A Generalized Bayesian Approach to Distribution-on-Distribution Regression (Supplementary Material)

A PROOFS

A.1 PROOF OF THEOREM 1

We first state a lemma which is a direct implication of Lemma 3.6 of Ghodrati and Panaretos [2022]. For $\boldsymbol{\eta} \in \mathbb{R}^{K+1}$, we define

$$T_{\boldsymbol{\eta}}(x) := \sum_{k=0}^{M} \eta_k G_{B(k,K-1+k)}(x).$$

Note that we have extended the definition of the map $T_{\boldsymbol{\eta}}$ from Θ to $\mathbb{R}^{K+1}$.

Lemma 1. *For $\epsilon > 0$, we have the following expansion of the expected risk around $\boldsymbol{\theta}_1$:*

$$R(\boldsymbol{\theta}_1 + \epsilon\boldsymbol{\eta}) = R(\boldsymbol{\theta}_1) + \epsilon D_{\boldsymbol{\eta}} R(\boldsymbol{\theta}_1) + \frac{\epsilon^2}{2}||T_{\boldsymbol{\eta}}||^2_{L^2(Q)}$$

where

$$D_{\boldsymbol{\eta}} R(\boldsymbol{\theta}_1) = \int_{\mathcal{W}_2(\Omega)\times\mathcal{W}_2(\Omega)} \int_0^1 (T_{\boldsymbol{\theta}}(F_\mu^{-1}(p)) - F_\nu^{-1}(p)) T_{\boldsymbol{\eta}}(F_\mu^{-1}(p)) dp dP(\mu,\nu)$$

is the directional derivative of $R(\boldsymbol{\theta}_1)$ in the direction of $\boldsymbol{\eta}$.

Proof of Theorem 1. We apply Theorem 3.2 of Syring and Martin [2023] to derive the stated contraction rate.

We need to show that the loss function $\ell_{\boldsymbol{\theta}}$ satisfies the sub-exponential condition:

There exists an interval $(0, \bar{\omega})$ and constant $K > 0$ such that for all $\omega \in (0, \bar{\omega})$ and for all sufficiently small $\delta > 0$, for $\boldsymbol{\theta} \in \Theta$,

$$||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||_{L^2(Q)} > \delta \implies Pe^{-\omega(\ell_{\boldsymbol{\theta}} - l_{\boldsymbol{\theta}_0})} < e^{-K\omega\delta^2}. \tag{1}$$

We also need to show that the prior Π puts sufficient amount of mass on "neighborhood" G_n of the true parameter $\boldsymbol{\theta}_0$:

$$\log \Pi(G_n) \gtrsim -n\epsilon_n^2, \tag{2}$$

where G_n is defined as

$$G_n := \{\boldsymbol{\theta} \in \Theta : u(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \le \epsilon_n^2, v(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \le \epsilon_n^2\}, \quad n = 1, 2, \dots,$$

and $u(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$ and $v(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$ are the mean and variance of excess risk:

$$u(\boldsymbol{\theta}, \boldsymbol{\theta}_0) := \frac{1}{2} P(d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\mu, \nu) - d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}_0}\#\mu, \nu)) = R(\boldsymbol{\theta}) - R(\boldsymbol{\theta}_0),$$

and

$$v(\boldsymbol{\theta},\boldsymbol{\theta}_0) := P\bigg(\bigg(\frac{1}{2}d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\mu,\nu) - \frac{1}{2}d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}_0}\#\mu,\nu)\bigg)^2\bigg) - u(\boldsymbol{\theta},\boldsymbol{\theta}_0)^2.$$

We first show that the sub-exponential condition (1) is satisfied. By compactness of Ω, we have that for all $\boldsymbol{\theta}\in\Theta$ and all (μ,ν) in the support of P,

$$\ell_{\boldsymbol{\theta}}(\mu,\nu) - \ell_{\boldsymbol{\theta}_0}(\mu,\nu) < C,$$

for some constant $C > 0$. Thus, by Section 3.4.1 of Syring and Martin [2023],

$$Pe^{-\omega(\ell_{\boldsymbol{\theta}} - l_{\boldsymbol{\theta}_0})} \leq \exp\big(-\omega u(\boldsymbol{\theta},\boldsymbol{\theta}_0) + C\omega^3 v(\boldsymbol{\theta},\boldsymbol{\theta}_0)\big),$$

for ω small enough.

Now, consider $\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\Theta$, and let $\boldsymbol{\eta} = \boldsymbol{\theta}_2 - \boldsymbol{\theta}_1$. By Lemma 1, we have

$$R(\boldsymbol{\theta}_1 + \epsilon\boldsymbol{\eta}) = R(\boldsymbol{\theta}_1) + \epsilon D_{\boldsymbol{\eta}}R(\boldsymbol{\theta}_1) + \frac{\epsilon^2}{2}||T_{\boldsymbol{\eta}}||^2_{L^2(Q)}$$

where $D_{\boldsymbol{\eta}}R(\boldsymbol{\theta}_1)$ is the directional derivative of $R(\boldsymbol{\theta}_1)$ in the direction of $\boldsymbol{\eta}$.

For any $\boldsymbol{\theta}\in\Theta$, let $\boldsymbol{\eta} = \boldsymbol{\theta} - \boldsymbol{\theta}_0$, applying the expansion above with $\epsilon = 1$, we have

$$R(\boldsymbol{\theta}) - R(\boldsymbol{\theta}_0) = D_{\boldsymbol{\eta}}R(\boldsymbol{\theta}_0) + \frac{1}{2}||T_{\boldsymbol{\eta}}||^2_{L^2(Q)}.$$

Since $\boldsymbol{\theta}_0$ is the minimizer of R, we have $D_{\boldsymbol{\eta}}R(\boldsymbol{\theta}_0) = 0$, and thus

$$u(\boldsymbol{\theta},\boldsymbol{\theta}_0) = R(\boldsymbol{\theta}) - R(\boldsymbol{\theta}_0) = \frac{1}{2}||T_{\boldsymbol{\eta}}||^2_{L^2(Q)} = \frac{1}{2}||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||^2_{L^2(Q)}. \tag{3}$$

We also have that

$$\begin{aligned}
|d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}_1}\#\mu,\nu) - d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}_2}\#\mu,\nu)| &\lesssim |d_{\mathcal{W}}(T_{\boldsymbol{\theta}_1}\#\mu,\nu) - d_{\mathcal{W}}(T_{\boldsymbol{\theta}_2}\#\mu,\nu)| \\
&\leq d_{\mathcal{W}}(T_{\boldsymbol{\theta}_1}\#\mu, T_{\boldsymbol{\theta}_2}\#\mu) \\
&= ||T_{\boldsymbol{\theta}_1} - T_{\boldsymbol{\theta}_2}||_{L^2(\mu)} \\
&\lesssim ||T_{\boldsymbol{\theta}_1} - T_{\boldsymbol{\theta}_2}||_{L^2(Q)}.
\end{aligned}$$

where the second inequality follows from triangle inequality, and the equality follows from that $\mathcal{W}_2(\Omega)$ is flat. Therefore, it follows that

$$v(\boldsymbol{\theta},\boldsymbol{\theta}_0) \lesssim ||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||^2_{L^2(Q)} \lesssim u(\boldsymbol{\theta},\boldsymbol{\theta}_0). \tag{4}$$

Combining (3) and (4), We obtain

$$\begin{aligned}
Pe^{-\omega(\ell_{\boldsymbol{\theta}} - l_{\boldsymbol{\theta}_0})} &\leq \exp\big(-C_1\omega u(\boldsymbol{\theta},\boldsymbol{\theta}_0)\big) \\
&= \exp\bigg(-\frac{1}{2}C_1\omega||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||^2_{L^2(Q)}\bigg)
\end{aligned}$$

for some constant $C_1 > 0$. It follows that $||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||_{L^2(Q)} > \delta$ implies that

$$Pe^{-\omega(\ell_{\boldsymbol{\theta}} - l_{\boldsymbol{\theta}_0})} \leq \exp\bigg(-\frac{1}{2}C_1\omega\delta^2\bigg),$$

and Condition (1) is verified.

We now verify the prior mass condition (2). We note that our prior specification satisfies

$$\Pi(\{||\boldsymbol{\theta} - \boldsymbol{\theta}_0||_2 \leq \delta) \gtrsim \delta^{K+1},$$

where $||\cdot||_2$ is the 2-norm on $\mathbb{R}^{K+1}$. Since $||\boldsymbol{\theta}-\boldsymbol{\theta}_0||_2 \le \delta$ implies $||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||_{L^2(Q)} \lesssim \delta$, it follows that

$$\Pi(\{||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||_{L^2(Q)} \le \delta) \gtrsim \delta^{K+1}.$$

Since $||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||_{L^2(Q)} \le \delta$ implies $\{u(\boldsymbol{\theta},\boldsymbol{\theta}_0) \lesssim \delta^2, v(\boldsymbol{\theta},\boldsymbol{\theta}_0) \lesssim \delta^2\}$, we have

$$\Pi(G_n) \gtrsim \Pi(\{\boldsymbol{\theta} : ||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||_{L^2(Q)} \le \epsilon_n\}) \gtrsim \epsilon_n^{K+1}.$$

Therefore, with $\epsilon_n = n^{-1/2}(\log n)^{1/2}$, we have

$$\log \Pi(G_n) \gtrsim -\log n \gtrsim -n\epsilon_n^2.$$

Thus, the prior mass condition is satisfied, and the proof is completed. $\square$

A.2 PROOF OF THEOREM 2

For each $m = 1, 2, \dots$, we define

$$\tilde{R}^m(\boldsymbol{\theta}) := \frac{1}{2}\int_{\mathcal{W}_2(\Omega)\times\mathcal{W}_2(\Omega)} d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \hat{\nu}^m)dP(\mu,\nu). \tag{5}$$

Also define the mean and variance of excess risk as

$$u_m(\boldsymbol{\theta},\boldsymbol{\theta}_0) := \frac{1}{2}P(d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \hat{\nu}^m) - d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}_0}\#\hat{\mu}^m, \hat{\nu}^m)) = \tilde{R}^m(\boldsymbol{\theta}) - \tilde{R}^m(\boldsymbol{\theta}_0),$$

and

$$v_m(\boldsymbol{\theta},\boldsymbol{\theta}_0) := P\left(\left(\frac{1}{2}d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \hat{\nu}^m) - \frac{1}{2}d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}_0}\#\hat{\mu}^m, \hat{\nu}^m)\right)^2\right) - u_m(\boldsymbol{\theta},\boldsymbol{\theta}_0)^2.$$

We first prove the following lemma bounding $u_m(\boldsymbol{\theta},\boldsymbol{\theta}_0)$ and $v_m(\boldsymbol{\theta},\boldsymbol{\theta}_0)$ in terms of $||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||^2_{L^2(Q)}$ and r_m^{-1} .

Lemma 2.

$$||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||^2_{L^2(Q)} - r_m^{-1} \lesssim u_m(\boldsymbol{\theta},\boldsymbol{\theta}_0) \lesssim ||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||^2_{L^2(Q)} + r_m^{-1},$$
$$v_m(\boldsymbol{\theta},\boldsymbol{\theta}_0) \lesssim ||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||^2_{L^2(Q)} + r_m^{-2}.$$

Proof. We first have the following decomposition of $u_m(\boldsymbol{\theta},\boldsymbol{\theta}_0)$:

$$\tilde{R}^m(\boldsymbol{\theta}) - \tilde{R}^m(\boldsymbol{\theta}_0) = \underbrace{\tilde{R}^m(\boldsymbol{\theta}) - R(\boldsymbol{\theta})} + \underbrace{R(\boldsymbol{\theta}) - R(\boldsymbol{\theta}_0)} + \underbrace{R(\boldsymbol{\theta}_0) - \tilde{R}^m(\boldsymbol{\theta}_0)}. \tag{6}$$

We bound each of the three terms on the RHS of (6).

$$\begin{aligned}
\tilde{R}^m(\boldsymbol{\theta}) - R(\boldsymbol{\theta}) &= P\Big(d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \hat{\nu}^m) - d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\mu, \nu)\Big) \\
&= P\Big(d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \hat{\nu}^m) - d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \nu) + d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \nu) - d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\mu, \nu)\Big).
\end{aligned}$$

We have that

$$\begin{aligned}
& P\Big(d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \hat{\nu}^m) - d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \nu)\Big) \\
= \; & P\Big(\big(d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \hat{\nu}^m) - d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \nu)\big)\big(d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \hat{\nu}^m) + d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \nu)\big)\Big) \\
\ge \; & P\big(-d_{\mathcal{W}}(\hat{\nu}^m, \nu)\big(d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \hat{\nu}^m) + d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \nu)\big)\big) \\
\gtrsim \; & P(-d_{\mathcal{W}}(\hat{\nu}^m, \nu)) \\
\gtrsim \; & -r_m^{-1},
\end{aligned}$$

where the first inequality follows from the reverse triangle inequality, and the last inequality follows from our assumption. We also have that

$$\begin{aligned}
& P\Big(d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \hat{\nu}^m) - d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \nu)\Big) \\
= \; & P\Big(\big(d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \hat{\nu}^m) - d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \nu)\big)\big(d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \hat{\nu}^m) + d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \nu)\big)\Big) \\
\leq \; & P\Big(d_{\mathcal{W}}(\hat{\nu}^m, \nu)\big(d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \hat{\nu}^m) + d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \nu)\big)\Big) \\
\lesssim \; & r_m^{-1}
\end{aligned}$$

by an application of the triangle inequality.

Similarly, we can show that

$$-r_m^{-1} \lesssim P\Big(d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \nu) - d^2_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\mu, \nu)\Big) \lesssim r_m^{-1}.$$

It follows that the first term on the RHS of (6) can be bounded as

$$-r_m^{-1} \lesssim \tilde{R}^m(\boldsymbol{\theta}) - R(\boldsymbol{\theta}) \lesssim r_m^{-1}.$$

Using the same calculation, we also have the bound for the third term on the RHS of (6):

$$-r_m^{-1} \lesssim R(\boldsymbol{\theta}_0) - \tilde{R}^m(\boldsymbol{\theta}_0) \lesssim r_m^{-1}.$$

For the second term on the RHS of (6), we recall from the proof of Theorem 1 that

$$R(\boldsymbol{\theta}) - R(\boldsymbol{\theta}_0) = \frac{1}{2}||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||^2_{L^2(Q)}.$$

Thus, we obtain the following bound for $u_m(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$:

$$||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||^2_{L^2(Q)} - r_m^{-1} \lesssim u_m(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \lesssim ||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||^2_{L^2(Q)} + r_m^{-1}.$$

Now we try to obtain the upper bound for $v_m(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$. By triangle inequality,

$$\begin{aligned}
& |d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \hat{\nu}^m) - d_{\mathcal{W}}(T_{\boldsymbol{\theta}_0}\#\hat{\mu}^m, \hat{\nu}^m)| \\
\leq \; & |d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \hat{\nu}^m) - d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\mu, \nu)| + |d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\mu, \nu) - d_{\mathcal{W}}(T_{\boldsymbol{\theta}_0}\#\mu, \nu)| \\
& + |d_{\mathcal{W}}(T_{\boldsymbol{\theta}_0}\#\mu, \nu) - d_{\mathcal{W}}(T_{\boldsymbol{\theta}_0}\#\hat{\mu}^m, \hat{\nu}^m)|
\end{aligned}$$

Similar calculations as above lead to

$$|d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \hat{\nu}^m) - d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\mu, \nu)| \lesssim r_m^{-1}$$

and

$$|d_{\mathcal{W}}(T_{\boldsymbol{\theta}_0}\#\mu, \nu) - d_{\mathcal{W}}(T_{\boldsymbol{\theta}_0}\#\hat{\mu}^m, \hat{\nu}^m)| \lesssim r_m^{-1}.$$

We also have that

$$\begin{aligned}
|d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\mu, \nu) - d_{\mathcal{W}}(T_{\boldsymbol{\theta}_0}\#\mu, \nu)| & \leq d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\mu, T_{\boldsymbol{\theta}_0}\#\mu) \\
& = ||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||_{L^2(\mu)} \\
& \lesssim ||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||_{L^2(Q)}.
\end{aligned}$$

Since

$$v_m(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \lesssim P(|d_{\mathcal{W}}(T_{\boldsymbol{\theta}}\#\hat{\mu}^m, \hat{\nu}^m) - d_{\mathcal{W}}(T_{\boldsymbol{\theta}_0}\#\hat{\mu}^m, \hat{\nu}^m)|^2),$$

it follows that

$$v_m(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \lesssim ||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||^2_{L^2(Q)} + r_m^{-2}.$$

□

We are now in a position to prove Theorem 2.

Proof of Theorem 2. Let $A_n := \{\boldsymbol{\theta} \in \Theta : ||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||_{L^2(Q)} > M\epsilon_n\}$. The Gibbs posterior probability of A_n is given by

$$\begin{aligned}\Pi_n(A_n) &= \frac{N_n^m(A_n)}{D_n^m} \\ &= \frac{\int_{A_n} \exp\big(-\omega n\big(\tilde{R}_n^m(\boldsymbol{\theta}) - \tilde{R}_n^m(\boldsymbol{\theta}_0)\big)\big)\Pi(d\boldsymbol{\theta})}{\int_{\Theta} \exp\big(-\omega n\big(\tilde{R}_n^m(\boldsymbol{\theta}) - \tilde{R}_n^m(\boldsymbol{\theta}_0)\big)\big)\Pi(d\boldsymbol{\theta})}.\end{aligned}$$

Note that $m(n)$ is assumed to be a deterministic function of n. We aim to show that $P^n\Pi_n(A_n) \to 0$ as $n \to \infty$. We define the set

$$G_n^m := \{\boldsymbol{\theta} \in \Theta : u_m(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \leq \epsilon_n^2, v_m(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \leq \epsilon_n^2\}.$$

By Lemma 2 and the assumption $\epsilon_n^2 > r_m^{-1}$, G_n^m is implied by the event

$$H_n^m := \{\boldsymbol{\theta} \in \Theta : ||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||^2_{L^2(Q)} \leq c(\epsilon_n^2 - r_m^{-1})\},$$

for some constant $c > 0$. We thus have

$$\Pi(G_n^m) \geq \Pi(H_n^m) \gtrsim \epsilon_n^{K+1},$$

from which it follows that

$$\log \Pi(G_n^m) \gtrsim -n\epsilon_n^2.$$

Since the excess loss

$$\ell_{\boldsymbol{\theta}}(\hat{\mu}^m, \hat{\nu}^m) - \ell_{\boldsymbol{\theta}_0}(\hat{\mu}^m, \hat{\nu}^m)$$

is bounded for all $\boldsymbol{\theta}$ and $(\hat{\mu}^m, \hat{\nu}^m)$, when

$$||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||^2_{L^2(Q)} > \epsilon_n,$$

we apply Section 3.4.1 of Syring and Martin [2023] to obtain

$$\begin{aligned}Pe^{-n\omega(\tilde{R}_n^m(\boldsymbol{\theta}) - \tilde{R}_n^m(\boldsymbol{\theta}_0))} &\leq \exp\Big(-nc_0\omega\Big(||T_{\boldsymbol{\theta}} - T_{\boldsymbol{\theta}_0}||^2_{L^2(Q)} - r^{-1}_{m(n)} - r^{-2}_{m(n)}\Big)\Big) \\ &\leq \exp\big(-nc_1\omega(\epsilon_n^2 - r_m^{-1})\big) \\ &\leq \exp\big(-nc_2\omega\epsilon_n^2\big)\end{aligned}$$

for some constants $c_0, c_1, c_2 > 0$. By Fubini's Theorem, we have

$$P^n N_n(A_n) = \int_{A_n} Pe^{-\omega n(\tilde{R}_n^m(\boldsymbol{\theta}) - \tilde{R}_n^m(\boldsymbol{\theta}_0))}\Pi(d\boldsymbol{\theta}) \leq \exp\Big(-nc_2\omega M^2\epsilon_n^2\Big).$$

Following essentially the same lines as the proof of Lemma 1 of Syring and Martin [2023], we obtain

$$P^n\left(D_n^m > \frac{1}{2}\Pi(G_n^m)e^{-2\omega n\epsilon_n^2}\right) \to 1,$$

as $n \to \infty$.

Let $b_n^m = \frac{1}{2}\Pi(G_n^m)e^{-2\omega n\epsilon_n^2}$, we have

$$P^n(D_n^m \leq b_n^m) \to 0$$

as $n \to \infty$. Since

$$\begin{aligned}\Pi_n(A_n) &\leq \frac{N_n^m(A_n)}{D_n^m}1(D_n^m > b_n^m) + 1(D_n^m \leq b_n^m) \\ &\leq b_n^{-1}N_n^m(A_n) + 1(D_n^m \leq b_n^m)\end{aligned}$$

It follows that

$$P^n\Pi_n(A_n) \to 0$$

as $n \to \infty$. The proof is completed.

□

B ADDITIONAL SIMULATION STUDIES

In order to evaluate the robustness of our proposed model and posterior sampling strategy, we conduct additional simulation studies aimed at investigating its behavior under mis-specification. Specifically, we replicate the simulation settings described in the main article, wherein the true optimal transport map is generated using BP basis functions with a polynomial order of $K = 50$. However, in the model fitting process, we set the polynomial order to $K = 20$. The outcomes of these experiments are illustrated in Figures 1, 2, and 3.

Upon examination of the results, we observe that despite the mis-specification in the model fitting, the true optimal transport maps are successfully recovered in all scenarios. This suggests that our proposed model and posterior sampling strategy exhibit robustness to mis-specification, demonstrating their effectiveness in capturing underlying patterns even when the model assumptions are not entirely met.

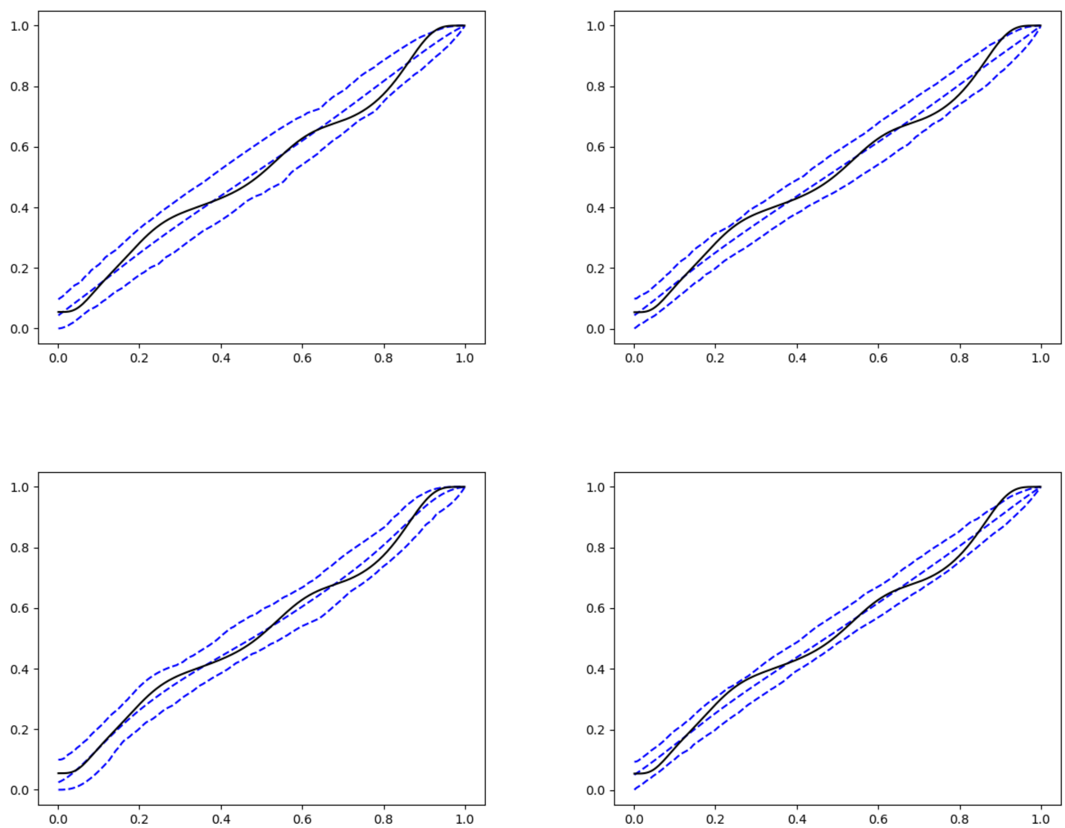

Figure 1: Simulation 1. Top left: n=5 ($\omega = 500$). Top right: n=20 ($\omega = 200$). Bottom left: n=50 ($\omega = 100$). Bottom right: n=100 ($\omega = 50$). Black curve: True optimal transport map. Blue dashed curves: estimated posterior mean and posterior credible intervals of optimal transport map.

References

Laya Ghodrati and Victor M Panaretos. Distribution-on-distribution regression via optimal transport maps. *Biometrika*, 109 (4):957–974, 01 2022.

Nicholas Syring and Ryan Martin. Gibbs posterior concentration rates under sub-exponential type losses. *Bernoulli*, 29(2): 1080 – 1108, 2023.

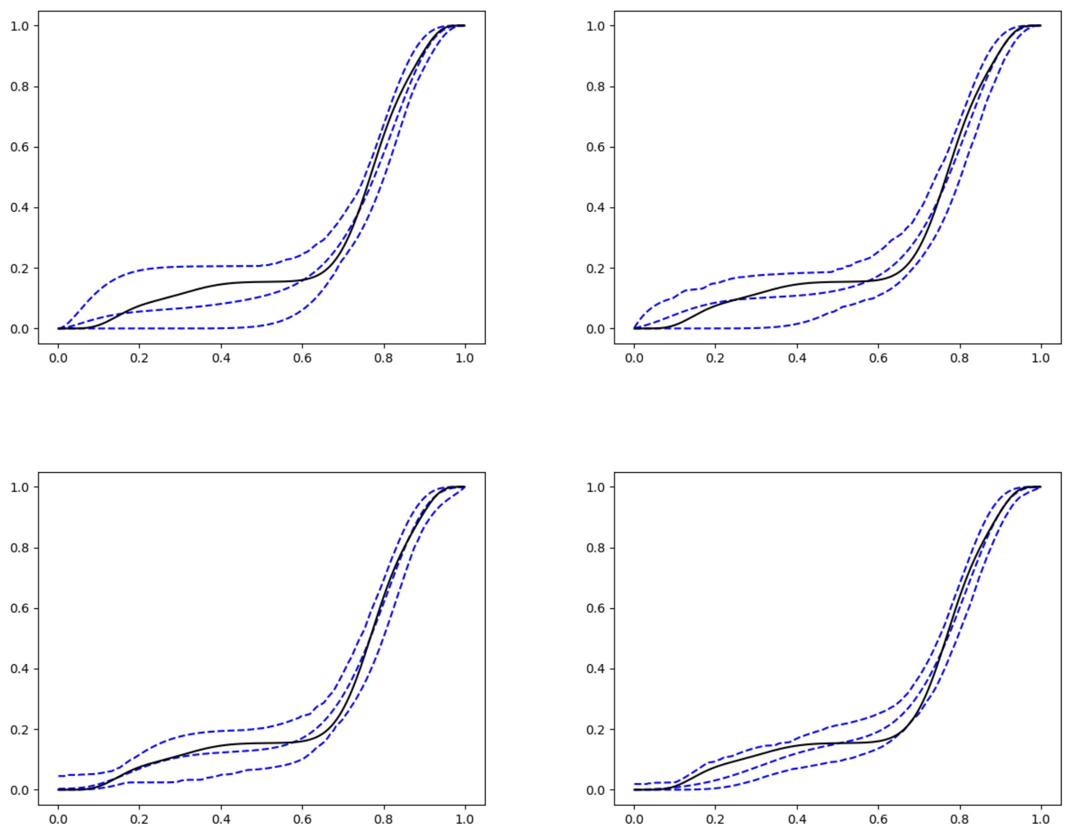

Figure 2: Simulation 2. Top left: n=5 ($\omega = 500$). Top right: n=20 ($\omega = 200$). Bottom left: n=50 ($\omega = 100$). Bottom right: n=100 ($\omega = 50$). Black curve: True optimal transport map. Blue dashed curves: estimated posterior mean and posterior credible intervals of optimal transport map.

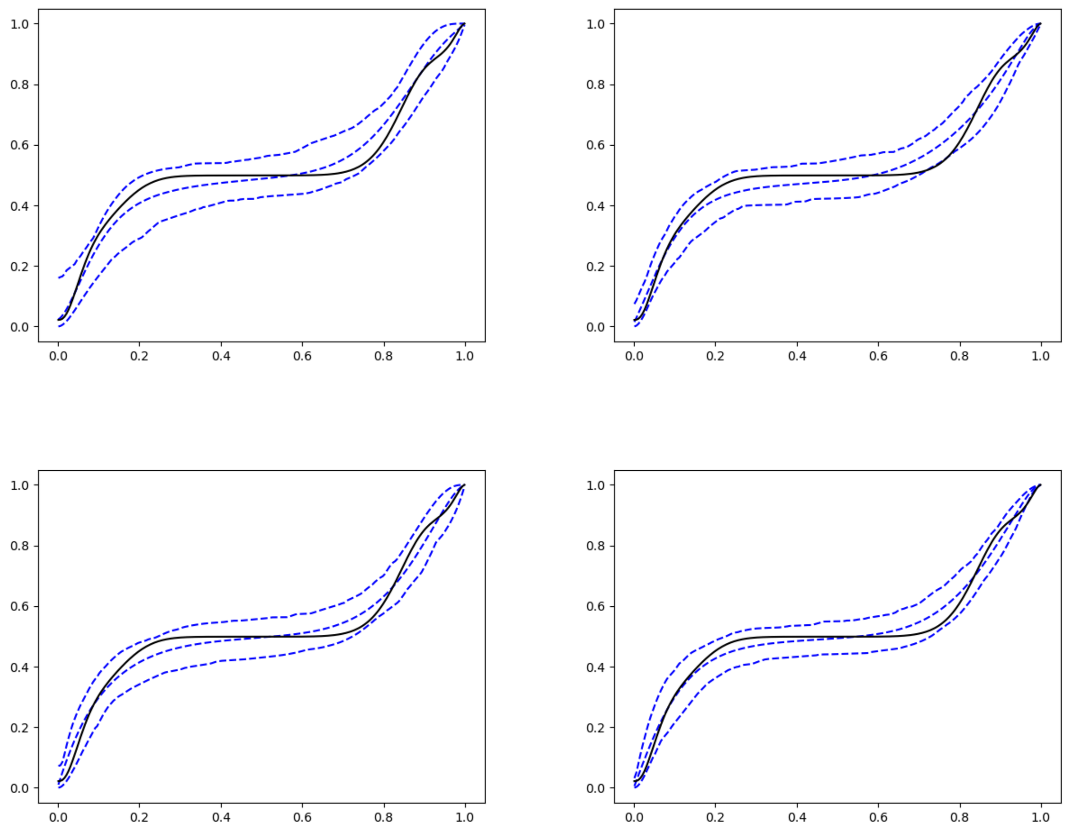

Figure 3: Simulation 3. Top left: n=5 ($\omega = 500$). Top right: n=20 ($\omega = 200$). Bottom left: n=50 ($\omega = 100$). Bottom right: n=100 ($\omega = 50$). Black curve: True optimal transport map. Blue dashed curves: estimated posterior mean and posterior credible intervals of optimal transport map.