Supplementary Material for "Avoiding Disparity Amplification under Different Worldviews"

Samuel Yeom
Carnegie Mellon University
syeom@cs.cmu.edu

Michael Carl Tschantz
International Computer Science Institute
mct@icsi.berkeley.edu

A PROOFS OF THEOREMS IN SECTION 9

Theorem 11. *Let the construct Y' be categorical with support $\mathcal{Y}'$, which has distance metric $d(u,v) = \mathbb{1}(u \neq v)$. If a model has disparity amplification under Definition 9, the model has disparity amplification under Definition 13 as well.*

Proof. We proceed by showing that $\rho_\ell^* \cdot d_{\mathrm{em}}(Y'|Z{=}0, Y'|Z{=}1) \leq d_{\mathrm{tv}}(Y'|Z{=}0, Y'|Z{=}1)$.

Since the likelihood function ℓ in Definition 13 is always between 0 and 1, we have $|\ell(u) - \ell(v)| \leq 1 = d(u,v)$ when $u \neq v$, so ℓ is 1-Lipschitz continuous. Therefore $\rho_\ell^* \leq 1$, and it suffices to show that $d_{\mathrm{em}}(Y'|Z{=}0, Y'|Z{=}1) \leq d_{\mathrm{tv}}(Y'|Z{=}0, Y'|Z{=}1)$.

By [1, Theorem 4], we get

$$\begin{aligned} d_{\mathrm{em}}(Y'|Z{=}0, Y'|Z{=}1) &\leq \left(\max_{u,v\in\mathcal{Y}'} d(u,v)\right) \cdot d_{\mathrm{tv}}(Y'|Z{=}0, Y'|Z{=}1) \\ &= d_{\mathrm{tv}}(Y'|Z{=}0, Y'|Z{=}1), \end{aligned}$$

so we are done. □

Theorem 12. *A model that passes the demographic parity test does not have disparity amplification under Definition 13.*

Proof. Under Definition 13, a model has disparity amplification when, for $\ell(y') = \Pr[\hat{Y}{=}1 \mid Y'{=}y']$ and ρ_ℓ^* being smallest nonnegative ρ such that ℓ is ρ-Lipschitz continuous,

$$d_{\mathrm{tv}}(\hat{Y}|Z{=}0, \hat{Y}|Z{=}1) > \rho_\ell^* \cdot d_{\mathrm{em}}(Y'|Z{=}0, Y'|Z{=}1).$$

The left-hand side of this inequality is $d_{\mathrm{tv}}(\hat{Y}|Z{=}0, \hat{Y}|Z{=}1) = 0$ when demographic parity holds. The right-hand side of this inequality is nonnegative since ρ_ℓ^* and $d_{\mathrm{em}}(Y'|Z{=}0, Y'|Z{=}1)$ are nonnegative. Thus, demographic parity ensures that the inequality cannot hold and the lack of disparity amplification under Definition 13. □

Theorem 13. *If the WYSIWYG worldview holds, then a model that passes the equalized odds test does not have disparity amplification under Definition 13.*

Proof. We present the proof for the case where Y' is continuous, but the proof for the discrete case is very similar. Let p_0 and p_1 be the probability density functions of $Y'|Z{=}0$ and $Y'|Z{=}1$, respectively. By Kantorovich duality [2, Equation 5.4], we have

$$\begin{aligned} &d_{\mathrm{em}}(Y'|Z{=}0, Y'|Z{=}1) \\ &\qquad \geq \int_{\mathcal{Y}'} \phi(v)\, p_1(v)\, dv - \int_{\mathcal{Y}'} \psi(u)\, p_0(u)\, du \end{aligned} \tag{1}$$

for all ϕ and ψ such that $\phi(v) - \psi(u) \leq d(u,v)$ for all $u,v \in \mathcal{Y}'$. We set $\phi(v) = \psi(v) = \ell(v)/\rho_\ell^*$, where ℓ and ρ_ℓ^* are defined as in Definition 13. Then, $\phi(v) - \psi(u) = (\ell(v) - \ell(u))/\rho_\ell^* \leq d(u,v)$ by Lipschitz continuity. Thus, (1) applies and implies that

$$\begin{aligned} &\rho_\ell^* \cdot d_{\mathrm{em}}(Y'|Z{=}0, Y'|Z{=}1) \\ &\qquad \geq \int_{\mathcal{Y}'} \ell(v)\, p_1(v)\, dv - \int_{\mathcal{Y}'} \ell(u)\, p_0(u)\, du. \end{aligned} \tag{2}$$

By the WYSIWYG worldview and equalized odds, we have $\ell(y) = \Pr[\hat{Y}{=}1 \mid Y'{=}y] = \Pr[\hat{Y}{=}1 \mid Y'{=}y, Z{=}0] = \Pr[\hat{Y}{=}1 \mid Y'{=}y, Z{=}1]$. Therefore, we can use the law of total probability to rewrite the first term on the right-hand side of (2) as $\Pr[\hat{Y}{=}1 \mid Z{=}1]$, and similarly the second term becomes $\Pr[\hat{Y}{=}1 \mid Z{=}0]$.

If we let $\phi(v) = \psi(v) = -\ell(v)/\rho_\ell^*$ in (1) instead, we get $\rho_\ell^* \cdot d_{\mathrm{em}}(Y'|Z{=}0, Y'|Z{=}1) \geq \Pr[\hat{Y}{=}1 \mid Z{=}0] - \Pr[\hat{Y}{=}1 \mid Z{=}1]$. Finally, combining this inequality with the previous one gives us

$$\begin{aligned} \rho_\ell^* \cdot d_{\mathrm{em}}(Y'|Z{=}0, Y'|Z{=}1) &\geq \left|\Pr[\hat{Y}{=}1 \mid Z{=}0] - \Pr[\hat{Y}{=}1 \mid Z{=}1]\right| \\ &= d_{\mathrm{tv}}(\hat{Y}|Z{=}0, \hat{Y}|Z{=}1), \end{aligned}$$

which is what we want. □

REFERENCES

[1] Alison L Gibbs and Francis Edward Su. 2002. On choosing and bounding probability metrics. *International Statistical Review* 70, 3 (2002), 419–435.

[2] Cédric Villani. 2008. *Optimal transport: old and new.* Grundlehren der mathematischen Wissenschaften: Comprehensive Studies in Mathematics, Vol. 338. Springer-Verlag Berlin Heidelberg.