \circ F_Q^{-1} \circ F_{P,t}} + o(1). \tag{15}$$

Combining (14) and (15) and the continuity of $v \circ F_Q^{-1}$ and $g \circ F_Q^{-1}$, we conclude the convergence in the Hadamard sense, that is, it holds that

$$\lim_{t \to 0} \frac{F_{Q,t}^{-1} \circ F_{P,t} - F_Q^{-1} \circ F_P}{t} = \frac{1}{f_Q \circ F_Q^{-1} \circ F_P} \left[u - v \circ F_Q^{-1} \circ F_P \right]$$

$$= \frac{1}{f_Q \circ T} \left[u - v \circ T \right]. \qquad \square$$

Proof of Proposition 6. We first recall that F_Q has a nonzero derivative at $F_Q^{-1}(F_P(x))$ for any $x \in [a, b]$; this implies that F_Q is locally invertible at $F_Q^{-1}(F_P(x))$, and so $F_Q(T_0(x)) = F_Q(F_Q^{-1}(F_P(x))) = F_P(x)$. This allows us to simplify the Hadamard derivative (11) evaluated at $u_i(x) = \mathbf{1}\{X_i \leq x\} - F_P(x)$ and $v_j(y) = \mathbf{1}\{Y_j \leq y\} - F_Q(y)$ as follows:

$$\phi'(u_i, v_j)(x)$$
$$= \frac{1}{f_Q(T_0(x))} \left[\mathbf{1}\{X_i \leq x\} - F_P(x) - \mathbf{1}\{Y_j \leq F_Q^{-1}(F_P(x))\} + F_Q(T_0(x)) \right]$$
$$= \frac{1}{g(T_0(x))} \left[\mathbf{1}\{X_i \leq x\} - F_P(x) - \mathbf{1}\{Y_j \leq T_0(x)\} + F_P(x) \right]$$
$$= \psi(X_i, x) - \zeta(Y_j, x).$$

Using the functional delta method (Lemma 9) and the linearity of ϕ', we arrive at the final approximation.

$$\sqrt{n+m} \left(\widetilde{T}_{n,m}(x) - T_0(x) \right)$$

$$= \sqrt{n+m} \left(\phi(\widetilde{F}_{P,n}, \widetilde{F}_{Q,m}) - \phi(F, G) \right)$$

$$= \sqrt{n+m}\phi'(\widetilde{F}_{P,n} - F_P, \widetilde{F}_{Q,m} - F_Q) + o_{P\otimes Q}(1)$$

$$= \frac{\sqrt{n+m}}{n} \sum_{i=1}^{n} \psi(X_i, x) + \frac{\sqrt{n+m}}{m} \sum_{j=1}^{m} \zeta(Y_j, x) + o_{P\otimes Q}(1).$$

$\square$

A.3. Gaussian approximation theorem

With the Bahadur representation (12), we can easily develop a Gaussian approximation on the estimation error. If F_Q an interval $[a, b]$ satisfy the conditions in Proposition 6, we obtain the following representation of the error on $[a, b]$:

$$\sqrt{n+m} \left(\widetilde{T}_{n,m}(x) - T_0(x) \right)$$

$$= \frac{\sqrt{n+m}}{n} \sum_{i=1}^{n} \psi(X_i, x) - \frac{\sqrt{n+m}}{m} \sum_{j=1}^{m} \zeta(Y_j, x) + o_{P\otimes Q}(1) \qquad (16)$$

Note that the two processes are independent, since $\{X_i\}_{t=1}^{n}$ and $\{Y_j\}_{j=1}^{m}$ are mutually independent.

We consider Gaussian limits of the terms in (16). In view of Proposition 6, Donsker's theorem tells us that there exist two independent Brownian bridges $\mathbb{G}_1$ and $\mathbb{G}_2$ such that the following convergences of $D[a, b]$ functions hold under the uniform norm:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i, x) = \frac{1}{f_Q(T_0(x))} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\mathbf{1}\{X_i \le x\} - F_P(x)]$$

$$\xrightarrow{d} \frac{1}{f_Q(T_0(x))} \mathbb{G}_1(F_P(x)),$$

and

$$\frac{1}{\sqrt{m}} \sum_{j=1}^{m} \xi(Y_j, x) = \frac{1}{f_Q(T_0(x))} \frac{1}{\sqrt{m}} \sum_{j=1}^{m} [\mathbf{1}\{Y_j \le T_0(x)\} - F_P(x)]$$

$$\xrightarrow{d} \frac{1}{f_Q(T_0(x))} \mathbb{G}_2(F_P(x)),$$

on $[a, b]$. Recalling the sample size condition $n/(n+m) \to \kappa$, we have the following convergence to a Gaussian process:

$$\sqrt{n+m}(\widetilde{T}_{n,m}(x) - T_0(x))$$

$$= \frac{\sqrt{n+m}}{n} \sum_{i=1}^{n} \psi(X_i; x) - \frac{\sqrt{n+m}}{m} \sum_{j=1}^{m} \zeta(Y_j; \cdot) + o_{P\otimes Q}(1)$$

$$\xrightarrow{d} \frac{1}{f_Q(T_0(x))} \left[\sqrt{1/\kappa}\mathbb{G}_1(x) - \sqrt{1/(1-\kappa)}\mathbb{G}_2(F_P(x)) \right].$$

We remark that the covariance function of $\mathbb{G}_1 \circ F_P$ and $\mathbb{G}_2 \circ F_P$ are can be written explicitly:

$$
\begin{aligned}
\mathrm{Cov}\left(\mathbb{G}_1(x), \mathbb{G}_1(x')\right) &= \mathrm{Cov}\left(\mathbb{G}_1(x), \mathbb{G}_2(x')\right) \\
&= \min\{F_P(x), F_P(x')\} - F_P(x)F_P(x').
\end{aligned}
$$

A.4. Bootstrap for pointwise confidence intervals

Recall the empirical distribution functions $\widetilde{F}_{P,n}(x) = n^{-1}\sum_{i=1}^{n} \mathbf{1}\{X_i \le x\}$ and $\widetilde{F}_{Q,m}(y) = m^{-1}\sum_{j=1}^{m} \mathbf{1}\{Y_j \le y\}$. We define a plug-in estimator as the optimal transport map from $\widetilde{P}_n$ to $\widetilde{Q}_m$:

Let $X_1^*, \ldots, X_n^* \sim \widetilde{P}_n$ and $Y_1^*, \ldots, Y_m^* \sim \widetilde{Q}_m$. Define the bootstrap distribution function $\widetilde{F}_{P,n}^*(x) := n^{-1}\sum_{i=1}^{n} \mathbf{1}\{X_i^* \le x\}$ and $\widetilde{F}_{Q,m}^*(y) := m^{-1}\sum_{j=1}^{m} \mathbf{1}\{Y_j^* \le y\}$. The bootstrap transport map is then given by $\widetilde{T}_{n,m}^*(x) := \widetilde{F}_{Q,m}^{*-1}(\widetilde{F}_{P,n}^*(x))$.

Theorem 7. *Let Φ be the distribution function of the standard normal distribution. If F_Q is continuously differentiable at $F_Q^{-1}(F_P(x))$ and the derivative $f_Q(F_Q^{-1}(F_P(x)))$ is nonzero, then*

$$
\sup_{x_0} \left| \widetilde{P}_{n,m}^*\left(\sqrt{n+m}(\widetilde{T}_{n,m}^*(x) - \widetilde{T}_{n,m}(x)) \le x_0\right) - \Phi\left(x_0/\sigma_T(x)\right) \right| \xrightarrow{P \otimes Q} 0, \quad (17)
$$

where $\sigma_T(x) = (f_Q(T_0(x))^{-1}\sqrt{(\kappa^{-1} + (1-\kappa)^{-1})F_P(x)(1 - F_P(x))}.$

Proof of Theorem 7. With $\sigma_P(x) := \sqrt{F_P(x)(1 - F_P(x))}$ and $\sigma_Q(x) := \sqrt{F_Q(x)(1 - F_Q(x))}$, we will show the weak conditional convergences of the bootstrap empirical distributions:

$$
\sup_{x_0} \left| \widetilde{P}_n\left(\sqrt{n}[\widetilde{F}_{P,n}^*(x) - \widetilde{F}_{P,n}(x)] \le x_0\right) - \Phi(x_0/\sigma_P(x)) \right| \xrightarrow{P} 0, \quad (18)
$$

and

$$
\sup_{x_0} \left| \widetilde{Q}_m\left(\sqrt{m}[\widetilde{F}_{Q,m}^*(x) - \widetilde{F}_{Q,m}(x)] \le x_0\right) - \Phi(x_0/\sigma_Q(x)) \right| \xrightarrow{Q} 0. \quad (19)
$$

Observe that the variance of $\widetilde{F}_{P,n}^*(x) - \widetilde{F}_{P,n}(x)$ is

$$
\sigma_{\widetilde{P}_n}^2 := \widetilde{F}_{P,n}(x)(1 - \widetilde{F}_{P,n}(x)) = \sigma_P^2 + o_P(1).
$$

Applying the Berry-Esseen on $\sqrt{n}[\widetilde{F}_{P,n}^*(x) - \widetilde{F}_{P,n}(x)]$ with respect to the empirical distribution $\widetilde{P}_n$, there is a constant $A > 0$ such that

$$
\sup_{x_0} \left| \widetilde{P}_n\left(\sqrt{n}[\widetilde{F}_{P,n}^*(x) - \widetilde{F}_{P,n}(x)] \le x_0\right) - \Phi(x_0/\sigma_{\widetilde{P}_n}^2(x)) \right|
$$

$$\leq \frac{A \sum_{i=1}^{n} |\widetilde{F}_{P,n}^*(x) - \widetilde{F}_{P,n}(x)|^3}{n^{3/2} \sigma_{\widetilde{P}_n}^2}$$

$$\leq \frac{A}{n^{1/2} \sigma_{\widetilde{P}_n}}$$

$$= \frac{A}{n^{1/2} \sigma_P} + o_P(1).$$

We now find an upper bound of $|\Phi(x_0/\sigma_{\widetilde{P}_n}(x)) - \Phi(x_0/\sigma_P(x))|$ via the method of calculus on the function $\sigma \mapsto \Phi(x_0/\sigma)$. By the boundedness of the Gaussian density, there is a constant B such that

$$\sup_{x_0} \left| \frac{\partial \Phi(x_0/\sigma)}{\partial \sigma} \right| \leq \frac{B}{\sigma}.$$

It follows from the mean value theorem that

$$|\Phi(x_0/\sigma_{\widetilde{P}_n}(x)) - \Phi(x_0/\sigma_P(x))| \leq \frac{B}{\min\{\sigma_{\widetilde{P}_n}(x), \sigma_P(x)\}} |\sigma_{\widetilde{P}_n}(x) - \sigma_P(x)| \xrightarrow{P} 0,$$

which allows us to conclude (18). The convergence (19) follows analogously.

For any cadlag function Λ, define a random function $Z_\Lambda(x) \sim \mathcal{N}(0, \Lambda(x)(1 - \Lambda(x)))$. By the weak law of large number,

$$\sqrt{n}[\widetilde{F}_{P,n}(x) - F_P(x)] \xrightarrow{d} Z_{F_P}(x), \tag{20}$$

and

$$\sqrt{m}[\widetilde{F}_{Q,m}(x) - F_Q(x)] \xrightarrow{d} Z_{F_Q}(x). \tag{21}$$

Consider the functional $\phi(u, v) = v^{-1} \circ u$. From Lemma 5, the Hadamard derivative of ϕ at (F_P, F_Q) is $\phi'(u, v) = (u - v \circ T)/f_Q \circ T$. We now apply the delta method for bootstrap Lemma 10 on ϕ: from (18), (19), (20), (21) and $n/(n + m) \to \kappa$, we have that $\sqrt{n + m}(\phi(\widetilde{F}_{P,n}^*, \widetilde{F}_{Q,m}^*) - \phi(\widetilde{F}_{P,n}, \widetilde{F}_{Q,m})) = \sqrt{n + m}(\widetilde{T}_{n,m}^* - \widetilde{T}_{n,m})$ converges in distribution to

$$\phi'(\sqrt{1/\kappa} Z_{F_P}, \sqrt{1/(1-\kappa)} Z_{F_Q}) = \frac{1}{f_Q \circ T_0} \left(\sqrt{1/\kappa} Z_{F_P} - \sqrt{1/(1-\kappa)} Z_{F_Q} \circ T_0 \right)$$

$$= \frac{1}{f_Q \circ T_0} \left(\sqrt{1/\kappa} Z_{F_P} - \sqrt{1/(1-\kappa)} Z_{F_Q \circ T_0} \right)$$

$$= \frac{1}{f_Q \circ T_0} \left(\sqrt{1/\kappa} - \sqrt{1/(1-\kappa)} \right) Z_{F_P},$$

conditional to $X_1, \ldots, X_n$ in probability. In other words, the convergence (17) holds true. $\square$

Corollary 8 (Asymptotic Validity of the Bootstrap Confidence Interval). *For any $\alpha \in (0, 1)$, let $\alpha' := \alpha/2$ and define*

$$\widetilde{q}_{n,m}(1-\alpha') := \text{the } (1 - \alpha')\text{-quantile of } \sqrt{n + m}\left(\widetilde{T}_{n,m}^*(x) - \widetilde{T}_{n,m}(x) \right) \text{ under } \widetilde{P}_{n,m}.$$

Then, for any $x \in \mathbb{R}$ such that F_Q is continuously differentiable at $F_Q^{-1}(F_P(x))$ and the derivative $f_Q(F_Q^{-1}(F_P(x)))$ is nonzero, the bootstrap confidence band for $T_0(x)$:

$$\mathcal{C}_{n,m}^{(\alpha)}(x) := \left[\widetilde{T}_{n,m}(x) \pm \frac{\widetilde{q}_{n,m}(1-\alpha')}{\sqrt{n+m}}\right],$$

is asymptotically consistent at level $1 - \alpha$, that is,

$$\mathbb{P}\left(T_0(x) \in \mathcal{C}_{n,m}^{(\alpha)}(x)\right) \to 1 - \alpha,$$

as $n, m \to \infty$ and $n/(n+m) \to \kappa$.

Proof of Corollary 8. For notational convenience, denote $\Phi_{T,x}(x_0) := \Phi(x_0/\sigma_T(x))$. Recall from (17) that the sequence $\widetilde{P}_{n,m}(\sqrt{n+m}(\widetilde{T}_{n,m}^*(x) - \widetilde{T}_{n,m}(x)) \le x_0)$ converges in $P \otimes Q$-probability to $\Phi(x_0/\sigma_T(x))$. By passing along any subsequence and a further subsequence, we can assume that the convergence is $P \otimes Q$-almost surely.

The almost-sure convergence of the distribution functions implies the almost-sure convergence of the corresponding quantile functions [vdV98, Lemma 21.2]. In particular, $\widetilde{q}_{n,m}(1-\alpha')$, which is the $1-\alpha'$-quantiles of $\sqrt{n+m}(\widetilde{T}_{n,m}^*(x) - \widetilde{T}_{n,m}(x))$, converges $P \otimes Q$-a.s. to $\Phi_{T,x}^{-1}(1-\alpha')$. It follows from Slutsky's theorem that $\sqrt{n+m}(\widetilde{T}_{n,m}(x) - T_{n,m}(x)) - \widetilde{q}_{n,m}(1-\alpha')$ converges in distribution to $Z_T(x) - \Phi_{T,x}^{-1}(1-\alpha')$, where $Z_T \sim \mathcal{N}(0, \sigma_T^2)$. Therefore, as $n, m \to \infty$ and $n/(n+m) \to \kappa$,

$$\mathbb{P}\left(T_0(x) \ge \widetilde{T}_{n,m}(x) - \widetilde{q}_{n,m}(1-\alpha')/\sqrt{n+m}\right)$$
$$= \mathbb{P}\left(\sqrt{n+m}(\widetilde{T}_{n,m}(x) - T_{n,m}(x)) \le \widetilde{q}_{n,m}(1-\alpha')\right)$$
$$\to \mathbb{P}\left(Z_T \le \Phi_{T,x}^{-1}(1-\alpha')\right)$$
$$= \Phi_{T,x}(\Phi_{T,x}^{-1}(1-\alpha'))$$
$$= 1 - \alpha'.$$

Similarly, by the symmetry of Z_T,

$$\mathbb{P}\left(T_0(x) \le \widetilde{T}_{n,m}(x) + \widetilde{q}_{n,m}(1-\alpha')/\sqrt{n+m}\right)$$
$$= \mathbb{P}\left(\sqrt{n+m}(\widetilde{T}_{n,m}(x) - T_{n,m}(x)) \ge -\widetilde{q}_{n,m}(1-\alpha')\right)$$
$$\to \mathbb{P}\left(Z_T \ge -\Phi_{T,x}^{-1}(1-\alpha')\right)$$
$$= \mathbb{P}\left(Z_T \le \Phi_{T,x}^{-1}(1-\alpha')\right)$$
$$= 1 - \alpha'.$$

We conclude that $\mathbb{P}\left(T_0(x) \in \mathcal{C}_{n,m}^{(\alpha)}(x)\right) \to 1 - 2\alpha' = 1 - \alpha.$ $\qquad\square$

TABLE 2

Evaluations of our $(1-\alpha)$-level pointwise confidence intervals of the optimal transport map from $N(0,1)$ to $\mathrm{Gamma}(5, 0.5)$ based on 1,000 pairs of samples from each distribution. The table displays the median of average widths and summary of the coverage probabilities of the pointwise confidence intervals over $[-2.5, 2.5]$.

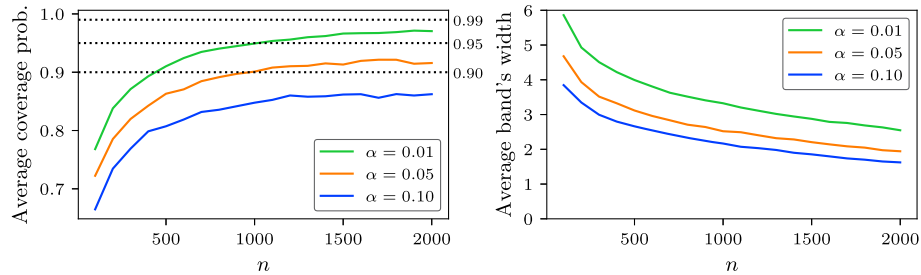| $1-\alpha$ | n | m | Average width | Pointwise coverage probabilities | | |
				Minimum	Maximum	Average
	200	50	3.35	0.008	0.865	0.73
0.90	700	175	2.44	0.362	0.881	0.83
	2000	500	1.62	0.774	0.889	**0.86**
	200	50	3.93	0.01	0.924	0.79
0.95	700	175	2.84	0.341	0.942	0.88
	2000	500	1.94	0.81	0.945	**0.92**
	200	50	4.93	0.003	0.979	0.84
0.99	700	175	3.63	0.365	0.986	0.93
	2000	500	2.55	0.847	0.99	**0.97**

FIG 4. *Plots of the average of the coverage probabilities (left) and the median of average widths (right) of the simulated pointwise confidence intervals over $[-2.5, 2.5]$ as functions of sample size n.*

A.5. Simulation

To validate our asymptotic intervals, we perform Monte Carlo simulation with the same design as in Section 5.1 for three sample sizes: $n = 200, 700$ and 2000. For each $x \in [-2.5, 2.5]$, the estimated coverage probability *at point x* is the proportion of 1000 runs in which our confidence interval at x contains $T_0(x)$. Table 2 displays the summary statistics for the pointwise coverage probabilities and the medians of average widths of the intervals across $x \in [-2.5, 2.5]$.

For small sample sizes ($n = 200$ and $n = 700$), the minimum coverage probabilities are much smaller than the nominal probabilities $(1 - \alpha)$, whereas the maximum coverage probabilities are close to the nominal probabilities. And as the sample size increases, both the minimum and maximum coverage probabilities converge to the nominal probabilities. Of course, the widths of the confidence intervals decreases as n increases and $1 - \alpha$ decreases.

In Figure 4, we plot the average coverage probability and the median of average width as a function of n. From the plots, we see that the average coverage probability converges to the nominal probability as n increases. However, the convergent is very slow for $1 - \alpha = 0.90$ and 0.95. Ss discussed in Section 5.2,
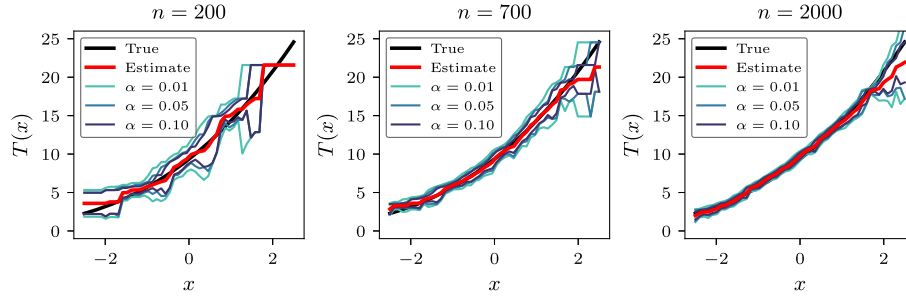
FIG 5. *Examples of* $(1 - \alpha)$-*level pointwise confidence intervals over* $[-2.5, 2.5]$, *for three different values of* α, *based on specific samples of size* $200, 700$ *and* 2000.

this is attributed to the insufficient number of sample points from the right tail of $N(0, 1)$, which hinders the estimation of the transport map between the two distributions.

The cause of the sub-optimal coverage probability becomes more apparent when examining the individual confidence intervals (Figure 5). We can see that for larger values of x, our plug-in estimator (the red curves) remains significantly distant from the true transport map (the black curves). Consequently, this discrepancy causes the confidence intervals to fail in capturing the transport map accurately.

Appendix B: Mathematical tools

B.1. Functional delta method

Consider stochastic processes $\mathbb{X}_n$ and $\mathbb{T}$ with values in a normed linear space $\mathcal{D}$ and a function $\phi : \mathcal{D} \to \mathcal{E}$ where $\mathcal{E}$ is another normed linear space. The functional delta method provides a way to turn the weak convergence of a sequence of stochastic processes $a_n(\mathbb{X}_n - \mathbb{T})$ into that of $a_n (\phi(\mathbb{X}_n) - \phi(\mathbb{T}))$. The idea is to realize that the latter can be written as $a_n \left(\phi(\mathbb{T} + a_n^{-1}h_n) - \phi(\mathbb{T})\right)$ where $h_n = a_n(\mathbb{X}_n - \mathbb{T})$. With sufficient differentiability condition on ϕ we expect this sequence to converge to $\phi'(\mathbb{T})$ as $a_n \to \infty$. To rigorously obtaining the convergence, we need the notion of *Hadamard differentiable* functions.

Definition 1 (Hadamard Derivative)**.** A function $\phi : \mathcal{D} \to \mathcal{E}$ is Hadamard differentiable at $\mathbb{T} \in \mathcal{D}$ if there exists a continuous linear map $\phi'_{\mathbb{T}} : \mathcal{D} \to \mathcal{E}$ such that for any $h_t, h \in \mathcal{D}$ satisfying $h_t \to h$ in $\mathcal{D}$ as $t \to 0$, we have

$$\left\| \frac{\phi(\mathbb{T} + th_t) - \phi(\mathbb{T})}{t} - \phi'_{\mathbb{T}}(h) \right\|_{\mathcal{E}} \to 0, \quad \text{as} \quad t \to 0.$$

In this case, we say that $\phi'_{\mathbb{T}}$ is the Hadamard derivative of ϕ at $\mathbb{T}$.

If ϕ is Hadamard differentiable then one can obtain the convergence of stochastic processes under ϕ—this is essentially the statement of the functional delta method.

Lemma 9 ([vdV98, Theorem 20.8])**.** *Let $\mathcal{D}$ and $\mathcal{E}$ be normed linear spaces. Let $\mathbb{X}_n, \mathbb{X}$ and $\mathbb{T}$ be stochastic processes with values in $\mathcal{D}$ such that $a_n(\mathbb{X}_n - \mathbb{T}) \xrightarrow{d} \mathbb{X}$. Let $\phi : \mathcal{D} \to \mathcal{E}$ be Hadamard differentiable at $\mathbb{T}$. Then,*

$$a_n\left(\phi(\mathbb{X}_n) - \phi(\mathbb{T})\right) \xrightarrow{d} \phi'_\mathbb{T}(\mathbb{X}). \tag{22}$$

If, in addition, $\phi'_\mathbb{T}$ is continuous on $\mathcal{D}$. Then,

$$a_n\left(\phi(\mathbb{X}_n) - \phi(\mathbb{T})\right) = \phi'_\mathbb{T}\left(a_n(\mathbb{X}_n - \mathbb{T})\right) + o_\mathbb{P}(1).$$

B.2. Functional delta method for bootstrap

In this work, we use the bootstrap procedure to estimate the distribution of stochastic processes of type $a_n\left(\phi(\mathbb{X}_n) - \phi(\mathbb{T})\right)$. Let $\widetilde{P}_n$ be the empirical distribution based on the sample $X_1, \dots, X_n$ and $\mathbb{X}_n^*$ be the bootstrap estimator based on a bootstrap sample $X_1^*, \dots, X_n^* \sim \widetilde{P}_n$. In view of (22), the asymptotic consistency of the bootstrap distribution requires that:

$$a_n\left(\phi(\mathbb{X}_n) - \phi(\mathbb{T})\right) \xrightarrow{d} \phi'_\mathbb{T}(\mathbb{X}) \quad \text{and} \quad a_n\left(\phi(\mathbb{X}_n^*) - \phi(\mathbb{X}_n)\right) \xrightarrow{\mathbb{P}|\widetilde{P}_n} \phi'_\mathbb{T}(\mathbb{X}), \tag{23}$$

where $\xrightarrow{\mathbb{P}|\widetilde{P}_n}$ denotes the convergence in probability, conditionally given $X_1, \dots,$ X_n in distribution, which can be formally written in terms of the bounded Lipschitz metric:

$$\sup_{h \in BL_1(\mathcal{E})} \left|\mathbb{E}_n\left[h\left(a_n(\phi(\mathbb{X}_n^*) - \phi(\mathbb{X}_n))\right)\right] - \mathbb{E}\left[h\left(\phi'(\mathbb{T})\right)\right]\right| \xrightarrow{\mathbb{P}} 0,$$

where $\mathbb{E}_n$ is the expectation with respect to $\widetilde{P}_n$ and $BL_1(\mathcal{E})$ is the space of Lipschitz functions $h : \mathcal{E} \to [-1, 1]$ with Lipschitz constants bounded by one.

The first convergence in (23) can be obtained via the functional delta method, and the second convergence can be obtained using the following lemma:

Lemma 10 ([vdV98, Theorem 23.9])**.** *Let $\mathcal{D}$ and $\mathcal{E}$ be normed linear spaces. Let $\mathbb{X}_n, \mathbb{X}$ and $\mathbb{T}$ be stochastic processes with values in $\mathcal{D}$ such that $a_n(\mathbb{X}_n - \mathbb{T}) \xrightarrow{d} \mathbb{X}$ and $a_n(\mathbb{X}_n^* - \mathbb{X}_n) \xrightarrow{\mathbb{P}|\widetilde{P}_n} \mathbb{X}$. Let $\phi : \mathcal{D} \to \mathcal{E}$ be Hadamard differentiable at $\mathbb{T}$. If $\mathbb{X}$ is tight, then*

$$a_n\left(\phi(\mathbb{X}_n^*) - \phi(\mathbb{X}_n)\right) \xrightarrow{\mathbb{P}|\widetilde{P}_n} \phi'_\mathbb{T}(\mathbb{X}).$$

Appendix C: Full proof of Theorem 3

Proof of Theorem 3. We divide the proof into seven steps.

Step 1. Find the stochastic limits of $\sqrt{n+m}(\widehat{F}_P^*(x) - \widehat{F}_P(x))$ **and** $\sqrt{n+m}(\widehat{F}_Q^*(x) - \widehat{F}_Q(x))$ **in the uniform norm** We use the result from [KMT76, Theorem 3] which states that, for any iid sequence $U_1, \ldots \sim P_U$ where P_U is the uniform random distribution on $[0,1]$ and $A_n(t) = n^{-1} \sum_{i=1}^n \mathbf{1}\{U_i \leq t\}$, there is a Brownian bridge $\mathbb{G}_1$ such that

$$P_U\left(\sup_{0 \leq t \leq 1} \left|\sqrt{n}(A_n(t) - t) - \mathbb{G}_1(t)\right| > \delta\right) \to 0, \tag{24}$$

as $n \to \infty$, for any $\delta > 0$.

Conditional on $X_1, \ldots, X_n$, $\widehat{F}_P$ is a non-random continuous function. So we can apply the above result with $U_i = \widehat{F}_P(X_i^*)$ and $t = \widehat{F}_P(x)$, yielding $A_n(t) = n^{-1} \sum_{i=1}^n \mathbf{1}\{U_i \leq \widehat{F}_P(x)\} = n^{-1} \sum_{i=1}^n \mathbf{1}\{\widehat{F}_P(X_i^*) \leq \widehat{F}_P(x)\} = n^{-1} \sum_{i=1}^n \mathbf{1}\{X_i^* \leq x\} = \widehat{F}_P^*(x)$. The fact that (24) holds for all sequences $X_1, \ldots, X_n$ allows us to obtain the P-almost sure convergence:

$$\widehat{P}_n\left(\sup_x \left|\sqrt{n}(\widehat{F}_P^*(x) - \widehat{F}_P(x)) - \mathbb{G}_1(\widehat{F}_P(x))\right| > \delta\right) \xrightarrow{P\text{-a.s.}} 0.$$

From [HHZ08, Lemma A.1], we have $\sup_x|\widehat{F}_P(x) - F_P(x)| \xrightarrow{P\text{-a.s.}} 0$. By the uniform continuity of the Brownian bridge, we have $\mathbb{P}\big(\sup_x|\mathbb{G}_1(\widehat{F}_P(x)) - \mathbb{G}_1(F_P(x))| > \delta\big) \xrightarrow{P\text{-a.s.}} 0$. Consequently, we have

$$\widehat{P}_n\left(\sup_x \left|\sqrt{n}(\widehat{F}_P^*(x) - \widehat{F}_P(x)) - \mathbb{G}_1(F_P(x))\right| > \delta\right) \xrightarrow{P\text{-a.s.}} 0.$$

From the condition $n/(n+m) \to \kappa$ as $n, m \to \infty$, we apply Slutsky's lemma and obtain $(\sqrt{1/\kappa} - \sqrt{(n+m)/n}) \sup_x \mathbb{G}_1(F_P(x)) \to 0$ in probability independent of $X_1, \ldots, X_n$. Therefore, we have

$$\widehat{P}_n\left(\sup_x \left|\sqrt{n+m}(\widehat{F}_P^*(x) - \widehat{F}_P(x)) - \sqrt{1/\kappa}\mathbb{G}_1(F_P(x))\right| > \delta\right) \xrightarrow{P\text{-a.s.}} 0. \tag{25}$$

Using a similar argument, we obtain

$$\widehat{Q}_m\left(\sup_x \left|\sqrt{n+m}(\widehat{F}_Q^*(x) - \widehat{F}_Q(x)) - \sqrt{1/(1-\kappa)}\mathbb{G}_2(F_Q(x))\right| > \delta\right) \xrightarrow{Q\text{-a.s.}} 0,$$

where $\mathbb{G}_2$ is another independent Browninan bridge.

Step 2. Convert the convergences in the uniform norm to those in the bounded Lipschitz metric Recall that $BL_1(\ell^\infty(\mathcal{F}))$ denotes the space of Lipschitz functions $h : \ell^\infty(\mathcal{F}) \to [-1,1]$ with Lipschitz constants bounded by one. For any $h \in BL_1(\ell^\infty(\mathcal{F}))$, we have $|h(\sqrt{n+m}(\widehat{F}_P^* - \widehat{F}_P)) - h(\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P)| \leq \sup_x|\sqrt{n+m}(\widehat{F}_P^*(x) - \widehat{F}_P(x)) - \sqrt{1/\kappa}\mathbb{G}_1(F_P(x))|$. Consequently, it holds that

$$\widehat{P}_n\left(\sup_x \left|\sqrt{n+m}(\widehat{F}_P^*(x) - \widehat{F}_P(x)) - \sqrt{1/\kappa}\mathbb{G}_1(F_P(x))\right| > \delta\right)$$

$$\geq \widehat{P}_n \left(|h(\sqrt{n+m}(\widehat{F}_P^* - \widehat{F}_P)) - h(\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P)| > \delta \right).$$

Define an event $S_{\delta,n,m} = \{h(\sqrt{n+m}(\widehat{F}_P^* - \widehat{F}_P)) - h(\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P) > \delta\}$. Denoting by $\mathbb{E}_n$ the expectation with respect to $\widehat{P}_n$, Jensen's inequality yields that

$$\left| \mathbb{E}_n h(\sqrt{n+m}(\widehat{F}_P^* - \widehat{F}_P)) - \mathbb{E}h(\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P) \right|$$
$$\leq \mathbb{E}_n \mathbb{E} \left| h(\sqrt{n+m}(\widehat{F}_P^* - \widehat{F}_P)) - h(\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P) \right|$$
$$\leq 2\widehat{P}_n(S_{\delta,n,m}) + \delta\widehat{P}_n(S_{\delta,n,m}^c)$$
$$\leq 2\widehat{P}_n(S_{\delta,n,m}) + \delta,$$

where we used the fact that h is bounded by one in the second to last inequality. From (25), we have $\widehat{P}_n(S_{\delta,n,m}) \to 0$ for any arbitrary δ; therefore, by taking supremum over all $h \in BL_1(\ell^\infty(\mathcal{F}))$, there is a Brownian bridge $\mathbb{G}_1$ such that

$$\sup_{h \in BL_1(\ell^\infty(\mathcal{F}))} \left| \mathbb{E}_n h(\sqrt{n+m}(\widehat{F}_P^* - \widehat{F}_P)) - \mathbb{E}h(\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P) \right| \xrightarrow{P\text{-a.s.}} 0. \quad (26)$$

Similarly, letting $\mathbb{E}_m$ be the expectation with respect to $\widehat{Q}_m$, there is a Brownian bridge $\mathbb{G}_2$ such that

$$\sup_{h \in BL_1(\ell^\infty(\mathcal{F}))} \left| \mathbb{E}_m h(\sqrt{n+m}(\widehat{F}_Q^* - \widehat{F}_Q)) - \mathbb{E}h(\sqrt{1/(1-\kappa)}\mathbb{G}_2 \circ F_Q) \right| \xrightarrow{Q\text{-a.s.}} 0.$$
$$(27)$$

Step 3. Prove the convergence of the sequence $\sqrt{n+m}(\widehat{F}_P^* - \widehat{F}_P, \widehat{F}_Q^* - \widehat{F}_Q)$ in the bounded Lipschitz metric Now we will show that

$$\sup_{h \in BL_1(\ell^\infty(\mathcal{F})^2)} \left| \mathbb{E}_{n,m} h\left(\sqrt{n+m}(\widehat{F}_P^* - \widehat{F}_P, \widehat{F}_Q^* - \widehat{F}_Q)\right) \right.$$
$$\left. - \mathbb{E}h\left(\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P, \sqrt{1/(1-\kappa)}\mathbb{G}_2 \circ F_Q\right) \right| \xrightarrow{P \otimes Q\text{-a.s.}} 0. \quad (28)$$

Consider any arbitrary $h \in BL_1(\ell^\infty(\mathcal{F})^2)$. For any $w \in \ell^\infty$, define $h_1^w, h_2^w \in BL_1(\ell^\infty)$ by $h_1^w(\cdot) = h(\cdot, w)$ and $h_2^w(\cdot) = h(w, \cdot)$. Denoting $E_{n,m}$ the expectation with respect to the empirical distribution $\widehat{\mathbb{P}}_{n,m}$, we have

$$\left| \mathbb{E}_{n,m} h\left(\sqrt{n+m}(\widehat{F}_P^* - \widehat{F}_P, \widehat{F}_Q^* - \widehat{F}_Q)\right) \right.$$
$$\left. - \mathbb{E}h\left(\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P, \sqrt{1/(1-\kappa)}\mathbb{G}_2 \circ F_Q\right) \right|$$
$$\leq \left| \mathbb{E}_{n,m} h\left(\sqrt{n+m}(\widehat{F}_P^* - \widehat{F}_P, \widehat{F}_Q^* - \widehat{F}_Q)\right) \right.$$
$$\left. - \mathbb{E}_{n,m} \mathbb{E}h\left(\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P, \sqrt{n+m}(\widehat{F}_Q^* - \widehat{F}_Q)\right) \right|$$
$$+ \left| \mathbb{E}_{n,m} \mathbb{E}h\left(\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P, \sqrt{n+m}(\widehat{F}_Q^* - \widehat{F}_Q)\right) \right.$$

$$- \mathbb{E}h\big(\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P, \sqrt{1/(1-\kappa)}\mathbb{G}_2 \circ F_Q\big)\Big|$$

$$\leq \sup_{w \in \ell^\infty(\mathcal{F})} \Big|\mathbb{E}_n h_1^w\big(\sqrt{n+m}(\widehat{F}_P^* - \widehat{F}_P)\big) - \mathbb{E}h_1^w(\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P)\Big|$$

$$+ \sup_{w \in \ell^\infty(\mathcal{F})} \Big|\mathbb{E}_m h_2^w\big(\sqrt{n+m}(\widehat{F}_Q^* - \widehat{F}_Q)\big) - \mathbb{E}h_2^w(\sqrt{1/(1-\kappa)}\mathbb{G}_2 \circ F_Q)\Big|$$

$$\leq \sup_{h_1 \in BL_1(\ell^\infty(\mathcal{F}))} \Big|\mathbb{E}_n h_1\big(\sqrt{n+m}(\widehat{F}_P^* - \widehat{F}_P)\big) - \mathbb{E}h_1(\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P)\Big|$$

$$+ \sup_{h_2 \in BL_1(\ell^\infty(\mathcal{F}))} \Big|\mathbb{E}_m h_2\big(\sqrt{n+m}(\widehat{F}_Q^* - \widehat{F}_Q)\big) - \mathbb{E}h_2(\sqrt{1/(1-\kappa)}\mathbb{G}_2 \circ F_Q)\Big|.$$

Using (26) and (27), the last expression converges $P \otimes Q$-a.s. to zero uniformly over $h \in BL_1(\ell^\infty(\mathcal{F})^2)$ as $n, m \to \infty$. Thus taking the supremum over such h yields (28) as desired.

In addition, under the required conditions on f_P, f_Q and K, we can use Corollary 2 of [GN07] to obtain $\sqrt{n}(\widehat{F}_P - F_P) \xrightarrow{d} \mathbb{G}_1 \circ F_P$ and $\sqrt{m}(\widehat{F}_Q - F_Q) \xrightarrow{d} \mathbb{G}_2 \circ F_Q$. Consequently, we obtain

$$\sqrt{n+m}(\widehat{F}_P - F_P, \widehat{F}_Q - F_Q) \xrightarrow{d} (\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P, \sqrt{1/(1-\kappa)}\mathbb{G}_2 \circ F_Q). \quad (29)$$

Step 4. Apply the delta method for bootstrap With (28) and (29), we can now apply the delta method for bootstrap (see Appendix B.2 for a brief exposition). Define a functional $\Psi : D[a,b] \times D[a,b] \to D[a,b]$

$$\Psi(u,v) = (v^{-1} \circ u(\cdot))/s_\kappa(\cdot).$$

We recall from (11) that the Hadamard derivative of Ψ at (F_P, F_Q) is

$$\Psi'(u,v) = (u(\cdot) - v \circ T_0(\cdot))/((f_Q \circ T)(\cdot)s_\kappa(\cdot)).$$

Using Lemma 10 on Ψ, the sequence

$$\sqrt{n+m}(\Psi(\widehat{F}_{P,n}^*, \widehat{F}_{Q,m}^*) - \Psi(\widehat{F}_P, \widehat{F}_Q)) = \sqrt{n+m}(\widehat{T}_{n,m}^* - \widehat{T}_{n,m})(\cdot)/s_\kappa(\cdot)$$

weakly converges to

$$\begin{aligned}
\Psi'&(\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P, \sqrt{1/(1-\kappa)}\mathbb{G}_2 \circ F_Q) \\
&= \Big(\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P - \sqrt{1/(1-\kappa)}\mathbb{G}_2 \circ F_Q \circ T_0\Big)/((f_Q \circ T_0)s_\kappa) \\
&= \Big(\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P - \sqrt{1/(1-\kappa)}\mathbb{G}_2 \circ F_P\Big)/((f_Q \circ T_0)s_\kappa) \\
&= \mathsf{Z}_\kappa,
\end{aligned}$$

under the bounded Lipschitz metric. More precisely, denoting $\widehat{\mathsf{Z}}_{n,m} := \sqrt{n+m}(\widehat{T}_{n,m}^* - \widehat{T}_{n,m})/s_\kappa$, we have

$$\sup_{h \in BL_1(\ell^\infty(\mathcal{F}))} \Big|\mathbb{E}_{n,m} h(\widehat{\mathsf{Z}}_{n,m}) - \mathbb{E}h(\mathsf{Z}_\kappa)\Big| \xrightarrow{P \otimes Q} 0, \quad (30)$$

Step 5. Prove that the sequence $\widehat{\mathbb{P}}_{n,m}(\sup_x |\widehat{\mathsf{Z}}_{n,m}(x)| \leq x_0)$ **converges to** $\mathbb{P}(\sup_x|\mathsf{Z}_\kappa(x)| \leq x_0)$ **in** $P \otimes Q$**-probability** From (30), we consider the supremum over the set of functions h that are of the form $h(f) = \widetilde{h}(\sup_x |f(x)|)$ for some $\widetilde{h} \in BL_1[0,\infty)$, which results in

$$\sup_{\widetilde{h} \in BL_1[0,\infty)} \left| \mathbb{E}_{n,m}\big[\widetilde{h}\big(\sup_x |\widehat{\mathsf{Z}}_{n,m}(x)|\big)\big] - \mathbb{E}\big[\widetilde{h}\big(\sup_x|\mathsf{Z}_\kappa(x)|\big)\big] \right| \xrightarrow{P\otimes Q} 0.$$

By scaling, we can extend the supremum to $BL_M[0,\infty)$, the set of bounded Lipschitz functions with the Lipschitz constants bounded by $M > 0$. Given any ε and $\delta > 0$, there exists n, m large enough so that

$$P\Big(\sup_{\widetilde{h} \in BL_M[0,\infty)} \left| \mathbb{E}_{n,m}\big[\widetilde{h}\big(\sup_x |\widehat{\mathsf{Z}}_{n,m}(x)|\big)\big] - \mathbb{E}\big[\widetilde{h}\big(\sup_x|\mathsf{Z}_\kappa(x)|\big)\big] \right| > \delta/2 \Big) \leq \varepsilon.$$

Fix $x_0 > 0$. For a sufficiently large M, we can find two functions $\widetilde{h}_L, \widetilde{h}_U \in BL_M[0,\infty)$ such that $\widetilde{h}_L(x) \leq \mathbf{1}\{0 \leq x \leq x_0\} \leq \widetilde{h}_U(x)$ and $\mathbb{E}[\widetilde{h}_U(\sup_x|\mathsf{Z}_\kappa(x)|) - \widetilde{h}_L(\sup_x|\mathsf{Z}_\kappa(x)|)] < \delta/2$. It follows that, with a probability greater than $1 - \varepsilon$,

$$\widehat{\mathbb{P}}_{n,m}\Big(\sup_x |\widehat{\mathsf{Z}}_{n,m}(x)| \leq x_0 \Big) - \mathbb{P}\Big(\sup_x|\mathsf{Z}_\kappa(x)| \leq x_0 \Big)$$

$$\leq \mathbb{E}_{n,m}\big[\widetilde{h}_U\big(\sup_x |\widehat{\mathsf{Z}}_{n,m}(x)|\big)\big] - \mathbb{E}\big[\widetilde{h}_L\big(\sup_x|\mathsf{Z}_\kappa(x)|\big)\big]$$

$$\leq \mathbb{E}\big[\widetilde{h}_U\big(\sup_x|\mathsf{Z}_\kappa(x)|\big)\big] - \mathbb{E}\big[\widetilde{h}_L\big(\sup_x|\mathsf{Z}_\kappa(x)|\big)\big] + \delta/2$$

$$\leq \delta/2 + \delta/2 = \delta.$$

Similarly, we have

$$\widehat{\mathbb{P}}_{n,m}\Big(\sup_x |\widehat{\mathsf{Z}}_{n,m}(x)| \leq x_0 \Big) - \mathbb{P}\Big(\sup_x|\mathsf{Z}_\kappa(x)| \leq x_0 \Big) \geq -\delta.$$

Therefore, we have the convergence in $P \otimes Q$-probability for each $x_0 > 0$:

$$\widehat{\mathbb{P}}_{n,m}\Big(\sup_x |\widehat{\mathsf{Z}}_{n,m}(x)| \leq x_0 \Big) \xrightarrow{P\otimes Q} \mathbb{P}\Big(\sup_x|\mathsf{Z}_\kappa(x)| \leq x_0 \Big).$$

Step 6. Use Lemma 2 to approximate s_κ **by** $\widehat{s}_{n,m}$ **and finish the proof of** (9) Let $\widehat{\mathsf{Z}}^*_{n,m} := \sqrt{n+m}(\widehat{T}^*_{n,m} - \widehat{T}_{n,m})/\widehat{s}_{n,m}$. So we have to show that

$$\widehat{\mathbb{P}}_{n,m}\Big(\sup_x |\widehat{\mathsf{Z}}^*_{n,m}(x)| \leq x_0 \Big) \xrightarrow{P\otimes Q} \mathbb{P}\Big(\sup_x|\mathsf{Z}_\kappa(x)| \leq x_0 \Big). \qquad (31)$$

First, observe that

$$\widehat{\mathbb{P}}_{n,m}\Big(\sup_x |\widehat{\mathsf{Z}}^*_{n,m}(x)| \leq x_0 \Big) = \widehat{\mathbb{P}}_{n,m}\left(\sup_x |\widehat{\mathsf{Z}}_{n,m}(x)| \cdot \left| \frac{s_\kappa(x)}{\widehat{s}_{n,m}(x)} \right| \leq x_0 \right).$$

This is where we use Lemma 2. For arbitrarily small $\varepsilon, \delta, \delta' > 0$, there are sufficiently large n, m such that, with $P \otimes Q$-probability greater than $1 - \varepsilon$, the following inequalities hold simulteneously: first,

$$\widehat{\mathbb{P}}_{n,m}\Big(\sup_x \big|\widehat{\mathsf{Z}}_{n,m}(x)\big| \leq x_0(1-\delta)\Big) \leq \widehat{\mathbb{P}}_{n,m}\Big(\sup_x \big|\widehat{\mathsf{Z}}^*_{n,m}(x)\big| \leq x_0\Big)$$
$$\leq \widehat{\mathbb{P}}_{n,m}\Big(\sup_x \big|\widehat{\mathsf{Z}}_{n,m}(x)\big| \leq x_0(1+\delta)\Big),$$

which is a result of (6). Second, we have

$$\left|\widehat{\mathbb{P}}_{n,m}\Big(\sup_x \big|\widehat{\mathsf{Z}}_{n,m}(x)\big| \leq x_0(1-\delta')\Big) - \mathbb{P}\Big(\sup_x |\mathsf{Z}_\kappa(x)| \leq x_0(1-\delta')\Big)\right| < \delta/2,$$

and lastly,

$$\left|\widehat{\mathbb{P}}_{n,m}\Big(\sup_x \big|\widehat{\mathsf{Z}}_{n,m}(x)\big| \leq x_0(1+\delta')\Big) - \mathbb{P}\Big(\sup_x |\mathsf{Z}_\kappa(x)| \leq x_0(1+\delta')\Big)\right| < \delta/2.$$

By the continuity of the distribution function of $\sup_x |\mathsf{Z}_\kappa(x)|$ (Lemma 11), we can choose δ' sufficiently small that $|\mathbb{P}(\sup_x |\mathsf{Z}_\kappa(x)| \leq x_0(1-\delta')) - \mathbb{P}(\sup_x |\mathsf{Z}_\kappa(x)| \leq x_0(1-\delta'))| < \delta/2$. Combining the above inequalities yields the following bound with $P \otimes Q$-probability greater than $1 - \varepsilon$,

$$\left|\widehat{\mathbb{P}}_{n,m}\Big(\sup_x \big|\widehat{\mathsf{Z}}^*_{n,m}(x)\big| \leq x_0\Big) - \mathbb{P}\Big(\sup_x |\mathsf{Z}_\kappa(x)| \leq x_0\Big)\right| < \delta,$$

which implies (31).

Step 7. Prove (8) **using the functional delta method** Starting with (29), it follows from the functional delta method (Lemma 9) that

$$\sqrt{n+m}\left(\frac{\widehat{T}_{n,m} - T_0}{s_\kappa}\right) = \sqrt{n+m}(\psi(\widehat{F}_P, \widehat{F}_Q) - \psi(F_P, F_Q))$$
$$\xrightarrow{d} \psi'_{F,G}\left(\sqrt{1/\kappa}\mathbb{G}_1 \circ F_P, \sqrt{1/(1-\kappa)}\mathbb{G}_2 \circ F_Q\right)$$
$$= \mathsf{Z}_\kappa.$$

From (6), we have $s_\kappa/\widehat{s}_{n,m} \xrightarrow{P \otimes Q} 1$ under the uniform norm, so it follows from Slutsky's theorem that

$$\sqrt{n+m}\left(\frac{\widehat{T}_{n,m} - T_0}{\widehat{s}_{n,m}}\right) \xrightarrow{d} \mathsf{Z}_\kappa.$$

The continuous mapping theorem yields

$$\sqrt{n+m}\sup_x \left|\frac{\widehat{T}_{n,m}(x) - T_0(x)}{\widehat{s}_{n,m}}\right| \xrightarrow{d} \sup_x |\mathsf{Z}_\kappa(x)|,$$

as desired. $\qquad\square$

Appendix D: Additional results and proofs

Proof of Proposition 5. The proof is analogous to that of Proposition 6. Assumption 1 and 2 guarantee that F_Q satisfies the conditions in Proposition 6. In addition, under Assumption 1, we have the following weak convergences under the uniform norm which is due to [GN07, Corollary 2]:

$$\sqrt{n}(\widehat{F}_P - F_P) \xrightarrow{d} \mathbb{G}_1 \circ F_P \quad \text{and} \quad \sqrt{m}(\widehat{F}_Q - F_Q) \xrightarrow{d} \mathbb{G}_2 \circ F_Q,$$

for some Brownian bridges $\mathbb{G}_1$ and $\mathbb{G}_2$. This allows us to apply the functional delta method. The rest of the proof follows as in the proof of Proposition 6. $\square$

Proof of Lemma 2. Denote

$$\widehat{A}_{n,m} = \sqrt{(n+m)(n^{-1} + m^{-1})\widehat{F}_P(1 - \widehat{F}_P)}$$

and

$$A_\kappa = \sqrt{(\kappa^{-1} + (1-\kappa)^{-1})F_P(1 - F_P)}$$

for notational convenience. Since f_Q is uniformly bounded above and A_κ is uniformly bounded away from zero on $[a, b]$, the function s is uniformly bounded away from zero on $[a, b]$, so it suffices to show that $\sup_{x \in [a,b]} |\widehat{s}_{n,m}(x) - s_\kappa(x)| \xrightarrow{P \otimes Q} 0$. Now we split the difference into two terms:

$$\widehat{s}_{n,m}(x) - s_\kappa(x) = \left[\frac{\widehat{A}_{n,m}(x)}{\widehat{f}_Q(\widehat{T}_{n,m}(x))} - \frac{\widehat{A}_{n,m}(x)}{f_Q(T_0(x))} \right] + \left[\frac{\widehat{A}_{n,m}(x) - A_\kappa(x)}{f_Q(T_0(x))} \right]. \quad (32)$$

For the first term in (32), we write

$$\frac{\widehat{A}_{n,m}(x)}{\widehat{f}_Q(\widehat{T}_{n,m}(x))} - \frac{\widehat{A}_{n,m}(x)}{f_Q(T_0(x))} = \frac{\widehat{A}_{n,m}(x) \left[f_Q(T_0(x)) - \widehat{f}_Q(\widehat{T}_{n,m}(x)) \right]}{f_Q(T_0(x))\widehat{f}_Q(\widehat{T}_{n,m}(x))}.$$

The conditions on K, r_n and the uniform continuity of f_Q on $[a, b]$ imply $\sup_{x \in \mathbb{R}} |\widehat{f}_Q(x) - f_Q(x)| \xrightarrow{Q\text{-a.s.}} 0$ (see [BR78] and [Sil86, p. 72]). Consequently, $\widehat{f}_Q \circ \widehat{T}_{n,m}$ is uniformly bounded away from zero $P \otimes Q$-a.s. Combining this with a simple observation that $\widehat{A}_{n,m}$ is uniformly bounded above, it suffices to show that

$$\sup_{x \in [a,b]} \left| \widehat{f}_Q(\widehat{T}_{n,m}(x)) - f_Q(T_0(x)) \right| \xrightarrow{P \otimes Q\text{-a.s.}} 0.$$

We split $\widehat{f}_Q \circ \widehat{T}_{n,m} - f_Q \circ T_0$ as follows:

$$\widehat{f}_Q \circ \widehat{T}_{n,m} - f_Q \circ T_0 = \{\widehat{f}_Q \circ \widehat{T}_{n,m} - f_Q \circ \widehat{T}_{n,m}\} + \{f_Q \circ \widehat{T}_{n,m} - f_Q \circ T_0\}.$$

For the second term on the right, we use the Bahadur representation (Proposition 6) on $[a, b]$, which allows us to obtain the convergence in probability

$\widehat{T}_{n,m} \xrightarrow{P \otimes Q} T_0$ of functions in $C[a,b]$ under the uniform norm. Since the mapping $T_0 \mapsto g \circ T_0$ is continuous in $C[a,b]$, it follows from the continuous mapping theorem that $f_Q \circ \widehat{T}_{n,m} \xrightarrow{P \otimes Q} f_Q \circ T_0$. In other words,

$$\sup_{x \in [a,b]} |f_Q(\widehat{T}_{n,m}(x)) - f_Q(T_0(x))| \xrightarrow{P \otimes Q} 0.$$

For the first term, we use $\sup_{x \in \mathbb{R}} |\widehat{f}_Q(x) - f_Q(x)| \xrightarrow{Q\text{-a.s.}} 0$ which implies the convergence $\sup_{x \in [a,b]} |\widehat{f}_Q(\widehat{T}_{n,m}(x)) - f_Q(\widehat{T}_{n,m}(x))| \xrightarrow{P \otimes Q\text{-a.s.}} 0$.

Now we consider the second term in (32). From [HHZ08, Lemma A.1], we have $\sup_x |\widehat{F}_P(x) - F_P(x)| \xrightarrow{P\text{-a.s.}} 0$. From this, we split the numerator as follows:

$$\begin{aligned}
\widehat{A}_{n,m} - A_\kappa = \frac{1}{\widehat{A}_{n,m} + A_\kappa} &\Big[(\kappa^{-1} + (1-\kappa)^{-1})(\widehat{F}_P - F_P) \\
&+ (\kappa^{-1} + (1-\kappa)^{-1})(F_P - \widehat{F}_P)(F_P + \widehat{F}_P) \\
&+ \big\{ (n+m)(n^{-1} + m^{-1}) - (\kappa^{-1} + (1-\kappa)^{-1}) \big\} \widehat{F}_P(1 - \widehat{F}_P) \Big].
\end{aligned}$$

On the closed interval $[a,b]$, the sequence $\widehat{A}_{n,m} + A_\kappa$ is uniformly bounded away from zero, and $F_P + \widehat{F}_P$ and $\widehat{F}_P(1 - \widehat{F}_P)$ are uniformly bounded above. Combining these with the fact that $f_Q \circ T$ is also uniformly bounded away from zero on $[a,b]$, we obtain

$$\sup_{x \in [a,b]} \left| \frac{\widehat{A}_{n,m}(x) - A_\kappa(x)}{f_Q(T_0(x))} \right| \xrightarrow{P\text{-a.s.}} 0. \qquad \square$$

We also need a result on the continuity of the distribution function of the suprema of Gaussian processes.

Lemma 11. *Let* $\mathsf{Z} = \{\mathsf{Z}(t) : t \in I\}$ *be a tight Gaussian process with* $\mathbb{E}\mathsf{Z}(t) = 0$ *and* $\mathbb{E}\mathsf{Z}^2(t) = 1$ *for all* $t \in I$. *Then the distribution function* $x_0 \mapsto \mathbb{P}(\sup_t |\mathsf{Z}(t)| \leq x_0)$ *is continuous.*

Proof of Lemma 11. We use an anti-concentration inequality for the suprema of Gaussian processes [CCK14, Lemma A.1]:

$$\sup_{x_0 \in \mathbb{R}} \mathbb{P}\left(\left| \sup_{t \in I} |\mathsf{Z}(t)| - x_0 \right| \leq \varepsilon \right) \leq 4\varepsilon \left(1 + \mathbb{E}\left[\sup_{t \in I} |\mathsf{Z}(t)| \right] \right),$$

which implies the continuity of the distribution function at any $x_0 \in \mathbb{R}$. $\qquad \square$

Proof of Corollary 4. For notational convenience, denote $\|\mathsf{Z}_\kappa\|_{[a,b]} = \sup_{x \in [a,b]} |\mathsf{Z}_\kappa(x)|$. Recall from (9) that the sequence $\widehat{\mathbb{P}}_{n,m}(\sqrt{n+m} \sup_x |\widehat{T}_{n,m}^*(x) - \widehat{T}_{n,m}(x)|/\widehat{s}_{n,m}(x) \leq x_0)$ converges in $P \otimes Q$-probability to $\mathbb{P}(\|\mathsf{Z}_\kappa\|_{[a,b]} \leq x_0)$. By passing along any subsequence and a further subsequence, we can assume that the convergence is $P \otimes Q$-almost surely.

The almost-sure convergence of the distribution functions implies the almost-sure convergence of the corresponding quantile functions [vdV98, Lemma 21.2]. Thus, the sequence of quantile functions $\widehat{q}_{n,m}$ of $\sqrt{n+m}\sup_x|\widehat{T}_{n,m}^*(x)-\widehat{T}_{n,m}(x)|/\widehat{s}_{n,m}(x)$ converges $P\otimes Q$-a.s. to the quantile function q_Z of $\|\mathsf{Z}_\kappa\|_{[a,b]}$ at the continuity points of q_Z. Let $1-\alpha$ be one of those continuity points. Then, $\widehat{q}_{n,m}(1-\alpha)$ converges $P\otimes Q$-a.s. to $q_\mathsf{Z}(1-\alpha)$, and it follows from Slutsky's theorem that $\sqrt{n+m}\sup_x|\widehat{T}_{n,m}(x)-T_{n,m}(x)|/\widehat{s}_{n,m}(x)-\widehat{q}_{n,m}(1-\alpha)$ converges in distribution to $\|\mathsf{Z}_\kappa\|_{[a,b]}-q_\mathsf{Z}(1-\alpha)$. Therefore,

$$\mathbb{P}\left(T_0(x)\in\left[\widehat{T}_{n,m}(x)\pm\widehat{s}_{n,m}(x)\widehat{q}_{n,m}(1-\alpha)/\sqrt{n+m}\right],\forall x\in[a,b]\right)$$
$$=\underbrace{\mathbb{P}\left(\sqrt{n+m}\sup_x\frac{|\widehat{T}_{n,m}(x)-T_{n,m}(x)|}{\widehat{s}_{n,m}(x)}\leq\widehat{q}_{n,m}(1-\alpha)\right)}_{\widehat{\mathsf{R}}_{n,m}(1-\alpha)}$$
$$\to\mathbb{P}\left(\|\mathsf{Z}_\kappa\|_{[a,b]}\leq q_\mathsf{Z}(1-\alpha)\right)=1-\alpha,\tag{33}$$

where the last equality holds because of the continuity of the distribution function of $\|\mathsf{Z}_\kappa\|_{[a,b]}$ (Lemma 11). With (33), we will show that the convergence (33) holds for all $\alpha\in(0,1)$. As $q_\mathsf{Z}(1-\alpha)$ is a monotone function of α, it has only countably many discontinuities. Consequently, given any $\alpha\in(0,1)$ and any $\varepsilon>0$, there exist continuity points $\alpha_L,\alpha_U\in(0,1)$ such that $\alpha_L<\alpha<\alpha_U$ and $\max\{\alpha_U-\alpha,\alpha-\alpha_L\}<\varepsilon/2$. Since the convergence (33) holds for α_L and α_U, we can find sufficiently large n and m such that $|\widehat{\mathsf{R}}_{n,m}(1-\alpha_L)-(1-\alpha_L)|<\varepsilon/2$ and $|\widehat{\mathsf{R}}_{n,m}(1-\alpha_U)-(1-\alpha_U)|<\varepsilon/2$. So we must have

$$\widehat{\mathsf{R}}_{n,m}(1-\alpha)\leq\widehat{\mathsf{R}}_{n,m}(1-\alpha_L)\leq1-\alpha_L+\varepsilon/2\leq1-\alpha+\varepsilon,$$

and similarly,

$$\widehat{\mathsf{R}}_{n,m}(1-\alpha)\geq\widehat{\mathsf{R}}_{n,m}(1-\alpha_U)\geq1-\alpha_U-\varepsilon/2\geq1-\alpha-\varepsilon,$$

which shows that $\widehat{\mathsf{R}}_{n,m}(1-\alpha)\to1-\alpha$ for all $\alpha\in(0,1)$ as desired. $\qquad\square$

Acknowledgments

Funding

References

[ACB17] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International Conference on Machine Learning*, pages 214–223. PMLR, 2017.

[BCP19] Jérémie Bigot, Elsa Cazelles, and Nicolas Papadakis. Central limit theorems for entropy-regularized optimal transport on finite spaces and statistical applications. *Electronic Journal of Statistics*, 13:5120–5150, 2019. MR4041704

[BL19] Sergey Bobkov and Michel Ledoux. *One-dimensional Empirical Measures, Order Statistics, and Kantorovich Transport Distances*, volume 261. American Mathematical Society, 2019. MR4028181

[BR78] Monique Bertrand-Retali. Convergence uniforme d'un estimateur de la densité par la méthode du noyau. *Rev. Roumaine Math. Pures Appl.*, 23:361–385, 1978. MR0494658

[BRPP15] Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and Radon Wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51:22–45, 2015. MR3300482

[CCK14] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Anti-concentration and honest, adaptive confidence bands. *The Annals of Statistics*, 42(5):1787–1818, 2014. MR3262468

[Cut13] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in Neural Information Processing Systems*, 26, 2013.

[DBCAMRR99] Eustasio Del Barrio, Juan A Cuesta-Albertos, Carlos Matrán, and Jesús M Rodríguez-Rodríguez. Tests of goodness of fit based on the L_2-Wasserstein distance. *Annals of Statistics*, 1230–1239, 1999. MR1740113

[dBGSL24] Eustasio del Barrio, Alberto González Sanz, and Jean-Michel Loubes. Central limit theorems for semi-discrete Wasserstein distances. *Bernoulli*, 30(1):554–580, 2024. MR4665589

[DBGU05] Eustasio Del Barrio, Evarist Giné, and Frederic Utzet. Asymptotics for l2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli*, 11(1):131–189, 2005. MR2121458

[dBL19] Eustasio del Barrio and Jean-Michel Loubes. Central limit theorem for empirical transportation cost in general dimension. *Annals of Probability*, 2, 2019. MR3916938

[dBSLNW23] Eustasio del Barrio, Alberto González Sanz, Jean-Michel Loubes, and Jonathan Niles-Weed. An improved central limit theorem and fast convergence rates for entropic transportation costs. *SIAM Journal on Mathematics of Data Science*, 5(3):639–669, 2023. MR4616887

[DGS21] Nabarun Deb, Promit Ghosal, and Bodhisattva Sen. Rates of es-

timation of optimal transport maps using plug-in estimators via barycentric projections. *Advances in Neural Information Processing Systems*, 34:29736–29753, 2021.

[DNWP22] Vincent Divol, Jonathan Niles-Weed, and Aram-Alexandre Pooladian. Optimal transport map estimation in general function spaces. *arXiv preprint arXiv:2212.03722*, 2022.

[DR23] Neil Deo and Thibault Randrianarisoa. On adaptive confidence sets for the Wasserstein distances. *Bernoulli*, 29(3):2119–2141, 2023. [MR4580910](MR4580910)

[Dud69] Richard Mansfield Dudley. The speed of mean Glivenko-Cantelli convergence. *The Annals of Mathematical Statistics*, 40(1):40–50, 1969. [MR0236977](MR0236977)

[Gal18] Alfred Galichon. *Optimal Transport Methods in Economics*. Princeton University Press, 2018. [MR3586373](MR3586373)

[GCB+19] Aude Genevay, Lénaic Chizat, Francis Bach, Marco Cuturi, and Gabriel Peyré. Sample complexity of Sinkhorn divergences. In *Artificial Intelligence and Statistics*, pages 1574–1583. PMLR, 2019.

[GKRS22a] Ziv Goldfeld, Kengo Kato, Gabriel Rioux, and Ritwik Sadhu. Limit theorems for entropic optimal transport maps and the Sinkhorn divergence. *arXiv preprint arXiv:2207.08683*, 2022.

[GKRS22b] Ziv Goldfeld, Kengo Kato, Gabriel Rioux, and Ritwik Sadhu. Statistical inference with regularized optimal transport. *arXiv preprint arXiv:2205.04283*, 2022.

[GN07] Evarist Giné and Richard Nickl. Uniform central limit theorems for kernel density estimators. *Probability Theory and Related Fields*, 141(3-4):333–387, July 2007. [MR2391158](MR2391158)

[HHZ08] Lajos Horváth, Zsuzsanna Horváth, and Wang Zhou. Confidence bands for ROC curves. *Journal of Statistical Planning and Inference*, 138(6):1894–1904, 2008. [MR2400485](MR2400485)

[HR21] Jan-Christian Hütter and Philippe Rigollet. Minimax estimation of smooth optimal transport maps. *The Annals of Statistics*, 49(2), 2021. [MR4255123](MR4255123)

[IOH22] Masaaki Imaizumi, Hirofumi Ota, and Takuo Hamaguchi. Hypothesis test and confidence analysis with Wasserstein distance on general dimension. *Neural Computation*, 34(6):1448–1487, 2022. [MR4431591](MR4431591)

[Kan60] Leonid V Kantorovich. Mathematical methods of organizing and planning production. *Management Science*, 6(4):366–422, 1960. [MR0129016](MR0129016)

[KMT76] János Komlós, Péter Major, and Gábor Tusnády. An approximation of partial sums of independent RV's, and the sample DF. II. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 34:33–58, 1976. [MR0402883](MR0402883)

[KTM20] Marcel Klatt, Carla Tameling, and Axel Munk. Empirical regularized optimal transport: Statistical theory and applications.

SIAM Journal on Mathematics of Data Science, 2(2):419–443, 2020. MR4105566

[Lei20] Jing Lei. Convergence and concentration of empirical measures under Wasserstein distance in unbounded functional spaces. *Bernoulli*, 26(1):767–798, 2020. MR4036051

[LFH+20] Tianyi Lin, Chenyou Fan, Nhat Ho, Marco Cuturi, and Michael Jordan. Projection robust Wasserstein distance and Riemannian optimization. *Advances in Neural Information Processing Systems*, 33:9383–9397, 2020.

[Max23] Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). Human Mortality Database. http://www.mortality.org, 2023. Accessed: June 9, 2023.

[MBNWW21] Tudor Manole, Sivaraman Balakrishnan, Jonathan Niles-Weed, and Larry Wasserman. Plugin estimation of smooth optimal transport maps. *arXiv preprint arXiv:2107.12364*, 2021.

[MBNWW23] Tudor Manole, Sivaraman Balakrishnan, Jonathan Niles-Weed, and Larry Wasserman. Central limit theorems for smooth optimal transport maps. *arXiv preprint arXiv:2312.12407*, 2023.

[MBW22] Tudor Manole, Sivaraman Balakrishnan, and Larry Wasserman. Minimax confidence intervals for the sliced Wasserstein distance. *Electronic Journal of Statistics*, 16(1):2252–2345, 2022. MR4402565

[MC98] Axel Munk and Claudia Czado. Nonparametric validation of similar distributions and assessment of goodness of fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):223–241, 1998. MR1625620

[MNW19] Gonzalo Mena and Jonathan Niles-Weed. Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. *Advances in Neural Information Processing Systems*, 32, 2019.

[MNW21] Tudor Manole and Jonathan Niles-Weed. Sharp convergence rates for empirical optimal transport with smooth costs. *arXiv preprint arXiv:2106.13181*, 2021. MR4700254

[NWR22] Jonathan Niles-Weed and Philippe Rigollet. Estimation of Wasserstein distances in the spiked transport model. *Bernoulli*, 28(4):2663–2688, 2022. MR4474558

[OI23] Ryo Okano and Masaaki Imaizumi. Inference for projection-based Wasserstein distances on finite spaces. *Statistica Sinica*, 2023.

[PC+19] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[PNW21] Aram-Alexandre Pooladian and Jonathan Niles-Weed. Entropic estimation of optimal transport maps. *arXiv preprint arXiv:2109.12004*, 2021.

[RGTC17] Aaditya Ramdas, Nicolás García Trillos, and Marco Cuturi. On Wasserstein two-sample testing and related families of nonparametric tests. *Entropy*, 19(2):47, 2017. MR3608466

[SGK23] Ritwik Sadhu, Ziv Goldfeld, and Kengo Kato. Limit theorems for semidiscrete optimal transport maps. *arXiv preprint arXiv: 2303.10155*, 2023.

[Sil86] Bernard W Silverman. *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London, 1986. MR0848134

[SM18] Max Sommerfeld and Axel Munk. Inference for empirical Wasserstein distances on finite spaces. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 80(1):219–238, 2018. MR3744719

[TGR21] William Torous, Florian Gunsilius, and Philippe Rigollet. An optimal transport approach to causal inference. *arXiv preprint arXiv:2108.05858*, 2021.

[TSM19] Carla Tameling, Max Sommerfeld, and Axel Munk. Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications. *The Annals of Applied Probability*, 29(5):2744–2781, 2019. MR4019874

[Tsy08] Alexandre B Tsybakov. *Introduction to Nonparametric Estimation.* Springer Science & Business Media, 2008. MR2724359

[V$^+$09] Cédric Villani et al. *Optimal Transport: Old and New*, volume 338. Springer, 2009. MR2459454

[vdV98] Aad W. van der Vaart. *Asymptotic Statistics.* Cambridge University Press, October 1998. MR1652247

[Vil21] Cédric Villani. *Topics in Optimal Transportation*, volume 58. American Mathematical Soc., 2021. MR1964483

[Was06] Larry Wasserman. *All of Nonparametric Statistics.* Springer Science & Business Media, 2006. MR2172729

[WB19] Jonathan Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 2019. MR4003560