Supplementary Material for "Wasserstein Hypothesis Test for Group Fairness"

Nian Si Karthyek Murthy Jose Blanchet Viet Anh Nguyen

Appendix A Proofs

We prove Theorems 1 and 2 under a more general assumption below.

Assumption 2′ (Continuous conditional measure)**.** *For the case where supp(U) is potentially an infinite set, the cost function c is decomposable as*

$$c\left((x,u),(x',u')\right) = \bar{c}(x,x') + \infty \cdot \|u-u'\|,$$

and the following conditions are satisfied:

a) *the moments $\mathbb{E}_{\mathbb{P}}\|U\|_2^2$, $\mathbb{E}_{\mathbb{P}}\|\phi(U,\mu)\|_2^2$ and $\mathbb{E}_{\mathbb{P}}\|\phi_z(U,\mu)\|_2$ are finite.*

b) *For z such that $\|z-\mu\|_2 < v$, the derivative $\phi_z(\cdot)$ satisfies,*

$$\|\phi_z(u,z) - \phi_z(u,\mu)\|_2 \le M(u)\,\|z-\mu\|_2\,, \tag{A.1}$$

where $\mathbb{E}_{\mathbb{P}}[M(U)] < +\infty$.

c) *The (regular) conditional probability measure ν_t of $\phi(U,\mu)|\Phi(X)=t$ converges in terms of the type 1-Wasserstein distance as $t \to 0$: i.e., there exist a set $B \subset \mathbb{R}$ with $\mathbb{P}(\Phi(X)\in B)=1$ and $\varepsilon_0 > 0$ such that*

$$\lim_{t\to 0} W_1\left(\nu_t, \nu_0\right)\mathbf{1}\{t\in B\} = 0$$

and $\sup_{t\in B}\mathbb{E}_{\mathbb{P}}[\|\phi(U,\mu)\|_2^{2+\epsilon_0}|\Phi(X)=t]$ is finite, where type 1-Wasserstein distance $W_1(\cdot,\cdot)$ is $W_c(\cdot,\cdot)$ with the cost function being a metric.

Remark 1. If supp(U) is a finite set, and U is completely dependent on (X,Y), the simpler Assumption 2 is equivalent to Assumption 2′.

Appendix A.1 Proofs of Section 3

Proof of Proposition 1. Since the cost to move U is $+\infty$, we have $\mathbb{E}_{\mathbb{Q}}[U] = \mathbb{E}_{\hat{\mathbb{P}}^N}[U]$. Then, consider any probability measure $\mathbb{Q}$ such that

$$\mathbb{E}_{\mathbb{Q}}[\mathcal{C}(X)\phi(U, \mathbb{E}_{\hat{\mathbb{P}}^N}[U])] = 0,$$

and let π be the optimal coupling between $\hat{\mathbb{P}}^N$ and $\mathbb{Q}$. Because $\hat{\mathbb{P}}^N$ is the empirical measure, the coupling π can be written as $\pi = \frac{1}{N}\sum_{i\in[N]} \pi_i \otimes \delta_{(x_i,u_i)}$. For any value $\varepsilon > 0$, construct now the measure

$$\mathbb{Q}^\varepsilon = \frac{1}{N}\sum_{i\in[N]} (1-p_i)\delta_{(x_i,u_i)} + p_i\delta_{(x_i^\varepsilon,u_i)}, \tag{A.2}$$

where the mass p_i is set to

$$p_i = \int_{\mathbb{X}_{1-\mathcal{C}(x_i)}} \pi_i(\mathrm{d}x) = \pi_i(\mathbb{X}_{1-\mathcal{C}(x_i)}) \in [0,1] \qquad \forall i \in [N].$$

and x_i^ε is an ε-optimizer of the problem $\inf_{x'\in\mathbb{X}_{1-\mathcal{C}(x)}} c(x_i, x')$. Then, it is easy to see

$$\begin{aligned}
&\mathbb{E}_{\mathbb{Q}^\varepsilon}[\mathcal{C}(X)\phi(U, \mathbb{E}_{\hat{\mathbb{P}}^N}[U])] = \mathbb{E}_{\mathbb{Q}}[\mathcal{C}(X)\phi(U, \mathbb{E}_{\hat{\mathbb{P}}^N}[U])] = 0, \text{ and}\\
&\mathbb{E}_{\mathbb{Q}^\varepsilon}[\mathcal{C}(X)\phi(U, \mathbb{E}_{\hat{\mathbb{P}}^N}[U])]\\
&= \frac{1}{N}\left(\sum_{i\in[N]} (1-p_i)\mathcal{C}(x_i)\phi(u_i, \mathbb{E}_{\hat{\mathbb{P}}^N}[U]) + p_i(1-\mathcal{C}(x_i))\phi(u_i, \mathbb{E}_{\hat{\mathbb{P}}^N}[U])\right)\\
&= \frac{1}{N}\left(\sum_{i\in[N]} (1-2\mathcal{C}(x_i))\phi(u_i, \mathbb{E}_{\hat{\mathbb{P}}^N}[U])p_i + \sum_{i\in[N]} \mathcal{C}(x_i)\phi(u_i, \mathbb{E}_{\hat{\mathbb{P}}^N}[U])\right).
\end{aligned}$$

Since $d(x_i) \le \|x_i^\varepsilon - x_i\| \le \|x - x_i\| + \varepsilon$ for any $x \in \mathbb{X}_{1-\mathcal{C}(x_i)}$, this implies that $\frac{1}{N}\sum_{i\in[N]} p_i d(x_i) \le W(\mathbb{Q}, \hat{\mathbb{P}}^N) + \varepsilon$. Since ε can be chosen arbitrarily, this implies that

$$\mathcal{P}(\hat{\mathbb{P}}^N) \ge \left\{\begin{array}{ll} \min & \frac{1}{N}\sum_{i\in[N]} p_i d(x_i)\\ \text{s.t.} & p \in [0,1]^N\\ & \sum_{i\in[N]} (1-2\mathcal{C}(x_i))\phi(u_i, \mathbb{E}_{\hat{\mathbb{P}}^N}[U])p_i = -\sum_{i\in[N]} \mathcal{C}(x_i)\phi(u_i, \mathbb{E}_{\hat{\mathbb{P}}^N}[U]). \end{array}\right. \tag{A.3}$$

On the other hand, for any $\{p_i\}_{i=1}^N$ satisfying the constraints in the linear programming (A.3), we can construct the measure $\mathbb{Q}^\varepsilon$ according to (A.2). Since $\mathbb{Q}^\varepsilon$ is a feasible solution of the primal problem (5), we have the other direction of the inequality.

□

Proof of Proposition 2. **Case 1:** $d(x_i) < +\infty$ for $i \in [N]$. The primal problem has a feasible solution

($p_i = 0$ if $\mathcal{C}(X) = 0$; $p_i = 1$, otherwise) and is bounded, thus it has an optimal solution and the strong duality holds. By the strong duality, we have

$$\mathcal{P}(\hat{\mathbb{P}}^N) = \left\{ \begin{array}{ll} \max & \dfrac{1}{N} \left\{ \displaystyle\sum_{i\in[N]} \alpha^i + \gamma^\top \mathcal{C}(x_i)\phi(u_i, \mathbb{E}_{\hat{\mathbb{P}}^N}[U]) \right\} \\ \text{s.t.} & \alpha_i \le 0 \quad \forall i \in [N] \\ & \alpha_i - (1 - 2\mathcal{C}(x_i))\,\gamma^\top \phi(u_i, \mathbb{E}_{\hat{\mathbb{P}}^N}[U]) \le d(x_i) \quad \forall i \in [N]. \end{array} \right. \tag{A.4}$$

Then, we have

$$\alpha_i = \Big(d(x_i) + (1 - 2\mathcal{C}(x_i))\,\gamma^\top \phi(u_i, \mathbb{E}_{\hat{\mathbb{P}}^N}[U]) \Big)^-,$$

which gives the desired results.

Case 2: $\exists i \in [N]$ such that $d(x_i) = +\infty$. The primal problem is equivalent to

$$\mathcal{P}(\hat{\mathbb{P}}^N) = \left\{ \begin{array}{ll} \min & \dfrac{1}{N} \displaystyle\sum_{i\in[N]} p_i d(x_i), \\ \text{s.t.} & p \in [0,1]^N, \\ & p_i = 0 \text{ for } d(x_i) = +\infty \\ & \displaystyle\sum_{i\in[N]} (1 - 2\mathcal{C}(x_i))\phi(u_i, \mathbb{E}_{\hat{\mathbb{P}}^N}[U]) p_i = -\sum_{i\in[N]} \mathcal{C}(x_i)\phi(u_i, \mathbb{E}_{\hat{\mathbb{P}}^N}[U]), \end{array} \right. \tag{A.5}$$

with the convention that if the problem is infeasible, the optimal value of the minimization problem is $+\infty$. We have the dual problem

$$\begin{array}{ll} \max & \dfrac{1}{N} \left\{ \displaystyle\sum_{d(x_i)<+\infty} \alpha^i + \gamma^\top \mathcal{C}(x_i)\phi(u_i, \mathbb{E}_{\hat{\mathbb{P}}^N}[U]) \right\} \\ \text{s.t.} & \alpha_i \le 0 \quad \text{for } d(x_i) < +\infty \\ & \alpha_i - (1 - 2\mathcal{C}(x_i))\,\gamma^\top \phi(u_i, \mathbb{E}_{\hat{\mathbb{P}}^N}[U]) \le d(x_i) \quad \text{for } d(x_i) < +\infty. \end{array} \tag{A.6}$$

Since the problem (A.6) is also feasible, if it is bounded, then the strong duality holds. If the problem (A.6) is unbounded, the primal problem (A.5) is infeasible, which means the primal and the dual both have optimal value $+\infty$. Finally, because

$$\Big(d(x_i) + (1 - 2\mathcal{C}(x_i))\,\gamma^\top \phi(u_i, \mathbb{E}_{\hat{\mathbb{P}}^N}[U]) \Big)^- = 0 \text{ for } d(x_i) = +\infty,$$

we have the optimal value of problem (A.6) equals to $\mathcal{D}(\hat{\mathbb{P}}^N)$. □

Proof of Lemma 1. Since $X_C \big| X_D = x$ has positive density in $\mathbb{R}^{d_2}$ for every $x \in \mathbb{D}$, we have $\theta_C^\top X_C \big| X_D =$

x has positive density in $\mathbb{R}$ for every $x \in \mathbb{D}$. Therefore, $\theta^\top X$ has a density

$$f_{\theta^\top X}(x) = \sum_{v \in \mathbb{D}} p_v f_{\theta_C^\top X_C \big| v}(x - \theta_D^\top v) > 0,$$

where $p_v = \mathbb{P}(X_D = v)$ and $f_{\theta_C^\top X_C \big| v}(\cdot)$ denotes the conditional density of $\theta_C^\top X_C \big| X_D = x$.

Further, let $w = \ell^{-1}(\tau)$. For the cost function $\bar{c}(\cdot)$ given by (3a), we have by Hölder inequality

$$d(x) = \inf_{\theta^\top x' = w} \|x - x'\| = \|\theta\|_*^{-1} |\theta^\top x - w|.$$

Therefore, $\mathbb{P}_\Phi$ has a continuous density $f(\cdot) = \|\theta\|_* f_{\theta^\top X}(\|\theta\|_* \times \cdot + w)$ with $f(0) > 0$.

For the cost function $\bar{c}(\cdot)$ given by (4), when $d(x) < \delta$, we have

$$d(x) = \inf_{\theta^\top x' = w} \bar{c}(x, x') = \inf_{x_D = x'_D, \theta_C^\top x' = w - \theta_D^\top x_D} \bar{c}(x, x') = \inf_{\theta_C^\top x' = w - \theta_D^\top x_D} \|x_C - x'_C\| = \|\theta_C\|_*^{-1} |\theta^\top x - w|.$$

The last equality is again due to Hölder inequality. Therefore, $\mathbb{P}_\Phi$ has a continuous density $f(\cdot) = \|\theta_C\|_* f_{\theta^\top X}(\|\theta_C\|_* \times \cdot + w)$ with $f(0) > 0$, which completes the proof. □

Lemmas A1 and A2 are useful for the proof of Theorem 1, whose proofs are presented in Section Appendix A.3.

Lemma A1. *Suppose Assumption 2′ is enforced. Then, we have*

$$\begin{aligned} & \sqrt{N} \left(\mathbb{E}_{\hat{\mathbb{P}}^N} \left[\phi \left(U, \mathbb{E}_{\hat{\mathbb{P}}^N}[U] \right) \mathcal{C}(X) \right] - \mathbb{E}_{\mathbb{P}} \left[\phi(U, \mu) \mathcal{C}(X) \right] \right) \\ \Rightarrow \quad & \mathcal{N}(0, \operatorname{cov} \left(\mathbb{E}_{\mathbb{P}} \left[\phi_z(U, \mu) \mathcal{C}(X) \right] U + \phi(U, \mu) \mathcal{C}(X) \right)). \end{aligned}$$

Lemma A2. *Suppose Assumption 1 and 2′ are enforced. Then, we have*

$$\begin{aligned} & \sqrt{N} \mathbb{E}_{\hat{\mathbb{P}}^N} \left[\left(-\gamma^\top \phi \left(U, \mathbb{E}_{\hat{\mathbb{P}}^N}[U] \right) + \sqrt{N} d(X) \right)^- \mathcal{C}(X) \right] \\ \xrightarrow{p} & -\frac{1}{2} f(0) \mathbb{E}_{\mathbb{P}} \left[\left(\gamma^\top \phi(U, \mu) \right)^2 \mathbb{I}\{ \gamma^\top \phi(U, \mu) \geq 0 \} \Big| d(X) = 0 \right], \end{aligned}$$

uniformly over $\|\gamma\|_2 \leq B$.

We are now ready to prove Theorem 1.

Proof of Theorem 1. Recall that

$$\mathcal{D}(\hat{\mathbb{P}}^N) \quad = \quad \max_{\gamma \in \mathbb{R}^m} \frac{1}{N} \left\{ \sum_{i \in [N]} \left(d(x_i) + (1 - 2\mathcal{C}(x_i)) \gamma^\top \phi(u_i, \mathbb{E}_{\hat{\mathbb{P}}^N}[U]) \right)^- + \gamma^\top \mathcal{C}(x_i) \phi(u_i, \mathbb{E}_{\hat{\mathbb{P}}^N}[U]) \right\}$$

$$
\begin{aligned}
&= \max_{\gamma\in\mathbb{R}^m}\left\{\frac{1}{N}\sum_{i\in[N]}\gamma^\top\mathcal{C}(x_i)\phi(u_i,\mathbb{E}_{\hat{\mathbb{P}}^N}[U])+\right.\\
&\left.\frac{1}{N}\sum_{i\in[N]}\left(d(x_i)-\gamma^\top\phi(u_i,\mathbb{E}_{\hat{\mathbb{P}}^N}[U])\right)^-\mathcal{C}(x_i)+\left(d(x_i)+\gamma^\top\phi(u_i,\mathbb{E}_{\hat{\mathbb{P}}^N}[U])\right)^-(1-\mathcal{C}(x_i))\right\}.
\end{aligned}
$$

We first rescale $\gamma\leftarrow\gamma\sqrt{N}$ and thus

$$
\begin{aligned}
&N\mathcal{D}(\hat{\mathbb{P}}^N)\\
=\;&\sqrt{N}\max_{\gamma\in\mathbb{R}^m}\left\{\gamma^\top\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\phi\left(U,\mathbb{E}_{\hat{\mathbb{P}}^N}[U]\right)\mathcal{C}(X)\right]+\right.\\
&\left.\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\left(\sqrt{N}d(X)-\gamma^\top\phi(U,\mathbb{E}_{\hat{\mathbb{P}}^N}[U])\right)^-\mathcal{C}(X)+\left(\sqrt{N}d(X)+\gamma^\top\phi(U,\mathbb{E}_{\hat{\mathbb{P}}^N}[U])\right)^-(1-\mathcal{C}(X))\right]\right\}.
\end{aligned}
$$

To ease the notation, we denote $\lambda_i=\phi\left(u_i,\mathbb{E}_{\hat{\mathbb{P}}^N}[U]\right)$. By Lemma A2, we have

$$
\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^N\left(-\gamma^\top\lambda_i+N^{1/2}d(x_i)\right)^-(1-\mathcal{C}(x_i))\\
\xrightarrow{p}\;&-\frac{1}{2}f(0)\mathbb{E}_{\mathbb{P}}\left[\left(\gamma^\top\phi\left(U,\mu\right)\right)^2\mathbb{I}\{\gamma^\top\phi\left(U,\mu\right)\geq 0\}\middle|\,d(X)=0\right]\\
=\;&-\frac{1}{2}f(0)\mathbb{E}_{\mathbb{P}}\left[\left(\gamma^\top\phi\left(U,\mu\right)\right)^2\mathbb{I}\{\gamma^\top\phi\left(U,\mu\right)\geq 0\}\middle|\,d(X)=0\right],
\end{aligned}
$$

and similarly, we have

$$
\frac{1}{\sqrt{N}}\sum_{i=1}^N\left(\gamma^\top\lambda_i+N^{1/2}d(x_i)\right)^-(1-\mathcal{C}(x_i))\xrightarrow{p}-\frac{1}{2}f(0)\mathbb{E}_{\mathbb{P}}\left[\left(\gamma^\top\phi\left(U,\mu\right)\right)^2\mathbb{I}\{\gamma^\top\phi\left(U,\mu\right)<0\}\middle|\,d(X)=0\right].
$$

Therefore, we have

$$
\begin{aligned}
&\sqrt{N}\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\left(\sqrt{N}d(X)-\gamma^\top\phi(U,\mathbb{E}_{\hat{\mathbb{P}}^N}[U])\right)^-\mathcal{C}(X)+\left(\sqrt{N}d(X)+\gamma^\top\phi(U,\mathbb{E}_{\hat{\mathbb{P}}^N}[U])\right)^-(1-\mathcal{C}(X))\right]\\
\xrightarrow{p}\;&-\frac{1}{2}f(0)\mathbb{E}_{\mathbb{P}}\left[\left(\gamma^\top\phi\left(U,\mu\right)\right)^2\middle|\,d(X)=0\right].
\end{aligned}
$$

We denote

$$
\begin{aligned}
V_N &= \sqrt{N}\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\phi\left(U,\mathbb{E}_{\hat{\mathbb{P}}^N}[U]\right)\mathcal{C}(X)\right],\ \text{and}\\
M_N(\gamma) &= \frac{1}{\sqrt{N}}\sum_{i=1}^N\left[\left(-\gamma^\top\lambda_i+N^{1/2}d(x_i)\right)^-\mathcal{C}(X)\}+\left(\gamma^\top\lambda_i+N^{1/2}d(x_i)\right)^-(1-\mathcal{C}(X))\right].
\end{aligned}
$$

To proceed, we rely on the following lemma.

Lemma A3. *Suppose Assumption 1 is enforced. Then, for every $\varepsilon>0$, there exists $N_0>0$ and*

$b \in (0, \infty)$ *such that for all* $N \geq N_0$,

$$\mathbb{P}\left(\sup_{\|\gamma\|_2 > b} \left\{ \gamma^\top V_N + M_N(\gamma) \right\} > 0 \right) \leq \varepsilon.$$

The proof of Lemma A3 is furnished in Section Appendix A.3. Notice that $\mathcal{D}(\hat{\mathbb{P}}^N) \geq 0$ (choosing $\gamma = 0$), Lemma A3 implies that when $N \geq N_0$,

$$\mathbb{P}\left\{ N\mathcal{D}(\hat{\mathbb{P}}^N) = \sup_{\|\gamma\|_2 \leq b} \left\{ \gamma^\top V_N + M_N(\gamma) \right\} \right\} \geq 1 - \varepsilon.$$

By Lemmas A1 and A2, we have

$$\begin{aligned} \sup_{\|\gamma\|_2 \leq b} \left\{ \gamma^\top V_N + M_N(\gamma) \right\} \quad &\Rightarrow \quad \sup_{\|\gamma\|_2 \leq b} \left\{ \gamma^\top V - \frac{1}{2} f(0) \mathbb{E}\left[\left(\gamma^\top \phi\left(U, \mu\right) \right)^2 \middle| d(X) = 0 \right] \right\} \\ &= \quad \sup_{\|\gamma\|_2 \leq b} \left\{ \gamma^\top V - \frac{1}{2} \gamma^\top S \gamma \right\}, \end{aligned}$$

where

$$S = f(0) \mathbb{E}_{\mathbb{P}}\left[\phi\left(U, \mu\right) \phi\left(U, \mu\right)^\top \middle| d(X) = 0 \right],$$

and V is normally distributed with mean zero and covariance matrix

$$\operatorname{cov}\left(\mathbb{E}_{\mathbb{P}}\left[\phi_z(U, \mu) \mathcal{C}(X) \right] U + \phi(U, \mu) \mathcal{C}(X) \right).$$

By the arbitrariness of ε, we have the desired result:

$$N \times \mathcal{D}(\hat{\mathbb{P}}^N) \Rightarrow \sup_{\gamma} \left\{ \gamma^\top V - \frac{1}{2} \gamma^\top S \gamma \right\}.$$

This completes the proof. □

Appendix A.2 Proofs of Section 4

The proofs of Propositions 4 and 5 are not presented because they follow the same lines as the proofs of Propositions 1 and 2.

Proof of Theorem 2. Let

$$\epsilon^* = \mathbb{E}_{\mathbb{P}}\left[\mathcal{C}(X) \phi\left(U, \mathbb{E}_{\mathbb{P}}[U]\right) \right].$$

By following the similar arguments with the proof of Theorem 1, we have

$N\mathcal{D}_\epsilon(\hat{\mathbb{P}}^N)$

$$
\begin{aligned}
&= N \sup_{\gamma\in\mathbb{R}^m_+}\left\{-\gamma^\top\epsilon+\frac{1}{N}\left\{\sum_{i\in[N]}(1-2\mathcal{C}(x_i))\phi(u_i,\mathbb{E}_{\hat{\mathbb{P}}^N}[U])+\sum_{i\in[N]}\mathcal{C}(x_i)\phi(u_i,\mathbb{E}_{\hat{\mathbb{P}}^N}[U])\right\}\right\}\\
&= \sqrt{N}\sup_{\gamma\in\mathbb{R}^m_+}\Big\{\gamma^\top\left(\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\phi\left(U,\mathbb{E}_{\hat{\mathbb{P}}^N}[U]\right)\mathcal{C}(X)\right\}-\epsilon\right])+\\
&\quad \mathbb{E}_{\hat{\mathbb{P}}^N}\left[\left(-\gamma^\top\phi(U,\mathbb{E}_{\hat{\mathbb{P}}^N}[U])+\sqrt{N}d(X)\right)^-\mathcal{C}(X)+\left(\gamma^\top\phi(U,\mathbb{E}_{\hat{\mathbb{P}}^N}[U])+\sqrt{N}d(X)\right)^-(1-\mathcal{C}(X))\right]\Big\}\\
&= \sup_{\gamma\in\mathbb{R}^m_+}\left\{\gamma^\top V_N+\sqrt{N}\gamma^\top(\epsilon^*-\epsilon)+M_N(\gamma)\right\}.
\end{aligned}
$$

Similarly, we still have

$$\gamma^\top V_N+M_N(\gamma)\Rightarrow\gamma^\top V-\frac{1}{2}\gamma^\top S\gamma,$$

uniformly over $\left\{\gamma:\gamma\in\mathbb{R}^m_+,\|\gamma\|_2\le B\right\}$. Therefore, we must enforce $\gamma^\top(\epsilon^*-\epsilon)=0$ here. Then, we have

$$N\mathcal{D}_\epsilon(\hat{\mathbb{P}}^N)\Rightarrow\max_{\gamma\in\mathbb{R}^m_+,\gamma^\top(\epsilon^*-\epsilon)=0}\left\{\gamma^\top V-\frac{1}{2}\gamma^\top S\gamma\right\}\preceq\max_{\gamma\in\mathbb{R}^m_+}\left\{\gamma^\top V-\frac{1}{2}\gamma^\top S\gamma\right\}.$$

This completes the proof. □

Appendix A.3 Proofs of Technical Results

Proof of Lemma A1. By adding and subtracting the term $\mathbb{E}_{\hat{\mathbb{P}}^N}[\phi(U,\mu)\mathcal{C}(X)]$, we find

$$
\begin{aligned}
&\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\phi(U,\mathbb{E}_{\hat{\mathbb{P}}^N}[U])\mathcal{C}(X)\right]-\mathbb{E}_{\mathbb{P}}\left[\phi(U,\mu)\mathcal{C}(X)\right]\\
=\;&\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\phi(U,\mathbb{E}_{\hat{\mathbb{P}}^N}[U])\mathcal{C}(X)-\phi(U,\mu)\mathcal{C}(X)\right]\\
&\quad+\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\phi(U,\mu)\mathcal{C}(X)\right]-\mathbb{E}_{\mathbb{P}}\left[\phi(U,\mu)\mathcal{C}(X)\right]
\end{aligned}
\tag{A.7}
$$

Under Assumption 2′ and the fundamental theorem of calculus, the first term in the right-hand side of (A.7) becomes

$$
\begin{aligned}
&\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\phi(U,\mathbb{E}_{\hat{\mathbb{P}}^N}[U])\mathcal{C}(X)-\phi(U,\mu)\mathcal{C}(X)\right]\\
=\;&\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\int_0^1\phi_z\left(U,\mu+t\left(\mathbb{E}_{\hat{\mathbb{P}}^N}[U]-\mu\right)\right)\left(\mathbb{E}_{\hat{\mathbb{P}}^N}[U]-\mu\right)\mathcal{C}(X)\mathrm{d}t\right].
\end{aligned}
$$

Thanks to Assumption 2′, we have that

$$
\begin{aligned}
&\left\|\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\int_0^1\phi_z\left(U,\mu+t\left(\mathbb{E}_{\hat{\mathbb{P}}^N}[U]-\mu\right)\right)\left(\mathbb{E}_{\hat{\mathbb{P}}^N}[U]-\mu\right)\mathcal{C}(X)\mathrm{d}t\right]\right.\\
&\qquad\left.-\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\int_0^1\phi_z\left(U,\mu\right)\left(\mathbb{E}_{\hat{\mathbb{P}}^N}[\psi(U)]-\mu\right)\mathcal{C}(X)\mathrm{d}t\right]\right\|_2\\
\le\;&\frac{1}{2}\mathbb{E}_{\hat{\mathbb{P}}^N}\left[M(U)\right]\left\|\mathbb{E}_{\hat{\mathbb{P}}^N}[U]-\mu\right\|_2^2,
\end{aligned}
$$

whenever $\left\|\mathbb{E}_{\hat{\mathbb{P}}^N}[U]-\mu\right\|_2<\varepsilon_\mu$. Then, notice that we have

$$\lim_{N\to\infty}\frac{1}{2}\sqrt{N}\mathbb{E}_{\hat{\mathbb{P}}^N}\left[M(U)\right]\left\|\mathbb{E}_{\hat{\mathbb{P}}^N}[\psi(U)]-\mu\right\|_2^2=0 \text{ almost surely,} \tag{A.8}$$

and

$$\begin{aligned}
&\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\int_0^1\phi_z\left(U,\mu\right)\left(\mathbb{E}_{\hat{\mathbb{P}}^N}[U]-\mu\right)\mathcal{C}(X)\mathrm{d}t\right]\\
=\;&\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\phi_z\left(U,\mu\right)\mathcal{C}(X)\right]\left(\mathbb{E}_{\hat{\mathbb{P}}^N}[U]-\mu\right)\}\\
=\;&\left(\mathbb{E}_{\mathbb{P}}\left[\phi_z\left(U,\mu\right)\mathcal{C}(X)\right]+o_p\left(1\right)\right)\left(\mathbb{E}_{\hat{\mathbb{P}}^N}[\psi(U)]-\mu\right).
\end{aligned}$$

By multiplying $\sqrt{N}$ to both sides of equation (A.7), we have

$$\begin{aligned}
&\sqrt{N}\left(\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\phi\left(U,\mathbb{E}_{\hat{\mathbb{P}}^N}[U]\right)\mathcal{C}(X)\right]-\mathbb{E}_{\mathbb{P}}\left[\phi(U,\mu)\mathcal{C}(X)\right]\right)\\
=\;&\sqrt{N}\left(\mathbb{E}_{\mathbb{P}}\left[\phi_z\left(U,\mu\right)\mathcal{C}(X)\right]+o_p\left(1\right)\right)\left(\mathbb{E}_{\hat{\mathbb{P}}^N}[U]-\mu\right)\\
&+\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\phi(U,\mu)\mathcal{C}(X)\right]-\mathbb{E}_{\mathbb{P}}\left[\phi(U,\mu)\mathcal{C}(X)\right]+o_p(1)\\
=\;&\sqrt{N}\mathbb{E}_{\hat{\mathbb{P}}^N}\Big[\mathbb{E}_{\mathbb{P}}\big[\phi_z\left(U,\mu\right)\mathcal{C}(X)\big]\left(U-\mu\right)+\phi(U,\mu)\mathcal{C}(X)-\mathbb{E}_{\mathbb{P}}\big[\phi(U,\mu)\mathcal{C}(X)\big]\Big]+o_p(1)\\
\Rightarrow\;&\mathcal{N}(0,\Sigma),
\end{aligned}$$

where Σ is the covariance matrix of $\mathbb{E}_{\mathbb{P}}\left[\phi_z(U,\mu)\mathcal{C}(X)\right]U+\phi(U,\mu)\mathcal{C}(X)$, namely

$$\Sigma=\mathrm{cov}\left(\mathbb{E}_{\mathbb{P}}\left[\phi_z(U,\mu)\mathcal{C}(X)\right]U+\phi(U,\mu)\mathcal{C}(X)\right).$$

□

Proof of Lemma A2. **Step 1:** we first show

$$\begin{aligned}
&\sqrt{N}\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\left(-\gamma^\top\phi\left(U,\mathbb{E}_{\hat{\mathbb{P}}^N}[U]\right)+\sqrt{N}d(X)\right)^-\mathcal{C}(X)\right]\\
&-\sqrt{N}\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\left(-\gamma^\top\phi\left(U,\mu\right)+\sqrt{N}d(X)\right)^-\mathcal{C}(X)\right]\xrightarrow{p}0,
\end{aligned}$$

uniformly over $\|\gamma\|_2\le B$. When $\left\|\mathbb{E}_{\hat{\mathbb{P}}^N}[U]-\mu\right\|_2<\varepsilon_\mu$, we have

$$\begin{aligned}
&\sqrt{N}\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\left(-\gamma^\top\phi\left(U,\mathbb{E}_{\hat{\mathbb{P}}^N}[U]\right)+\sqrt{N}d(X)\right)^-\mathcal{C}(X)\right]\\
&-\sqrt{N}\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\left(-\gamma^\top\phi\left(U,\mu\right)+\sqrt{N}d(X)\right)^-\mathcal{C}(X)\right]\\
\le\;&N^{-1/2}\left\|\gamma\right\|_2\sum_{i=1}^N\left[\left\|\phi\left(u_i,\mathbb{E}_{\hat{\mathbb{P}}^N}[\psi(U)]\right)-\phi\left(u_i,\mu\right)\right\|_2\mathbb{I}\{\mathcal{E}_i\}\right],
\end{aligned}$$

where the events $\mathcal{E}_i$ are defined by

$$\mathcal{E}_i = \{ \|\gamma\|_2 \left(\|\phi\left(u_i, \mu\right)\|_2 + \left(\|\phi_z\left(u_i, \mu\right)\|_2 + M(u_i)\varepsilon_\mu \right) \varepsilon_\mu \right) \geq \sqrt{N} d(x_i) \}.$$

By a similar derivation with the proof of Lemma A1, we have

$$\begin{aligned}
& N^{-1/2} \|\gamma\|_2 \sum_{i=1}^{N} \left[\left\| \phi\left(u_i, \mathbb{E}_{\hat{\mathbb{P}}^N}[U]\right) - \phi\left(u_i, \mu\right) \right\|_2^- \mathbb{I}\{\mathcal{E}_i\} \right] \\
= & \frac{\|\gamma\|_2}{\sqrt{N}} \sum_{i=1}^{N} \left[\int_0^1 \left(\phi_z\left(u_i, \mu + t\left(\mathbb{E}_{\hat{\mathbb{P}}^N}[U] - \mu\right)\right) \left(\mathbb{E}_{\hat{\mathbb{P}}^N}[U] - \mu\right) \mathrm{d}t \right) \mathbb{I}\{\mathcal{E}_i\} \right] \\
\leq & \|\gamma\|_2 \sqrt{N} \left(\mathbb{E}_{\hat{\mathbb{P}}^N}[U] - \mu\right) \mathbb{E}_{\hat{\mathbb{P}}^N}\left[\phi_z\left(U, \mu\right) \mathbb{I}\{\mathcal{E}_i\}\right] + \frac{1}{2} \|\gamma\|_2 \sqrt{N} \mathbb{E}_{\hat{\mathbb{P}}^N}[M(U)] \left\| \mathbb{E}_{\hat{\mathbb{P}}^N}[U] - \mu \right\|_2^2.
\end{aligned}$$

Since $\mathbb{I}\{\mathcal{E}_i\} \to 0$ almost surely and $\mathbb{E}_{\mathbb{P}}[\phi_z(U, \mu)] < +\infty$, we have

$$\|\gamma\|_2 \sqrt{N} \left(\mathbb{E}_{\hat{\mathbb{P}}^N}[U] - \mu\right) \mathbb{E}_{\hat{\mathbb{P}}^N}\left[\phi_z\left(U, \mu\right) \mathbb{I}\{\mathcal{E}_i\}\right] \xrightarrow{p} 0,$$

uniformly over $\|\gamma\|_2 \leq B$. By combining

$$\frac{1}{2} \|\gamma\|_2 \sqrt{N} \mathbb{E}_{\hat{\mathbb{P}}^N}[M(U)] \left\| \mathbb{E}_{\hat{\mathbb{P}}^N}[U] - \mu \right\|_2^2 \to 0 \text{ almost surely,}$$

uniformly over $\|\gamma\|_2 \leq B$, we finish step 1.

Step 2: We claim that

$$\begin{aligned}
& \sqrt{N} \mathbb{E}_{\mathbb{P}} \left[\left(-\gamma^\top \phi\left(U, \mu\right) + \sqrt{N} d(X) \right)^- \mathcal{C}(X) \right] \\
\longrightarrow & -\frac{1}{2} f(0) \mathbb{E}\left[\left(\gamma^\top \phi\left(U, \mu\right) \right)^2 \mathbb{I}\{\gamma^\top \phi\left(U, \mu\right) \geq 0\} \middle| d(X) = 0 \right].
\end{aligned}$$

Notice that for any $c > 0$, we have

$$\begin{aligned}
& \sqrt{N} \mathbb{E}_{\mathbb{P}} \left[\left(-\gamma^\top \phi\left(U, \mu\right) + \sqrt{N} d(X) \right)^- \mathcal{C}(X) \right] \\
= & \sqrt{N} \int_0^{+\infty} \mathbb{E}_{\mathbb{P}} \left[\left(-\gamma^\top \phi\left(U, \mu\right) + \sqrt{N} d(X) \right)^- \middle| d(X)\left(2\mathcal{C}(X) - 1\right) = t \right] \mathrm{d}\mathbb{P}_\Phi(t). \\
= & \sqrt{N} \int_0^{c/\sqrt{N}} \mathbb{E}_{\mathbb{P}} \left[\left(-\gamma^\top \phi\left(U, \mu\right) + \sqrt{N} d(X) \right)^- \middle| \Phi(X) = t \right] \mathrm{d}\mathbb{P}_\Phi(t) \qquad \text{(A.9)} \\
& + \sqrt{N} \int_{c/\sqrt{N}}^{+\infty} \mathbb{E}_{\mathbb{P}} \left[\left(-\gamma^\top \phi\left(U, \mu\right) + \sqrt{N} d(X) \right)^- \middle| \Phi(X) = t \right] \mathrm{d}\mathbb{P}_\Phi(t) \qquad \text{(A.10)}
\end{aligned}$$

We first analyze the first term in (A.9). By Assumption 1.a), when N is sufficient large such that

$c/\sqrt{N} < v$, we have

$$\begin{aligned}
&\sqrt{N}\int_0^{c/\sqrt{N}} \mathbb{E}_{\mathbb{P}}\left[\left(-\gamma^\top \phi\left(U,\mu\right)+\sqrt{N}d(X)\right)^{-}\middle|\,\Phi(X)=t\right]\mathrm{d}\mathbb{P}_{\Phi}(t)\\
=\;&\sqrt{N}\int_0^{c/\sqrt{N}} \mathbb{E}_{\mathbb{P}}\left[\left(-\gamma^\top \phi\left(U,\mu\right)+\sqrt{N}d(X)\right)^{-}\middle|\,\Phi(X)=t\right] f(t)\mathrm{d}t.
\end{aligned}$$

By changing of the variable $s=\sqrt{N}t$, we have

$$\begin{aligned}
&\sqrt{N}\int_0^{c/\sqrt{N}} \mathbb{E}_{\mathbb{P}}\left[\left(-\gamma^\top \phi\left(U,\mu\right)+\sqrt{N}d_{\mathbb{B}}(X)\right)^{-}\middle|\,\Phi(X)=t\right] f(t)\mathrm{d}t\\
=\;&\int_0^{c} \mathbb{E}_{\mathbb{P}}\left[\left(-\gamma^\top \phi\left(U,\mu\right)+s\right)^{-}\middle|\,\Phi(X)=N^{-1/2}s\right] f(N^{-1/2}s)\mathrm{d}s
\end{aligned}$$

By Assumption 1.c), we have for any $\varepsilon>0$, any $0<c<+\infty$, there exists N_0, such that for $N>N_0$ and $s\le c$,

$$\begin{aligned}
&\left|\mathbb{E}_{\mathbb{P}}\left[\left(-\gamma^\top \phi\left(U,\mu\right)+s\right)^{-}\middle|\,\Phi(X)=N^{-1/2}s\right]-\mathbb{E}_{\mathbb{P}}\left[\left(-\gamma^\top \phi\left(U,\mu\right)+s\right)^{-}\middle|\,\Phi(X)=0\right]\right|\\
\le\;&\left\|\gamma\right\|_*\left\|\mathbb{E}_{\mathbb{P}}\left[\phi\left(U,\mu\right)|\,\Phi(X)=N^{-1/2}s\right]-\mathbb{E}_{\mathbb{P}}\left[\phi\left(U,\mu\right)|\,\Phi(X)=0\right]\right\|\\
\le\;&\left\|\gamma\right\|_* W_1\left(\phi\left(U,\mu\right)|\,\Phi(X)=N^{-1/2}s,\,\phi\left(U,\mu\right)|\,\Phi(X)=0\right)\le\varepsilon.
\end{aligned}$$

Therefore, by taking $\varepsilon\downarrow 0$, we have

$$\begin{aligned}
&\left|\int_0^{c} \mathbb{E}_{\mathbb{P}}\left[\left(-\gamma^\top \phi\left(U,\mu\right)+s\right)^{-}\mathbb{I}\{h\left(X\right)\ge\tau\}\middle|\,\Phi(X)=N^{-1/2}s\right] f(N^{-1/2}s)\mathrm{d}s\right.\\
&\left.-\int_0^{c} \mathbb{E}_{\mathbb{P}}\left[\left(-\gamma^\top \phi\left(U,\mu\right)+s\right)^{-}\mathbb{I}\{h\left(X\right)\ge\tau\}\middle|\,\Phi(X)=0\right] f(N^{-1/2}s)\mathrm{d}s\right|\xrightarrow{p}0.
\end{aligned}$$

Then, the basic algebra and the mean value theorem for integrals give us

$$\begin{aligned}
&\int_0^{c} \mathbb{E}_{\mathbb{P}}\left[\left(-\gamma^\top \phi\left(U,\mu\right)+s\right)^{-}\middle|\,\Phi(X)=0\right] f(N^{-1/2}s)\mathrm{d}s\\
=&\int_0^{c} \mathbb{E}_{\mathbb{P}}\left[\left(-\gamma^\top \phi\left(U,\mu\right)+s\right)^{-}\middle|\,d(X)=0\right] f(N^{-1/2}s)\mathrm{d}s\\
=&f(\xi)\mathbb{E}_{\mathbb{P}}\left[\int_0^{c}\left(\left(-\gamma^\top \phi\left(U,\mu\right)+s\right)^{-}\middle|\,d(X)=0\right)\mathrm{d}s\right]\\
=&f(\xi)\mathbb{E}_{\mathbb{P}}\left[\int_0^{\min(c,\gamma^\top\phi(U,\mu)|\Phi(X)=0)}\left(\left(-\gamma^\top \phi\left(U,\mu\right)+s\right)\mathbb{I}\{\gamma^\top\phi\left(U,\mu\right)\ge 0\}\middle|\,d(X)=0\right)\mathrm{d}s\right]\\
\to&-\frac{1}{2}f(0)\mathbb{E}_{\mathbb{P}}\left[\min\{c,\gamma^\top\phi\left(U,\mu\right)\}\left(\gamma^\top\phi\left(U,\mu\right)+\left(\gamma^\top\phi\left(U,\mu\right)-c\right)^{+}\right)\mathbb{I}\{\gamma^\top\phi\left(U,\mu\right)\ge 0\}\middle|d(X)=0\right],
\end{aligned}\tag{A.11}$$

where $\xi \in [0, N^{-1/2}c]$ and $(x)^+ = \max\{x, 0\}$.

We then deal with the second term in (A.9). Let

$$M_\gamma = \operatorname*{ess\,sup}_{t\geq 0} \mathbb{E}_{\mathbb{P}}\left[\left|\gamma^\top \phi\left(U,\mu\right)\right|^{2+\epsilon_0} |\Phi(X)=t\right].$$

For any $c \geq 0$, we have

$$\begin{aligned}
& \sqrt{N}\int_{c/\sqrt{N}}^{+\infty} \mathbb{E}_{\mathbb{P}}\left[\left(-\gamma^\top \phi\left(U,\mu\right)+\sqrt{N}t\right)^- \middle| \Phi(X)=t\right] \mathrm{d}\mathbb{P}_\Phi(t) \\
\geq\ & -\sqrt{N}\int_{c/\sqrt{N}}^{+\infty} \mathbb{E}_{\mathbb{P}}\left[\gamma^\top \phi\left(U,\mu\right) \mathbb{I}\{\gamma^\top \phi\left(U,\mu\right) \geq \sqrt{N}t\middle| \Phi(X)=t\right] \mathrm{d}\mathbb{P}_\Phi(t) \\
\geq\ & -\sqrt{N}\int_{c/\sqrt{N}}^{+\infty} \left(\frac{1}{\sqrt{N}t}\right)^{1+\epsilon_0} \mathbb{E}_{\mathbb{P}}\left[\left(\gamma^\top \phi\left(U,\mu\right)\right)^{2+\epsilon_0} \mathbb{I}\{\gamma^\top \phi\left(U,\mu\right) \geq \sqrt{N}t\middle| \Phi(X)=t\right] \mathrm{d}\mathbb{P}_\Phi(t) \\
\geq\ & -\left(\sqrt{N}\right)^{-\epsilon_0}\int_{c/\sqrt{N}}^{+\infty} \frac{1}{t^{1+\epsilon_0}} \mathbb{E}_{\mathbb{P}}[\left|\gamma^\top \phi\left(U,\mu\right)\right|^{2+\epsilon_0} |\Phi(X)=t] \mathrm{d}\mathbb{P}_\Phi(t) \\
\geq\ & -\left(\sqrt{N}\right)^{-\epsilon_0} M_\gamma \int_{c/\sqrt{N}}^{+\infty} \frac{1}{t^{1+\epsilon_0}} \mathrm{d}\mathbb{P}_\Phi(t).
\end{aligned}$$

We pick $\varepsilon > 0$ such that $\mathbb{P}_\Phi(\cdot)$ has density in $[0, \varepsilon]$. Then, we have

$$\begin{aligned}
& \left(\sqrt{N}\right)^{-\epsilon_0} M_\gamma \int_{c/\sqrt{N}}^{+\infty} \frac{1}{t^{1+\epsilon_0}} f(t) \mathrm{d}t \\
=\ & M_\gamma \left(\sqrt{N}\right)^{-\epsilon_0} \left(\int_{\varepsilon}^{+\infty} \frac{1}{t^{1+\epsilon_0}} f(t) \mathrm{d}\mathbb{P}_\Phi(t) + \int_{c/\sqrt{N}}^{\varepsilon} \frac{1}{t^{1+\epsilon_0}} f(t) \mathrm{d}t\right) \\
\leq\ & M_\gamma \left(\left(\sqrt{N}\right)^{-\epsilon_0} \frac{1}{\varepsilon^{1+\epsilon_0}} + \frac{1}{\epsilon_0}\left(\sqrt{N}\right)^{-\epsilon_0} \left(\sqrt{N}/c\right)^{\epsilon_0} f(\xi)\right) \\
=\ & M_\gamma \left(\left(\sqrt{N}\right)^{-\epsilon_0} \frac{1}{\varepsilon^{1+\epsilon_0}} + \frac{1}{c^{\epsilon_0}\epsilon_0} f(\xi)\right),
\end{aligned}$$

where $\xi \in (c/\sqrt{N}, \varepsilon)$. By taking $\varepsilon \downarrow 0$, we have

$$\liminf_{N\to+\infty} \sqrt{N}\int_{c/\sqrt{N}}^{+\infty} \mathbb{E}_{\mathbb{P}}\left[\left(-\gamma^\top \phi\left(U,\mu\right)+\sqrt{N}t\right)^- \middle| \Phi(X)=t\right] \mathrm{d}\mathbb{P}_\Phi(t) \geq -\frac{M_\gamma f(0)}{c^{\epsilon_0}\epsilon_0}.$$

Finally, by taking $c \uparrow +\infty$, we conclude step 2.

Step 3: We then apply weak law of triangular arrays Durrett [2, Theorem 2.2.11]. We need to check

$$N \times \left[\mathbb{P}\left(-\left(-\gamma^\top \phi\left(U,\mu\right)+\sqrt{N}d(X)\right)\mathcal{C}(X) > \sqrt{N}\right)\right] \to 0, \text{ and} \tag{A.12a}$$

$$\mathbb{E}\left[\left(\left(-\gamma^\top \phi\left(U,\mu\right)+\sqrt{N}d(X)\right)^-\right)^2 \mathcal{C}(X)\right] \to 0. \tag{A.12b}$$

For condition (A.12a), we have

$$\begin{aligned} & N\mathbb{P}\left(-\left(-\gamma^{\top}\phi\left(U,\mu\right)+\sqrt{N}d(x_i)\right)^{-}\mathcal{C}(X)>\sqrt{N}\right) \\ \leq\ & N\mathbb{P}\left(\gamma^{\top}\phi\left(U,\mu\right)\geq\sqrt{N}\right) \\ \leq\ & \frac{\mathbb{E}\left[\left(\gamma^{\top}\phi\left(U,\mu\right)\right)^{2+\epsilon_0}\right]}{\left(\sqrt{N}\right)^{\epsilon_0}}\leq\frac{M_\gamma}{\left(\sqrt{N}\right)^{\epsilon_0}}\to 0. \end{aligned}$$

For condition (A.12b), we have

$$\begin{aligned} & \mathbb{E}\left[\left(\left(-\gamma^{\top}\phi\left(U,\mu\right)+\sqrt{N}d(X)\right)^{-}\right)^2\mathcal{C}(X)\right] \\ \leq\ & \mathbb{E}\left[\left(\gamma^{\top}\phi\left(U,\mu\right)\right)^2\mathbb{I}\{\gamma^{\top}\phi\left(U,\mu\right)\geq\sqrt{N}d(X)\}\right] \\ =\ & \int_0^{+\infty}\mathbb{E}_{\mathbb{P}}\left[\left(\gamma^{\top}\phi\left(U,\mu\right)\right)^2\mathbb{I}\left\{\gamma^{\top}\phi\left(U,\mu\right)\geq\sqrt{N}t\right\}\middle|\,\Phi(X)=t\right]\mathrm{d}\mathbb{P}_{\Phi}(t). \end{aligned}$$

We pick $\varepsilon>0$ such that $\mathbb{P}_{\Phi}\left(\cdot\right)$ has density in $[0,\varepsilon]$. Then, we have

$$\begin{aligned} & \int_0^{+\infty}\mathbb{E}_{\mathbb{P}}\left[\left(\gamma^{\top}\phi\left(U,\mu\right)\right)^2\mathbb{I}\left\{\gamma^{\top}\phi\left(U,\mu\right)\geq\sqrt{N}t\right\}\middle|\,\Phi(X)=t\right]\mathrm{d}\mathbb{P}_{\Phi}(t) \\ =\ & \int_0^{\varepsilon}\mathbb{E}_{\mathbb{P}}\left[\left(\gamma^{\top}\phi\left(U,\mu\right)\right)^2\mathbb{I}\left\{\gamma^{\top}\phi\left(U,\mu\right)\geq\sqrt{N}t\right\}\middle|\,\Phi(X)=t\right]f(t)\mathrm{d}t \qquad \text{(A.13)} \\ & +\int_{\varepsilon}^{+\infty}\mathbb{E}_{\mathbb{P}}\left[\left(\gamma^{\top}\phi\left(U,\mu\right)\right)^2\mathbb{I}\left\{\gamma^{\top}\phi\left(U,\mu\right)\geq\sqrt{N}t\right\}\middle|\,\Phi(X)=t\right]\mathrm{d}\mathbb{P}_{\Phi}(t). \qquad \text{(A.14)} \end{aligned}$$

For the first term (A.13), we have

$$\int_0^{\varepsilon}\mathbb{E}_{\mathbb{P}}\left[\left(\gamma^{\top}\phi\left(U,\mu\right)\right)^2\mathbb{I}\{\,\gamma^{\top}\phi\left(U,\mu\right)\geq\sqrt{N}t\middle|\,\Phi(X)=t\right]f(t)\mathrm{d}t\leq M_{\gamma}^{2/(2+\epsilon_0)}\varepsilon f(\xi),$$

where $\xi\in[0,\varepsilon]$. For the second term (A.14) we have

$$\begin{aligned} & \int_{\varepsilon}^{+\infty}\mathbb{E}_{\mathbb{P}}\left[\left(\gamma^{\top}\phi\left(U,\mu\right)\right)^2\mathbb{I}\{\,\gamma^{\top}\phi\left(U,\mu\right)\geq\sqrt{N}t\middle|\,\Phi(X)=t\right]\mathrm{d}\mathbb{P}_{\Phi}(t) \\ \leq\ & \int_{\varepsilon}^{+\infty}\frac{\mathbb{E}_{\mathbb{P}}\left[\left(\gamma^{\top}\phi\left(U,\mu\right)\right)^{2+\epsilon_0}\Phi(X)=t\right]}{\left(\sqrt{N}t\right)^{\epsilon_0}}\mathrm{d}\mathbb{P}_{\Phi}(t) \\ \leq\ & \frac{M_0}{\left(\sqrt{N}\varepsilon\right)^{\epsilon_0}}\to 0. \end{aligned}$$

By taking $\varepsilon \downarrow 0$, we have

$$\int_0^{+\infty} \mathbb{E}_{\mathbb{P}}\left[\left(\gamma^\top \phi\left(U,\mu\right)\right)^2 \mathbb{I}\{\gamma^\top \phi\left(U,\mu\right) \geq \sqrt{N}t\Big|\, \Phi(X)=t\right] \mathrm{d}\mathbb{P}_\Phi(t) \to 0.$$

We then apply Durrett [2, Theorem 2.2.11] to obtain the weak law for each γ.

Step 4: We establish the Lipschitz continuity of

$$\sqrt{N}\mathbb{E}_{\mathbb{P}}\left[\left(-\gamma^\top \phi\left(U,\mu\right)+\sqrt{N}d(X)\right)^- \mathcal{C}(X)\right]$$

for $\|\gamma\|_2 \leq B$, which ensures the tightness. For any γ_1, γ_2 satisfying $\|\gamma_1\|_2 \leq B$ and $\|\gamma_2\|_2 \leq B$, we have

$$\begin{aligned}
& \sqrt{N}\left|\mathbb{E}_{\mathbb{P}}\left[\left(-\gamma_1^\top \phi\left(U,\mu\right)+\sqrt{N}d(X)\right)^- \mathcal{C}(X)\right]\right. \\
- \quad & \left.\mathbb{E}_{\mathbb{P}}\left[\left(-\gamma_2^\top \phi\left(U,\mu\right)+\sqrt{N}d(X)\right)^- \mathcal{C}(X)\}\right]\right| \\
\leq \quad & \sqrt{N}\left\|\gamma_1-\gamma_2\right\|_2 \left\|\phi\left(U,\mu\right)\right\|_2 \mathcal{C}(X)\mathbb{I}\left\{B\left\|\phi\left(U,\mu\right)\right\|_2 \geq \sqrt{N}d(X)\right\}.
\end{aligned}$$

By following similar lines with steps 2 and 3, we have

$$\begin{aligned}
& \sqrt{N}\mathbb{E}_{\hat{\mathbb{P}}^N}\left[\left\|\phi\left(U,\mu\right)\right\|_2 \mathcal{C}(X)\mathbb{I}\left\{B\left\|\phi\left(U,\mu\right)\right\|_2 \geq \sqrt{N}d(X)\right\}\right] \\
& \xrightarrow{p} f(0)\mathbb{E}_{\mathbb{P}}\left[\left\|\phi\left(U,\mu\right)\right\|_2^2 \Big| d(X)=0\right].
\end{aligned}$$

Then, by Billingsley [1, Theorem 7.5], we have the desired uniform convergence result. □

Proof of Lemma A3. Due to $\mathbb{E}\left[\phi\left(U,\mu\right)\phi\left(U,\mu\right)^\top \Big| d(X)=0\right] \succ 0$, there exists $\delta > 0$ and $c_0 \in (0,+\infty)$ such that

$$\inf_{\|\gamma\|_2=1} \mathbb{E}\left[\min\left\{c_0, \gamma^\top \phi\left(U,\mu\right)\right\}\left|\gamma^\top \phi\left(U,\mu\right)\right|\right] > \delta.$$

for all $\|\gamma\|_2 = 1$. And

$$\inf_{\|\gamma\|_2=1} \mathbb{E}\left|\gamma^\top \phi\left(U,\mu\right)\right| > 0,$$

since the unit circle is compact. Let $\delta = \inf_{\|\gamma\|_2=1} \mathbb{E}\left|\gamma^\top \phi\left(U,\mu\right)\right|$. For any $\varepsilon > 0$, there exists $N_1 > 0$ and $b' < +\infty$, such that

$$\mathbb{P}(\|V_N\|_2 \geq b') < \varepsilon/2,$$

for any $N > N_0$. Recalling Lemma A2 and equation (A.11), there exists $N_0 > N_1$ such that

$$\mathbb{P}\left(\exists \gamma : \|\gamma\|_2 = b \text{ such that } M_N(\gamma) \geq -\frac{1}{4}\mathbb{E}\left[\min\left\{bc_0, \left|\gamma^\top \phi\left(U,\mu\right)\right|\right\}\left|\gamma^\top \phi\left(U,\mu\right)\right|\right]\right) < \varepsilon/2$$

for any $N > N_0$. Then, we have

$$\begin{aligned} & \inf_{\|\gamma\|_2=b} \mathbb{E}\left[\min\left\{bc_0, \left|\gamma^\top \phi\left(U,\mu\right)\right|\right\}\left|\gamma^\top \phi\left(U,\mu\right)\right|\right] \\ \geq \; & b^2 \inf_{\|\gamma\|_2=1} \mathbb{E}\left[\min\left\{c_0, \left|\gamma^\top \phi\left(U,\mu\right)\right|\right\}\left|\gamma^\top \phi\left(U,\mu\right)\right|\right] > b^2\delta. \end{aligned}$$

Let $b = 4b'/\delta$. We have

$$\mathbb{P}\left(\sup_{\|\gamma\|_2=b} M_N(\gamma) \geq -bb'\right) < \varepsilon/2. \tag{A.15}$$

Notice that for any $\|\gamma\|_2 > b$,

$$M_N(\gamma) \leq \frac{\|\gamma\|_2}{b} M_N\left(\frac{b}{\|\gamma\|_2}\gamma\right) \leq \frac{\|\gamma\|_2}{b} \sup_{\|\gamma\|_2=b} M_N(\gamma). \tag{A.16}$$

By combining inequalities (A.15) and (A.16), we have

$$\mathbb{P}\left(\exists\gamma : \|\gamma\|_2 > b, \text{ such that } M_N(\gamma) \geq -\|\gamma\|_2 b'\right) < \varepsilon/2.$$

Therefore,

$$\begin{aligned} & \mathbb{P}\left(\sup_{\|\gamma\|_2>b}\left\{\gamma^\top V_N + M_N(\gamma)\right\} > 0\right) \\ \leq \; & \mathbb{P}\left(\sup_{\|\gamma\|_2>b}\left\{\|\gamma\|_2 \|V_N\|_2 + M_N(\gamma)\right\} > 0\right) \\ \leq \; & \mathbb{P}(\|V_N\|_2 \geq b') + \mathbb{P}\left(\exists\gamma : \|\gamma\|_2 > b, \text{ such that } M_N(\gamma) \geq -\|\gamma\|_2 b'\right) \\ \leq \; & \varepsilon. \end{aligned}$$

This completes the proof. □

Appendix B Additional Details for Numerical Experiments

Appendix B.1 Validation of the Hypothesis Test

In this section, we empirically validate the convergence result in Theorem 1 and our proposed hypothesis test method. we use a simple logistic classifier in the form

$$\mathcal{C}(x) = \mathbb{I}\left\{\frac{1}{1+\exp\left(-\theta^\top x\right)} \geq \tau\right\}.$$

Then, the decision boundary is $\left\{x : \theta^\top x = -\log\left(\frac{1}{\tau} - 1\right)\right\}$. We denote $w = -\log\left(\frac{1}{\tau} - 1\right)$. Then, we borrowed the example in Taskesen et al. [4]. Let

$$p_{11} = 0.4, p_{01} = 0.1, p_{10} = 0.4, p_{00} = 0.1.$$

Moreover, conditioning on (A, Y), the feature X follows a Gaussian distribution of the form

$$\begin{aligned}
X|A = 1, Y = 1 \sim & \mathcal{N}([6, 0], [3.5, 0; 0, 5]), \\
X|A = 0, Y = 1 \sim & \mathcal{N}([-2, 0], [5, 0; 0, 5]), \\
X|A = 1, Y = 0 \sim & \mathcal{N}([6, 0], [3.5, 0; 0, 5]), \\
X|A = 0, Y = 0 \sim & \mathcal{N}([-4, 0], [5, 0; 0, 5]).
\end{aligned}$$

The true distribution $\mathbb{P}$ is thus a mixture of Gaussian. A simple algebraic calculation indicates that a logistic classifier with $\theta = (0, 1)^\top$ and $\tau = 0.5$ is fair with respect to the equal opportunity criterion in Example 1. Let $\varphi(\cdot)$ denotes the density of the standard normal distribution and we denote μ_{ay} and Σ_{ay} to be the conditional mean and variable defined above, respectively. For any θ, the density of $\theta^\top X$ becomes

$$\sum_{a,y \in \{0,1\}^2} \left(\theta^\top \Sigma_{ay} \theta\right)^{-1/2} p_{ay} \varphi\left(\left(\theta^\top \Sigma_{ay} \theta\right)^{-1/2} \left(\theta^\top x - \theta^\top \mu_{ay}\right)\right).$$

And thus the density of $\Phi(\cdot)$ becomes

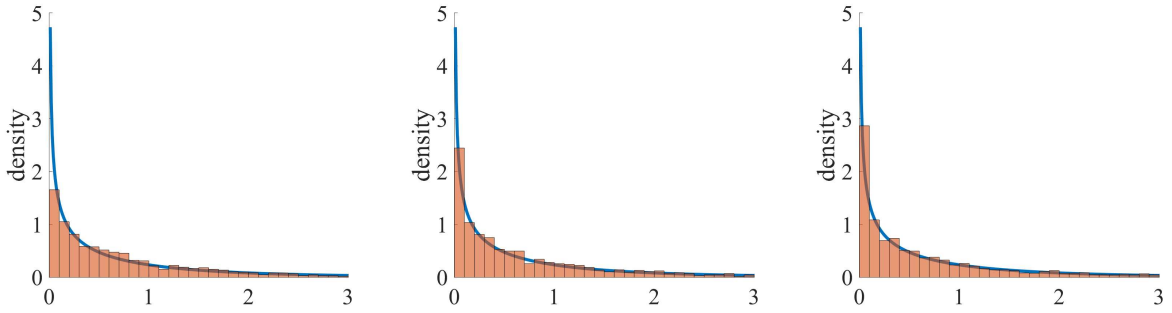
$$f(z) = \|\theta\|_* \sum_{a,y \in \{0,1\}^2} \left(\theta^\top \Sigma_{ay} \theta\right)^{-1/2} p_{ay} \varphi\left(\left(\theta^\top \Sigma_{ay} \theta\right)^{-1/2} \left((z \|\theta\|_* + w) - \theta^\top \mu_{ay}\right)\right).$$

By Bayes formula, we have

$$p_{ay|d(X)=0} = f(0)^{-1} \left(\theta^\top \Sigma_{ay} \theta\right)^{-1/2} \|\theta\|_* p_{ay} \varphi\left(\left(\theta^\top \Sigma_{ay} \theta\right)^{-1/2} \left(w - \theta^\top \mu_{ay}\right)\right)$$

for $a \in \{0, 1\}$ and $y \in \{0, 1\}$, where $p_{ay|d(X)=0} = \mathbb{E}[\mathbb{I}_{(a,y)}(A, Y)|d(X) = 0]$. In the first experiments, we generate $N \in \{30, 100, 500\}$ i.i.d. samples from $\mathbb{P}$ and then calculate $N \times \mathcal{D}(\hat{\mathbb{P}}^N)$. We replicate this process for 2,000 times and compare the empirical distribution of $N \times \mathcal{D}(\hat{\mathbb{P}}^N)$ with the limiting distribution defined in Theorem 1. Figure 1 shows that finite-sample empirical estimates are closed to the theoretical limiting distributions even when N is as small as 30.

In the second experiments, we show that our proposed Wasserstein projection hypothesis test has the desired coverage property. We generate $N \in \{30, 100, 500, 1000, 2000\}$ i.i.d. samples from $\mathbb{P}$ and compute the estimate $\hat{S}$ defined in Section 5.2 and the empirical covariance using the sample data. For the kernel estimator $\hat{S}$, we use the standard Gaussian kernel and choose the bandwidth $h = N^{-1/5}$, where the results listed below are not sensitive to the constant. We repeat the procedure for 2,000 replications and report the rejection probability at different significant values of $\alpha \in \{0.1, 0.05, 0.01\}$ in Table 1. We can observe that when $N > 100$, the rejection probability is closed to the desired level α.



(a) $N = 30$ (b) $N = 100$ (c) $N = 500$

Figure 1: Empirical distribution $N \times \mathcal{D}(\hat{\mathbb{P}}^N)$ over 2,000 replications (histogram) versus the limiting Chi-square distribution (blue curve) with different sample sizes N.

Table 1: Comparison of the null rejection probabilities of probabilistic equal opportunity tests with different significance levels α and test sample sizes N.

α	0.10	0.05	0.01
$N = 30$	0.2875	0.2255	0.1415
$N = 100$	0.0945	0.0540	0.0250
$N = 500$	0.0895	0.0450	0.0085
$N = 1000$	0.0900	0.0430	0.0065
$N = 2000$	0.0870	0.0460	0.0080

Appendix B.2 The Description of Datasets

Followings show brief descriptions of datasets: Arrhythmia, COMPAS and Drug [3] provided in Section 6.

- **Arrhythmia** is from UCI repository[1], where the aim of this data set is to distinguish between the presence and absence of cardiac arrhythmia and classify it in one of the 16 groups. The dataset consists of 452 samples and we use the first 12 features among which the gender is the sensitive feature. For our purpose, we construct binary labels between 'class 01' ('normal') and all other classes (different classes of arrhythmia and unclassified ones).

- **COMPAS** (Correctional Offender Management Profiling for Alternative Sanctions) [2] is a commerical tool used by judges, probation and parole officers to estimate a criminal defendant's likelihood to re-offend algorithmically. The COMPAS dataset contains the criminal records within 2 years after the decision. We use race (African-American and Caucasian, which accounts for 5278 samples) as the sensitive attribute.

- **Drug** [3] contains answers of 1885 participants on their use of 17 legal and illegal drugs. We concern the cannabis usage as a binary problem, where the label is 'Never used' VS 'Others'

[1] `https://archive.ics.uci.edu/ml/datasets/arrhythmia`

[2] `https://www.propublica.org/datastore/dataset/compas-recidivism-risk-score-data-and-analysis`

('used'). There are 12 features including age, gender, education, country, ethnicity, NEO-FFI-R measurements, impulsiveness measured by BIS-11 and sensation seeing measured by ImpSS. Among those, we choose ethnicity (black vs others) as the sensitive attribute.

References

[1] Patrick Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, 2013.

[2] Rick Durrett. *Probability: Theory and Examples*. Cambridge University Press, 2019.

[3] Elaine Fehrman, Awaz K Muhammad, Evgeny M Mirkes, Vincent Egan, and Alexander N Gorban. The five factor model of personality and evaluation of drug consumption risk. In *Data Science*, pages 231–242. Springer, 2017.

[4] Bahar Taskesen, Jose Blanchet, Daniel Kuhn, and Viet Anh Nguyen. A statistical test of probabilistic fairness. *Accepted to ACM Conference on Fairness, Accountability, and Transparency*, 2021.