Estimating Certain Integral Probability Metrics (IPMs) Is as Hard as Estimating under the IPMs

Tengyuan Liang

OCTOBER 2020
Estimating Certain Integral Probability Metrics (IPMs) Is as Hard as Estimating under the IPMs

Tengyuan Liang∗

1University of Chicago

Abstract

We study the minimax optimal rates for estimating a range of Integral Probability Metrics (IPMs) between two unknown probability measures, based on \( n \) independent samples from them. Curiously, we show that estimating the IPM itself between probability measures is not significantly easier than estimating the probability measures under the IPM. We prove that the minimax optimal rates for these two problems are multiplicatively equivalent, up to a \( \log \log(n)/\log(n) \) factor.

1 Introduction

In this note, we study the minimax optimal rates for estimating the Integral Probability Metrics (IPMs) between probability measures based on samples. IPMs are widely used in both statistics and machine learning, with applications in nonparametric two-sample tests (Sriperumbudur et al., 2012; Gretton et al., 2012), inferring the transportation cost (the Wasserstein-1 metric) from one set of samples to another (Sommerfeld and Munk, 2018; Peyré and Cuturi, 2019), and with more recent appearances in rigorous investigations of the generative adversarial networks (GANs) (Arjovsky et al., 2017; Liu et al., 2017; Liang, 2018; Singh et al., 2018; Uppal et al., 2019; Bińkowski et al., 2018).

Let \( \mu, \nu \) be two probability measures supported on \( \Omega = [0,1]^d \), and \( d_\mathcal{F}(\mu, \nu) \) denote a certain IPM induced by a set of functions \( \mathcal{F} \), defined as

\[
d_\mathcal{F}(\mu, \nu) := \sup_{f \in \mathcal{F}} \left| \int_{\Omega} f \, d\mu - \int_{\Omega} f \, d\nu \right|.
\]

Consider that \( X_1, \ldots, X_m \) are i.i.d samples from \( \mu \), and \( Y_1, \ldots, Y_n \) i.i.d from \( \nu \). We study the minimax optimal rate for estimating \( d_\mathcal{F}(\mu, \nu) \) based on \( \{X_i\}_{i=1}^m, \{Y_j\}_{j=1}^n \), for some class of probability measures \( \mathcal{G} \) of interest

\[
\inf_{\tilde{T}_{m,n}} \sup_{\mu, \nu \in \mathcal{G}} \mathbb{E}[|\tilde{T}_{m,n} - d_\mathcal{F}(\mu, \nu)|].
\]

∗tengyuan.liang@chicagobooth.edu. Liang gratefully acknowledges support from the George C. Tiao Fellowship. The paper was previously titled ‘‘On the minimax optimality of estimating Wasserstein metric’’ (Liang, 2019). The previous version is no longer intended for publication.
It turns out that using the empirical measure $\hat{\nu}_n$ to estimate is a bad idea when $F$ is complex enough, regardless of how simple $G$ is. To see this, let’s consider a simple case with $F = \text{Lip}(1)$. In such a case, $d_F$ reduces to the Wasserstein-1 metric $W(1.10)$. Due to results by Dudley (1969); Ajtai et al. (1984), even for infinitely smooth $G = \{\text{Unif}(\Omega)\}$ and $d \geq 3$,

$$\sup_{\nu \in G} |W(\mu, \nu) - W(\mu, \hat{\nu}_n)| \asymp n^{-\frac{1}{d}}.$$

(1.3)

A natural question arises: can one obtain faster rates, for estimating the IPM with other estimators $\tilde{T}_{m,n}$ leveraging certain regularity of $G$ such as smoothness?

A related yet different problem studied in the current literature is estimating a probability measure under certain IPMs (Sriperumbudur et al., 2012; Weed and Bach, 2017; Liang, 2018; Singh et al., 2018; Weed and Berthet, 2019), in the following sense

$$\inf_{\tilde{T}_{m,n}} \sup_{\nu \in G} \mathbb{E} d_F(\tilde{T}_{m,n}, \nu) .$$

(1.4)

The two problems are closely related: “estimating the metric itself” is usually an easier problem than “estimating under the metric.” In fact, the solution of the latter problem $\tilde{\mu}_m, \tilde{\nu}_n$ naturally induces a plug-in answer to the former, since

$$\mathbb{E} |d_F(\tilde{\mu}_m, \tilde{\nu}_n) - d_F(\mu, \nu)| \leq \mathbb{E} d_F(\tilde{\mu}_m, \tilde{\nu}_n) - d_F(\mu, \nu) .$$

However, it is unclear whether such a plug-in estimator is optimal. In fact, it is well-known that estimating specific functional of density $F(\nu)$ is usually strictly easier than estimating the density $\nu$ itself. For example, in estimating quadratic functionals of a smooth density vs. estimating under the quadratic functionals, the plug-in approach is strictly sub-optimal, where the rates can be much-improved (Bickel and Ritov, 1988; Donoho and Nussbaum, 1990; Fan, 1991). In recent practical applications such as GANs, one is curious to understand if evaluating and inferring how well we do in terms of learning the probability measure, could be simpler than learning the measure itself (Lucic et al., 2018; Liang and Stokes, 2018).

In this paper, however, we prove that “estimating the IPMs,” is not significantly easier than “estimating under the IPMs,” for a wide range of measures and metrics. Specifically, the plug-in approach is minimax optimal up to a log log(n)/log(n) factor

$$\log \log(n \wedge m) \cdot (n \wedge m)^{-\frac{\beta \gamma}{d(\beta + \gamma)}} \leq \inf_{\tilde{T}_{m,n}} \sup_{\nu \in G_\beta} \mathbb{E} |\tilde{T}_{m,n} - d_{F_\gamma}(\mu, \nu)|$$

$$\leq \inf_{\tilde{\mu}_m, \tilde{\nu}_n} \sup_{\nu \in G_\beta} \mathbb{E} |d_{F_\gamma}(\tilde{\mu}_m, \tilde{\nu}_n) - d_{F_\gamma}(\mu, \nu)| \lesssim (n \wedge m)^{-\frac{\beta \gamma}{d(\beta + \gamma)}} .$$

Here $G_\beta$ contains probability measures with densities in the Hölder space with smoothness $\beta \in \mathbb{R}_{\geq 0}$, and the IPMs are induced by $F_\gamma$, the Hölder space with smoothness $\gamma \in \mathbb{R}_{\geq 0}$, with $\gamma < d/2$. Note that when $\gamma \geq d/2$, the parametric rate $n^{-1/2}$ is attainable. The result informs us that (1) seeking for other forms of estimators for $d_{F_\gamma}(\mu, \nu)$ would only improve the rates logarithmically, and (2) estimating the IPM between two measures is fundamentally just as hard as estimating the measure under the IPM.
1.1 Preliminaries

We introduce the notations used in the paper. For a function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ and $p \geq 1$, $\|f\|_{L^p}$ denotes the $L_p$ norm w.r.t. the Lebesgue measure. For a finite dimensional vector $\theta$ and $q \geq 1$, $\|\theta\|_q$ is the vector $L_q$ norm, and $\|\theta\|$ is the $L_2$ norm. For an integer $K$, $[K] := \{0, 1, \ldots, K-1\}$.

Let $C^\beta := C[\beta, \beta] \to \mathbb{R}$ to be Hölder space with smoothness $\beta > 0$.

$$
C^\beta := \left\{ f : \Omega \to \mathbb{R} : \max_{|\alpha| \leq \beta} \sup_{x \in \Omega} |D^\alpha f| + \max_{|\alpha| = \beta} \sup_{x \neq y \in \Omega} \frac{|D^\alpha f(x) - D^\alpha f(y)|}{\|x - y\|^\beta} < \infty \right\}
$$

(1.5)

where $\alpha = [\alpha_1, \ldots, \alpha_d] \in \mathbb{N}^d$ ranges over multi-indices, and $|\alpha| := \sum_{i=1}^d \alpha_i$. We only consider the bounded case with $\Omega = [0, 1]^d$.

Define the forward difference operator, given a function $f : \Omega \to \mathbb{R}$, for every $h \in \mathbb{R}^d$

$$
\Delta_h f(x) := f(x + h) - f(x),
$$

$$
\Delta_h^m f(x) := \Delta_h (\Delta_h^{m-1} f(x)), m \geq 2.
$$

Let $1 \leq p, q \leq \infty$ and $\beta > 0$, the Besov-Lipschitz space semi-norm $\| \cdot \|_{B^\beta}^p$ is defined in the following way (Leoni, 2017, Chapter 17, Proposition 17.21).

$$
\|f\|_{B^\beta}^p := \left( \sum_{j=0}^{\infty} \left( (2^j)^\beta \sup_{\|\alpha\| \leq 1/2} \|\Delta^{|\beta|+1}_h f\|_{L^p} \right)^q \right)^{1/q}.
$$

(1.6)

Wavelets are used to provide an equivalent characterization of the Besov spaces, if the basis $\{h_{jk}, j \in \mathbb{N}, 0 \leq k < 2^{jd}\}$ satisfies certain regularity conditions (Härdle et al., 2012, Chapter 9, Theorem 9.2). For function $f(x) = \sum_{j=0}^{\infty} \sum_{k \leq 2^j} \theta_{j,k} h_{j,k}(x)$, define the Besov space norm in terms of the wavelet coefficients (Weed and Berthet, 2019)

$$
\|f\|_{B^\beta}^p := \left( \sum_{j=0}^{\infty} \left( (2^j)^\beta (2^j)^{\frac{1}{2} - \frac{1}{\beta}} \|\theta_j\|_p \right)^q \right)^{1/q}.
$$

(1.7)

In this paper, we assume such regularity conditions throughout so that the Besov space norms $\| \cdot \|_{B^\beta}^p$ and $\| \cdot \|_{B^\beta}^q$ are equivalent. We refer the readers to (Weed and Berthet, 2019, Appendix C) and (Cohen, 2003, Chapter 2.12) for details on the regularity conditions that we assume on the wavelets.

Besov spaces subsume Hölder spaces as special cases ($p = q = \infty$), in the following sense (Triebel, 1980; Donoho et al., 1996; Härdle et al., 2012): under regularity conditions, the following equivalence holds between the Besov space and Hölder space

$$
B^{\beta, \infty}_\infty = C^\beta, \text{for } \beta \in \mathbb{N}.
$$

In particular, when $\beta = 1$, $B^{1, \infty}_\infty$ is called the Zygmund space, which contains the Lipschitz space $B^{1, \infty}_\infty \supseteq \text{Lip} \supseteq B^{1, \infty}_1$.

Now we are ready to formally state the parameter spaces, and the IPMs to study.
Parameter Spaces. For some $M > 0$, the class of probability measures of interest is
\[
\mathcal{G}_\beta := \left\{ \mu : \int_{\Omega} d\mu = 1, \mu \geq 0, \frac{d\mu}{dx} \in B^\infty_{\infty}(M) \right\},
\] (1.8)
with
\[
B^\infty_{\infty}(M) := \left\{ f : \|f\|_{B^\infty_{\infty}} \leq M \right\}.
\]
Again, for non-integer $\beta$, we are considering densities that are Hölder smooth.

Integral Probability Metric. The class of IPMs considered is induced by the Besov space, for some $\gamma > 0$
\[
\mathcal{F}_\gamma := B^\gamma_{\infty,\infty}(1),
\]
\[
d_{\mathcal{F}_\gamma}(\mu, \nu) = \sup_{f \in \mathcal{F}_\gamma} \left| \int f d\mu - \int f d\nu \right|.
\] (1.9)
As a special case for the IPMs, the Wasserstein-1 metric (for measures supported on bounded $\Omega$) is
\[
W(\mu, \nu) := \sup_{f \in \text{Lip}(1)} \left| \int f d\mu - \int f d\nu \right|.
\] (1.10)

2 Optimal Rates for Estimating IPMs

Theorem 2.1 (Minimax Rate). Consider the domain $\Omega = [0, 1]^d$. Given $m$ i.i.d. samples $X_1, \ldots, X_m$ from $\mu$, and $n$ i.i.d. samples $Y_1, \ldots, Y_n$ from $\nu$, the minimax optimal rate for estimating $d_{\mathcal{F}_\gamma}(\mu, \nu)$ satisfies
\[
\frac{\log \log(n \wedge m)}{\log(n \wedge m)} \cdot (n \wedge m)^{-\frac{\beta + \gamma}{2d}} \leq \inf_{\tilde{T}_{m,n, \mu, \nu} \in \mathcal{G}_\beta} \mathbb{E} |\tilde{T}_{m,n} - d_{\mathcal{F}_\gamma}(\mu, \nu)| \lesssim (n \wedge m)^{-\frac{\beta + \gamma}{2d}}.
\] (2.1)
Here $\mu, \nu$ lie in $\mathcal{G}_\beta, \beta \in \mathbb{R}_{\geq 0}$ as in (1.8) whose densities are $\beta$-Hölder smooth. The function class $\mathcal{F}_\gamma, \gamma \in \mathbb{R}_{\geq 0}$ for the metric is defined in (1.9), with $\gamma < d/2$.

Remark. Here the $\beta$ quantifies the regularity of the measures, and $\gamma$ quantifies the regularity of the metrics.

A few remarks are in order. First, we emphasize that the main technicality is in deriving the lower bound. We construct two composite/fuzzy hypotheses using delicate priors with matching $\log(n \wedge m)$ moments. However, the IPMs to estimate differs sufficiently under the null vs. the alternative. Then we calculate the Total Variation (TV) metric directly on the posterior of data samples defined by the composite hypothesis, using some telescoping techniques involving sum-products. The transparent technique could be of independent interest in handling TV-type calculations in proving lower bounds. Second, as a direct corollary, the following extension holds true. Suppose $\mu \in \mathcal{G}_{\beta_1}$ and $\nu \in \mathcal{G}_{\beta_2}$, then define $\beta := \beta_1 \wedge \beta_2$,
\[
\frac{\log \log(n \wedge m)}{\log(n \wedge m)} \cdot (n \wedge m)^{-\frac{\beta + \gamma}{2d}} \leq \inf_{\tilde{T}_{m,n, \mu, \nu} \in \mathcal{G}_{\beta_1}, \nu \in \mathcal{G}_{\beta_2}} \mathbb{E} |\tilde{T}_{m,n} - d_{\mathcal{F}_\gamma}(\mu, \nu)| \lesssim (n \wedge m)^{-\frac{\beta + \gamma}{2d}}.
\]
Third, the $\gamma < d/2$ condition is effectively equivalent to that $\mathcal{F}_y$ is beyond the Donsker’s class. This is the complex regime since within the Donsker’s class $\gamma \geq d/2$, the parametric rate $n^{-1/2}$ is attainable (Gretton et al., 2012; Liang, 2018; Singh et al., 2018). Finally, we would like to remark that for the special case $\beta = 0$ and $\gamma = 1$, Niles-Weed and Rigollet (2019) independently obtains a lower bound with a similar logarithmic factor gap, using a distinct approach. Admittedly, their logarithmic factor is better than ours. We note that closing the $\log n$ gap is an interesting question for future work.

2.1 Proof of the Lower Bound

Without the loss of generality, consider the case when $m \geq n$. The lower bound construction is divided into six logical steps, for better organization. We make use of multi-resolution analysis in the construction.

**Step 1: reduction to Besov space semi-norm.** For any $p \geq 1$, define $p_* \geq 1$ such that $1/p_* + 1/p = 1$. For simplicity, define Radon-Nikodym derivative of measure $\mu$ w.r.t. the Lebesgue measure as $\rho_\mu(x) := d\mu/dx$. Define explicitly the wavelet coefficients $f_{jk} := \langle f, h_{jk} \rangle$, and $u_{jk} := \langle d\mu/dx, h_{jk} \rangle$, $v_{jk} := \langle d\nu/dx, h_{jk} \rangle$. Under such notations, the integral probability metric reduces to the following

$$d_{\mathcal{B}_q}(\mu, \nu) := \sup_{f \in \mathcal{B}_q} \left| \int f d\mu - \int f d\nu \right|$$

$$= \sup_{f \in \mathcal{B}_q} \left| \sum_{j \geq 0} \sum_{k=0}^{2^j-1} f_{jk} (u_{jk} - v_{jk}) \right|$$

$$= \sup_{f \in \mathcal{B}_q} \left| \sum_{j \geq 0} \|f_j\|_{p} \|u_j - v_j\|_{p_*} \right|$$

$$= \sup_{f \in \mathcal{B}_q} \left| \sum_{j \geq 0} (2^{dj})^{\gamma d + 1/p} \|f_j\|_p \cdot (2^{-dj})^{\gamma d + 1/p} \|u_j - v_j\|_{p_*} \right|$$

$$= \left\{ \sum_{j \geq 0} \left[ (2^{dj})^{\gamma d + 1/p} \|f_j\|_p \right]^{q/p} \right\}^{1/q} \left\{ \sum_{j \geq 0} \left[ (2^{-dj})^{\gamma d + 1/p} \|u_j - v_j\|_{p_*} \right]^{q/p} \right\}^{1/q}.$$ 

Take $p = q = \infty$ (in this case $p_* = q_* = 1$), we know that the IPM can be regarded as a type of Besov space semi-norm

$$d_{\mathcal{B}_q}(\mu, \nu) = \sum_{j \geq 0} (2^{-dj})^{\gamma d + 1/p} \sum_{k=0}^{2^j-1} |u_{jk} - v_{jk}|.$$ 

**Step 2: composite hypothesis and prior construction.** Next we are going to construct two priors on $\nu$ such that the difference

$$\left| \mathbb{E}_{\nu \sim P_0} d_{\mathcal{F}_y}(\mu, \nu) - \mathbb{E}_{\nu \sim P_1} d_{\mathcal{F}_y}(\mu, \nu) \right|$$

5
is large, while at the same time one can not distinguish the following two distributions

\[ p_0(Y_1, \ldots, Y_n) = \mathbb{E}_{\nu \sim P_0} \left[ \prod_{i=1}^{n} \rho_\nu(Y_i) \right], \quad p_1(Y_1, \ldots, Y_n) = \mathbb{E}_{\nu \sim P_1} \left[ \prod_{i=1}^{n} \rho_\nu(Y_i) \right]. \]  

(2.2)

Here \( P_0, P_1 \) are two prior distributions on \( \nu \) which we will construct. Consider \( \mu \) to be the same distribution under the null \( H_0 \) and the alternative \( H_1 \). Set two values \( K, \tau \) to be used in the construction

\[ K \asymp \log n \log \log n, \quad \tau \asymp 1. \]  

(2.3)

The choice will be apparent in the latter part of the proof. The following prior construction is inspired by Lepski et al. (1999), where they study the estimation of functionals under the Gaussian white noise model. This prior was also used in Cai and Low (2011) for studying non-smooth functional estimation in Gaussian sequence models.

**Proposition 2.2 (Lepski et al. (1999), Proposition 4.2).** For any given positive integer \( K \) and \( \tau \in \mathbb{R}_{\geq 0} \), there exist two symmetric probability measures \( q_0 \) and \( q_1 \) on \([-\tau, \tau]\) such that

\[ \int_{-\tau}^{\tau} t^l q_0(dt) = \int_{-\tau}^{\tau} t^l q_1(dt), \quad l = 0, 1, \ldots, 2K, \]  

(2.4)

\[ \int_{-\tau}^{\tau} |t| q_1(dt) - \int_{-\tau}^{\tau} |t| q_0(dt) = 2 \kappa \cdot K^{-1} \tau, \]  

(2.5)

where \( \kappa \) is some universal constant that does not depend on \( K, \tau \).

Now let’s construct \( P_0 \) and \( P_1 \) as follows. Take \( \mu \sim \text{Unif}([0, 1]^d) \). Choose \( J \in \mathbb{N} \) such that \( 2^d J \asymp n^{\frac{1}{1+2\beta/d}} \), first we embed a parametrized class of densities into

\[ G_\beta = \{ \mu(x) + \frac{1}{\sqrt{n}} \sum_{k=0}^{2^d J-1} \theta_k h_{jk}(x) \} \]  

with each \( \theta_k \in [-\tau, \tau] \) for all \( k \). Now we show that the construction lies inside the space of interest, i.e., \( \nu_0 \in \mathcal{G}_\beta \). First observe that for the wavelet basis that satisfy the regularity condition \( \int_{\Omega} h_{jk} d\mu = 0 \), we have \( \int_{\Omega} d\nu_0 = 1 \) and \( \|d\nu_0/dx\|_{L_\infty} \geq 1 - \sqrt{2^d J/n} > 0 \). Hence \( \nu_0 \) is a valid probability measure. Let’s then verify that the density \( \rho_{\nu_0} \in B_{\infty, \infty}^{\beta, \infty} \) lies in the Besov space, since

\[ \frac{1}{\sqrt{n}} |\theta_k| \leq (2^{dJ})^{-\left(\frac{\gamma}{2} + \frac{1}{2}\right)}, \quad \forall k. \]  

(2.7)

For any \( \gamma \geq 0 \), it is then easy to verify via Step 1 that

\[ d_{\mathcal{F}_\gamma}(\mu, \nu_0) := (2^{-dJ})^{\frac{\gamma}{2} + \frac{1}{2}} \frac{1}{\sqrt{n}} \sum_{k=0}^{2^{dJ}-1} |\theta_k| \]  

\[ = (2^{-dJ})^{\frac{\gamma}{2} + \frac{1}{2}} (2^{dJ})^{-\left(\frac{\gamma}{2} + \frac{1}{2}\right)} \sum_{k=0}^{2^{dJ}-1} |\theta_k| \]  

\[ = (2^{-dJ})^{\frac{\gamma}{2} + \frac{1}{2}} \frac{1}{2^{dJ}} \sum_{k=0}^{2^{dJ}-1} |\theta_k|. \]  

(2.8)
Making use of the probability measures \( q_0 \) and \( q_1 \) on \([−τ, τ]\) claimed by Proposition 2.2, we define a collection of measures
\[
S_0 := \{ν_0 : θ_k \sim q_0 \text{ i.i.d. for each } k \in [2^d] \}.
\]

Then \( P_0 \) can be viewed as an uniform prior over this set \( S_0 \). Similar construction holds for \( P_1 \) via \( q_1 \).

**Step 3: polynomials and matching moments.** For the lower bound construction, we consider a subset of the wavelet family with disjoint support. Remark that due to the localization of the support, i.e., \( h_{jk}(x)h_{jk'}(x) = 0 \) for \( k \neq k' \), we have the equivalent expression as in (2.6)
\[
\frac{dν_0}{dx} = \prod_{k=0}^{2^d-1} (1 + θ_k n^{-1/2} h_{jk}(x)) .
\]  

(2.9)

Use \( θ \sim q_0^{\otimes 2^d} \) to denote that \( θ_k \sim q_0 \text{ i.i.d. for each } k \in [2^d] \), we know
\[
p_0(Y_1, \ldots, Y_n) = \mathbb{E}_{θ \sim q_0^{\otimes 2^d}} \prod_{i=1}^{n} \rho_θ(Y_i) = \mathbb{E}_{θ \sim q_0^{\otimes 2^d}} \prod_{i=1}^{n} \prod_{k=0}^{2^d-1} (1 + θ_k n^{-1/2} h_{jk}(Y_i))
\]
\[
= \mathbb{E}_{θ \sim q_0^{\otimes 2^d}} \prod_{k=0}^{2^d-1} \prod_{i=1}^{n} (1 + θ_k n^{-1/2} h_{jk}(Y_i)) \quad \text{by (2.9)}
\]
\[
= \prod_{k=1}^{2^d} \mathbb{E}_{θ_k \sim q_0} \prod_{i=1}^{n} (1 + θ_k n^{-1/2} h_{jk}(Y_i)) .
\]  

(2.10)

Remark that we can not further interchange the ordering of \( \mathbb{E}_{θ_k} \) and \( \prod_{i=1}^{n} \).

Let’s introduce the polynomial \( f(θ_k; h_{jk}(Y_1), \ldots, h_{jk}(Y_n)) \) in \( θ_k \) (and \( h_{jk}(Y_i) \)) with degree at most \( n \) appearing in the above expression, which will be used extensively in the next step,
\[
f(θ_k; h_{jk}(Y_1), \ldots, h_{jk}(Y_n)) := \prod_{i=1}^{n} (1 + θ_k \frac{h_{jk}(Y_i)}{\sqrt{n}})
\]  

(2.11)

\[
= \sum_{l=0}^{n} \theta_k^l \frac{\sum_{i_1 < \cdots < i_l} h_{jk}(Y_{i_1}) \cdots h_{jk}(Y_{i_l})}{\sqrt{n}^l} =: \sum_{l=0}^{n} \theta_k^l H_{jk}^{(l)}(Y_1, \ldots, Y_n) .
\]

Here \( H_{jk}^{(l)}(Y_1, \ldots, Y_n) \) is a sum of monomials of order \( l \), i.e., \( \binom{n}{l} \) terms with each of the form \( h_{jk}(Y_{i_1}) \cdots h_{jk}(Y_{i_l}) \). Denote \( f^{[<K]}, f^{[>K]} \) to be the corresponding truncated polynomial according to the degree. In this convenient notation, we know
\[
p_0(Y_1, \ldots, Y_n) = \prod_{k \in [2^d]} \mathbb{E}_{θ_k \sim q_0} f(θ_k; h_{jk}(Y_1), \ldots, h_{jk}(Y_n)).
\]  

(2.12)

Later, we shall use the following properties of the polynomial \( f \) of degree at most \( n \),
\[
∀θ_k, \int_{\mathbb{R}^n} f(θ_k; h_{jk}(y_1), \ldots, h_{jk}(y_n))dy_1 \cdots dy_n = 1 .
\]  

(2.13)
And the following property according to \( q_0 \) and \( q_1 \) constructed in Proposition 2.2: \( \forall y_1, \ldots, y_n \)

\[
\mathbb{E}_{\theta_k \sim q_1} f(\theta_k; h_j(y_1), \ldots, h_j(y_n)) - \mathbb{E}_{\theta_k \sim q_0} f(\theta_k; h_j(y_1), \ldots, h_j(y_n))
= \int_{[-\tau, \tau]} f^{[2K]}(\theta_k; h_j(y_1), \ldots, h_j(y_n))(q_1 - q_0)(d\theta_k).
\]

The above explains the intuition on why we construct matching moments up to degree \( 2K \) in our lower bound.

**Step 4: total variation, telescoping and the sum-product trick.** When there is no confusion, we use \( f(\theta_k; h_j(y^{\otimes n})) \) to abbreviate \( f(\theta_k; h_j(y_1), \ldots, h_j(y_n)) \). Recall (2.10), we have

\[
TV(p_1, p_0) := \frac{1}{2} \int_{\Omega^{2n}} \left| p_1(y_1, \ldots, y_n) - p_0(y_1, \ldots, y_n) \right| dy_1 \ldots dy_n
= \frac{1}{2} \int_{\Omega^{2n}} \left| \prod_{k \in [2^d]} \mathbb{E}_{\theta_k \sim q_1} f(\theta_k; h_j(y^{\otimes n})) - \prod_{k \in [2^d]} \mathbb{E}_{\theta_k \sim q_0} f(\theta_k; h_j(y^{\otimes n})) \right| dy_1 \ldots dy_n.
\]

We claim that the following telescoping Lemma holds. The proof can be seen clearly through writing the left hand side as a telescoping sum and using the triangle inequality.

**Proposition 2.3 (Telescoping).** For \( N \in \mathbb{N}, N \geq 2, \) and \( a_i, b_i \geq 0, 1 \leq i \leq N, \)

\[
\prod_{k \in [1, N]} a_k - \prod_{k \in [1, N]} b_k \leq \sum_{i \in [1, N]} |a_i - b_i| \cdot \prod_{k \in [1, i]} b_k \cdot \prod_{k \in (i, N]} a_k.
\tag{2.14}
\]

To make use of the above Lemma, define

\[
a_k(h_j(y_1), \ldots, h_j(y_n)) := \mathbb{E}_{\theta_k \sim q_1} f(\theta_k; h_j(y^{\otimes n}))
\tag{2.15}
\]
\[
b_k(h_j(y_1), \ldots, h_j(y_n)) := \mathbb{E}_{\theta_k \sim q_0} f(\theta_k; h_j(y^{\otimes n}))
\tag{2.16}
\]

Using the the above telescoping proposition, we claim

\[
TV(p_1, p_0) \leq \sum_{k \in [2^d]} \int \left| a_k - b_k \right| \cdot \left( \prod_{k' \in [0, k]} b_k' \prod_{k'' \in [k, 2^d-1]} a_k'' \right) dy_1 \ldots dy_n
= \sum_{k \in [2^d]} \mathbb{E}_{\theta_k \sim q_0, k' \in [0, k], Y_{1, \ldots, Y_n \sim \theta_k}} \mathbb{E}_{\theta_k' \sim q_1, k'' \in (k, 2^d-1]} |a_k(h_j(Y_1), \ldots, h_j(Y_n)) - b_k(h_j(Y_1), \ldots, h_j(Y_n))|.
\tag{2.17}
\]

The reasoning behind the last line is as follows. Firstly, we need to define a tilted measure \( \nu_{\theta, k} \) without the influence of the \( k \)-the coordinate \( \theta_k \),

\[
\rho_{\nu_{\theta, k}}(x) = \frac{d\nu_{\theta, k}}{dx} := 1 + \frac{1}{\sqrt{n}} \sum_{0 \leq k' \leq 2^d-1 \atop k \neq k'} \theta_k' h_{j'k'}(x) = \prod_{0 \leq k' \leq 2^d-1 \atop k \neq k'} \left( 1 + \frac{1}{\sqrt{n}} \theta_k' h_{j'k'}(x) \right).
\tag{2.18}
\]
From the properties established in Step 3, one can verify that

\[
\mathbb{E}_{\theta_0 \sim q_0, k \in [0,k]} \prod_{i=1}^{n} \rho_{Y_{0,k}}(y_i) = \mathbb{E}_{\theta_0 \sim q_0, k \in [0,k]} \prod_{i=1}^{n} \prod_{k' \leq k, 0 \leq k' \leq 2^d_i-1} \left( 1 + \frac{1}{\sqrt{n}} \theta_k \cdot h_{j,k}(y_i) \right)
\]

\[
= \mathbb{E}_{\theta_0 \sim q_0, k \in [0,k]} \prod_{k' \leq k, 0 \leq k' \leq 2^d_i-1} \left( 1 + \frac{1}{\sqrt{n}} \theta_k \cdot h_{j,k}(y_i) \right)
\]

\[
= \prod_{k' \in [0,k]} b_{k'} \prod_{k'' \in (k,2^d_i-1]} a_{k''}.
\]

(2.19)

Now we have proved (2.17), since by using Fubini’s theorem,

\[
\int |a_k - b_k| \left( \prod_{k' \in [0,k]} b_{k'} \prod_{k'' \in (k,2^d_i-1]} a_{k''} \right) \, \sqrt{n} \, dy_1 \ldots dy_n
\]

\[
= \mathbb{E}_{\theta_0 \sim q_0, k \in [0,k]} \int |a_k(h_{j,k}(y_1), \ldots, h_{j,k}(y_n)) - b_k(h_{j,k}(y_1), \ldots, h_{j,k}(y_n))| \prod_{i=1}^{n} \rho_{Y_{0,k}}(y_i) \, dy_1 \ldots dy_n.
\]

Let’s analyze the term

\[
\mathbb{E}_{Y_1, \ldots, Y_n \sim \nu_{\theta,i}} |a_k(h_{j,k}(Y_1), \ldots, h_{j,k}(Y_n)) - b_k(h_{j,k}(Y_1), \ldots, h_{j,k}(Y_n))|
\]

where \(Y_1, \ldots, Y_n\) are i.i.d. sampled from a measure \(\nu_{\theta,i}\). We emphasize that \(\nu_{\theta,i}\) agrees with the uniform measure \(\mu\) on the domain associated with \(h_{j,k}(x)\). Due to the separation of support for the wavelet basis, we know that the random variables

\(h_{j,k}(Y_i)\)

are only determined by \(\nu_{\theta,i}\) restricted to the domain of \(h_{j,k}\). Equivalently, the distributions of \(h_{j,k}(Y)\)'s are the same when \(Y \sim \nu_{\theta,i}\) and \(Y \sim \mu\). Hence for \(Y_1, \ldots, Y_n \sim \nu_{\theta,i}\),

\[
\mathbb{E}_{Y_1, \ldots, Y_n \sim \nu_{\theta,i}} |a_k(h_{j,k}(Y_1), \ldots, h_{j,k}(Y_n)) - b_k(h_{j,k}(Y_1), \ldots, h_{j,k}(Y_n))|
\]

\[
= \mathbb{E}_{Y_1, \ldots, Y_n \sim \mu} |a_k(h_{j,k}(Y_1), \ldots, h_{j,k}(Y_n)) - b_k(h_{j,k}(Y_1), \ldots, h_{j,k}(Y_n))|.
\]

Now one can directly bound the TV metric between the complex sum-product distribution \(p_0\) and \(p_1\) defined in (2.10),

\[
2\text{TV}(p_1, p_0) \leq \sum_{k=0}^{2^d_i-1} \mathbb{E}_{Y_1, \ldots, Y_n \sim \mu} |a_k(h_{j,k}(Y_1), \ldots, h_{j,k}(Y_n)) - b_k(h_{j,k}(Y_1), \ldots, h_{j,k}(Y_n))|
\]

\[
= \sum_{k=0}^{2^d_i-1} \int | \mathbb{E}_{\theta_k \sim q_1} f(\theta_k; h_{j,k}(y^{0^n})) - \mathbb{E}_{\theta_k \sim q_0} f(\theta_k; h_{j,k}(y^{0^n})) | \, dy_1 \ldots dy_n.
\]

(2.20)
Step 5: $\ell_2$ bound. In this section, we are going to bound, for a fixed $k$, the following expression using the properties of the $q_1$ and $q_0$ constructed with matching moments up to $2K$ (claimed by Proposition 2.2),

$$
\left| \int \left[ \mathbb{E} f(\theta_k; h_k(y^{\otimes n})) - \mathbb{E} f(\theta_k; h_k(y^{\otimes n})) \right] \, dy_1 \ldots dy_n \right|
$$

First, observe the $\ell_2$ bound

$$
\int |g_1 - g_2| d\mu \leq \left( \int (g_1 - g_2)^2 d\mu \right)^{1/2} .
$$

(2.21)

Let’s bound the $\ell_2$ norm, which takes the form

$$
\left( \int \left( \mathbb{E} f(\theta_k; h_k(y^{\otimes n})) - \mathbb{E} f(\theta_k; h_k(y^{\otimes n})) \right)^2 \, dy_1 \ldots dy_n \right)^{1/2} .
$$

(2.22)

Note now each $f(\theta; h_k(y^{\otimes n}))f(\theta'; h_k(y^{\otimes n}))$ for fixed $\theta, \theta'$ takes the following product form

$$
f(\theta; h_k(y^{\otimes n}))f(\theta'; h_k(y^{\otimes n})) = \prod_{i=1}^n \left( 1 + (\theta + \theta') \frac{h_k(Y_i)}{\sqrt{n}} + \theta \theta' \frac{h_k^2(Y_i)}{n} \right)
$$

and

$$
\int f(\theta; h_k(y^{\otimes n}))f(\theta'; h_k(y^{\otimes n})) \, dy^{\otimes n} = \left( 1 + \theta \theta' \frac{1}{n} \right)^n .
$$

Therefore we have for (2.22)

$$
(2.22) = \sum_{l=1}^{[n/2]} \left( \mathbb{E}_{\theta, \theta' \sim q_1} \left( \mathbb{E}_{\omega, \omega' \sim q_0} \left[ (\theta \theta')^{2l} \right] \right) + \mathbb{E}_{\theta, \theta' \sim q_1} \left( \mathbb{E}_{\omega, \omega' \sim q_0} \left[ (\theta \omega')^{2l} \right] \right) - \mathbb{E}_{\theta, \theta' \sim q_1, \omega, \omega' \sim q_0} \left[ (\theta \omega')^{2l} \right] \right) \left( \frac{n}{2l} \right) \frac{n}{2l} .
$$

Recall the crucial property that for all $l \leq K$, we know

$$
\mathbb{E}_{\theta \sim q_1} \left[ \theta^{2l} \right] = \mathbb{E}_{\theta \sim q_0} \left[ \theta^{2l} \right] \Rightarrow \left( \mathbb{E}_{\theta \sim q_1} \left[ \theta^{2l} \right] \right)^2 + \left( \mathbb{E}_{\theta \sim q_0} \left[ \theta^{2l} \right] \right)^2 - 2 \mathbb{E}_{\theta \sim q_1} \left[ \theta^{2l} \right] \mathbb{E}_{\theta \sim q_0} \left[ \theta^{2l} \right] = 0
$$

(2.23)
therefore the above summation equals
\[ (2.22) = \sum_{l=K+1}^{\lfloor n/2 \rfloor} \left( \sum_{q_0}^{q_1} \left( \frac{\Theta^{2l}}{q_1^{2l}} - 2 \frac{\Theta^{2l}}{q_0^{2l}} \right) \right)^{(\frac{n}{2l})} \]
\[ \leq \sum_{l=K+1}^{\lfloor n/2 \rfloor} 4^{4l} \frac{1}{(2l)!} \]
\[ \leq 4 \frac{\tau^4K}{(2K)!} \exp(\tau^4) . \]

Assemble the two bounds, we have
\[ \int \left| \mathbb{E} f(\theta_k; h_j_k(y^{\otimes n})) - \mathbb{E} f(\theta_k; h_j_k(y^{\otimes n})) \right| dy_1 \ldots dy_n \]
\[ \leq 2 \frac{\tau^2K}{\sqrt{(2K)!}} \exp(\tau^4/2) \quad (2.24) \]
\[ \leq 2 \frac{\tau^2K}{\sqrt{(2K)!}} \exp(\tau^4/2) \quad (2.25) \]

Step 6: combine all pieces. Now continuing (2.20), we have
\[ 2TV(p_1, p_0) \leq \sum_{k=0}^{2^{dJ}-1} \mathbb{E} \left| a_k(h_j_k(Y_1), \ldots, h_j_k(Y_n)) - b_k(h_j_k(Y_1), \ldots, h_j_k(Y_n)) \right| \]
\[ = \sum_{k=0}^{2^{dJ}-1} \int \left| \mathbb{E} f(\theta_k; h_j_k(y^{\otimes n})) - \mathbb{E} f(\theta_k; h_j_k(y^{\otimes n})) \right| dy_1 \ldots dy_n \]
\[ \leq 2^{dJ} \cdot \frac{\tau^2K}{\sqrt{(2K)!}} \exp(\tau^4/2) \leq \exp(c \log n - K \log K) . \]

Recall \( \tau \approx 1 \) and that \( 2^{dJ} = n^c \). Therefore by taking \( K = \frac{\log n}{\log \log n} \), we know
\[ 2TV(p_1, p_0) \leq \exp(-\frac{c}{2} \log n) \leq n^{-c/2} . \quad (2.26) \]

By the construction of the composite hypothesis, we have
\[ \left| \mathbb{E} f_j(\mu, \nu, \theta) - \mathbb{E} f_j(\mu, \nu, \theta) \right| \]
\[ = (2^{-dJ})^{\beta_{\gamma \alpha}} \left| \mathbb{E} \left[ \frac{1}{2^{dJ}} \sum_{k \in [2^{dJ}]} |\theta_k| \right] - \mathbb{E} \left[ \frac{1}{2^{dJ}} \sum_{k \in [2^{dJ}]} |\theta_k| \right] \right| \]
\[ = n^{-\beta_{\gamma \alpha}} \left| \mathbb{E} \left[ |\theta| \right] - \mathbb{E} \left[ |\theta| \right] \right| \]
\[ \geq n^{-\beta_{\gamma \alpha}} \cdot 2K^{-1} \tau \approx n^{-\beta_{\gamma \alpha}} \cdot \frac{\log \log(n)}{\log(n)} . \]
Denote $D_n$ to be the collection of data $(Y_1, \cdots, Y_n)$, which is drawn from the distribution $Pr(y^{\otimes n}|\theta) := \prod_{i=1}^{n} p_{\nu_i}(y_i)$. For any functional of $\theta$, and for any estimator based on $n$-i.i.d. samples, we know

$$
\sup_{\nu_0} \mathbb{E}_{D_n \sim Pr(y^{\otimes n}|\theta)} |\hat{T}_n - F(\theta)| \geq \mathbb{E}_{\theta \sim Q_0} \mathbb{E}_{D_n \sim Pr(y^{\otimes n}|\theta)} |\hat{T}_n - F(\theta)| 
\geq \mathbb{E}_{\theta \sim Q_0} \mathbb{E}_{D_n \sim Pr(y^{\otimes n}|\theta)} |\hat{T}_n - \mathbb{E}_{\theta \sim Q_0} F(\theta)| - \delta_{Q_0}
$$

where $\delta_{Q_0} := \mathbb{E}_{\theta \sim Q_0} |\mathbb{E}_{\theta \sim Q_0} F(\theta) - F(\theta)|$. Here $Q_0$ is some prior distribution on $\theta$. Repeat the same argument for $Q_1$, and by Le Cam’s argument on two composite hypothesis

$$
\sup_{\nu_0} \mathbb{E}|\hat{T}_n - F(\theta)| \geq \frac{1}{2} \left( \mathbb{E}_{\theta \sim Q_0} |\hat{T}_n - F(\theta)| + \mathbb{E}_{\theta \sim Q_1} |\hat{T}_n - F(\theta)| \right) - \frac{\delta_{Q_0} + \delta_{Q_1}}{2}
= \frac{1}{2} \left( \mathbb{E}_{D_n \sim p_0} |\hat{T}_n - F(\theta)| + \mathbb{E}_{D_n \sim p_1} |\hat{T}_n - F(\theta)| \right) - \frac{\delta_{Q_0} + \delta_{Q_1}}{2}
\geq \frac{1}{4} \left( \mathbb{E}_{\theta \sim Q_0} F(\theta) - \mathbb{E}_{\theta \sim Q_1} F(\theta) \right) (p_0(T = 1) + p_1(T = 0)) - \frac{\delta_{Q_0} + \delta_{Q_1}}{2}
\geq \frac{1}{4} \left( \mathbb{E}_{\theta \sim Q_0} F(\theta) - \mathbb{E}_{\theta \sim Q_1} F(\theta) \right) \int p_0(y^{\otimes n}) \wedge p_1(y^{\otimes n}) dy^{\otimes n} - \frac{\delta_{Q_0} + \delta_{Q_1}}{2}
= \frac{1}{4} \left(1 - d_{TV}(p_0, p_1)\right) - \frac{\delta_{Q_0} + \delta_{Q_1}}{2}
$$

where the posterior distribution $p_i(y^{\otimes n}) = \int Pr(y^{\otimes n}|\theta)Q_i(d\theta)$, for $i = 0, 1$. Here the test $T = 1$ if and only if $\hat{T}_n$ is closer to $\mathbb{E}_{\theta \sim Q_1} F(\theta)$. In our case,

$$
F(\theta) := d_{\mathcal{F}_y}(\mu, \nu_0) = (2^{-d/4}) \frac{1}{2^{df}} \sum_{k \in [2^d]} |\theta_k|,
$$

hence we know

$$
| \mathbb{E}_{\theta \sim Q_0} F(\theta) - \mathbb{E}_{\theta \sim Q_1} F(\theta)| = | \mathbb{E}_{\nu_0 \sim p_0} d_{\mathcal{F}_y}(\mu, \nu_0) - \mathbb{E}_{\nu_0 \sim p_1} d_{\mathcal{F}_y}(\mu, \nu_0)| 
\geq n^{-\frac{\beta + \gamma}{2\beta + \gamma}} \cdot \frac{\log \log(n)}{\log(n)}
\geq 1 - d_{TV}(p_0, p_1) \geq 1 - n^{-c/2} \text{ by (2.26)}
\geq \frac{\delta_{Q_0} + \delta_{Q_1}}{2} \leq n^{-\frac{\beta + \gamma}{2\beta + \gamma}} \cdot \frac{1}{\sqrt{2d}} \leq n^{-\frac{\beta + \gamma}{2\beta + \gamma}} \cdot \frac{\log \log(n)}{\log(n)} .
$$

Therefore we have

$$
\inf \sup_{\hat{T}_n, \nu \in \mathcal{C}_\beta} \mathbb{E} |\hat{T}_n - d_{\mathcal{F}_y}(\mu, \nu)| \geq n^{-\frac{\beta + \gamma}{2\beta + \gamma}} \cdot \frac{\log \log(n)}{\log(n)} . \tag{2.27}
$$

### 2.2 Proof of the Upper Bound

The upper bound can be obtained through similar derivations as in Liang (2018); Singh et al. (2018); Weed and Berthet (2019); Singh and Páczos (2018). We include here for completeness.
The estimator is of the plug-in form, with

\[ d_{\gamma}(\bar{\mu}_m, \bar{\nu}_n) := \sup_{f \in \mathcal{F}_\gamma} \int f d\bar{\mu}_m - \int f d\bar{\nu}_n \]  

(2.28)

where \( \bar{\mu}_m \) and \( \bar{\nu}_n \) are smoothed empirical measures based on truncation on Wavelets. It is clear that

\[ |d_{\gamma}(\bar{\mu}_m, \bar{\nu}_n) - d_{\gamma}(\mu, \nu)| \leq \sup_{f \in \mathcal{F}_\gamma} \int f d\bar{\mu}_m - \int f d\mu + \sup_{f \in \mathcal{F}_\gamma} \int f d\bar{\nu}_n - \int f d\nu. \]  

(2.29)

Now let’s bound \( \sup_{f \in \mathcal{F}_\gamma} \int f d\bar{\nu}_n - \int f d\nu \) via expanding under the Wavelet basis. Denote \( \bar{\nu}[h_{jk}]:=1/n \sum_{i=1}^n h_{jk}(Y_i) \), the smoothed empirical estimate \( \bar{\nu}_n \) is defined as

\[ \frac{d\bar{\nu}_n}{dx} := \sum_{j=0}^l \sum_{k=0}^{2^j-1} \bar{\nu}[h_{jk}]h_{jk}(x), \]  

(2.30)

with \( f \) to be specified later. Expand \( f(x) = \sum_{j \geq 0} \sum_{k=0}^{2^j-1} f_{jk}h_{jk}(x) \), we have

\[ \sup_{f \in \mathcal{F}_\gamma} |\int f d\bar{\nu}_n - \int f d\nu| \leq \sup_{f \in \mathcal{B}_\infty^{\gamma,\infty}(1)} |\int f d\bar{\nu}_n - \int f d\nu| \]

\[ = \sup_{f \in \mathcal{B}_\infty^{\gamma,\infty}(1)} \sum_{j \geq 0} \sum_{k=0}^{2^j-1} f_{jk}(\bar{\nu}[h_{jk}] - \nu[h_{jk}]) + \sup_{f \in \mathcal{B}_\infty^{\gamma,\infty}(1)} \sum_{j > J} \sum_{k=0}^{2^j-1} f_{jk} \nu[h_{jk}] \]

For the first term, since \( f \in \mathcal{B}_\infty^{\gamma,\infty}(1) \Rightarrow \forall j, k, |f_{jk}| \leq (2^{-d_j})^{\gamma + \frac{1}{2}} \)

\[ \mathbb{E} \sup_{f \in \mathcal{B}_\infty^{\gamma,\infty}(1)} \sum_{j \geq 0} \sum_{k=0}^{2^j-1} f_{jk}(\bar{\nu}[h_{jk}] - \nu[h_{jk}]) \leq \sum_{j \geq 0} (2^{-d_j})^{\gamma + \frac{1}{2}} \sum_{k=0}^{2^j-1} \mathbb{E} \nu[h_{jk}] - \mathbb{E}[h_{jk}] \]

\[ \leq \sum_{j \geq 0} (2^{-d_j})^{\gamma + \frac{1}{2}} 2^{dj} 1 - \frac{1}{\sqrt{n}} \times \frac{1}{\sqrt{n}} (2^{dj})^{\frac{1}{2} - \frac{1}{2}} \]

for \( d \geq 2\gamma \). Here the last inequality follows since \( \mathbb{E} \nu[h_{jk}] - \mathbb{E}[h_{jk}]^2 \leq \mathbb{E}[h_{jk}^2]/n \times 2^{2dj(1/2 - 1/2)}/n \times 1/n \) due to Appendix C in Weed and Berthet (2019).

For the second term, recall \( \mathbb{E}_{Y \sim \nu}[h_{jk}(Y)] = \langle d\nu/dx, h_{jk} \rangle =: v_{jk} \). Due to the fact that

\[ d\nu/dx \in \mathcal{B}_\infty^{\beta,\infty} \Rightarrow \forall j, k, |v_{jk}| \leq (2^{-d_j})^{\gamma + \frac{1}{2}} \]  

(3.31)

\[ f \in \mathcal{B}_\infty^{\gamma,\infty} \Rightarrow \forall j, k, |f_{jk}| \leq (2^{-d_j})^{\gamma + \frac{1}{2}} \]  

(3.32)
Balancing the two terms, we have
\[
\sup_{\nu \in \mathcal{G}} \mathbb{E} \sup_{f \in \mathcal{F}_\gamma} | \int f \, d\tilde{\nu} - \int f \, d\nu | \leq (2^{dJ} - \beta + \gamma d) \frac{1}{n^{1/2}} - \gamma d \leq (2^{dJ} - \beta + \gamma d) \frac{1}{n^{1/2}}.
\]

Put everything together, we know
\[
\mathbb{E} | d_{\mathcal{F}_\gamma}(\tilde{\mu}_m, \tilde{\nu}_n) - d_{\mathcal{F}_\gamma}(\mu, \nu) | \leq (n \wedge m)^{-\beta/2} \frac{1}{\sqrt{n}}.
\]

\textbf{References}


