Mitigating Discrimination in **Insurance**

via Wasserstein Barycenters

Arthur Charpentier [1], François HU [2] and Philipp Ratz [1]
[1] Université du Québec à Montréal
[2] Université de Montréal

3rd Workshop on Bias and Fairness in AI (BIAS 2023)

September 22, 2023

Outline

1 Introduction

2 Quantification of the Bias

3 Mitigation of the Bias

4 Case Study on Motor Insurance

5 Conclusion

Introduction

Insurance is characterised by the **need to segment its customers**, or as put by Avraham,“**the core of insurance business lies discrimination between risky and non-risky insureds**”, [Avr17], to ask for an **actuarially fair premium**.

However, problems arise when these characteristics are strongly linked to sensitive variables, as illustrated in Figure 1, which shows the practice of **redlining**.

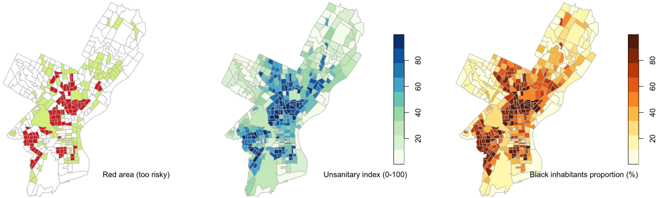


Figure 1: Fictitious maps, (freely) inspired by a Home Owners’ Loan Corporation map from 1937.

Left: Locations where investments and **loans** have been discouraged or encouraged.

Middle: **risk** related variable (fictitious "unsanitary index") per neighborhood.

Right: **sensitive** variable (the proportion of Black people in the neighborhood).

Discrimination in Insurance: examples from the US and Canada

	CA	HI	GA	NC	NY	MA	PA	FL	TX	AL	ON	NB	NL	QC
Gender	x	x	●	x	●	x	x	●	●	●	●	x	x	●
Age	x	x	●	x*	●	x	●	●	●	●*	●	x	x	●
Driving experience	●	x	●	●	●	●	●	●	●	●	●	●	●	●
Credit history	x	x	●	●	●	x	●*	●	●	x*	x	●*	x	●
Education	x	x	x	x	x	x	●	●	●	●	●	●	●	●
Profession	x	x	x	●	x	x	●	●	●	●	●	●	●	●
Employment	x	x	x	●	x	x	●	●	●	●	●	●	●	●
Family	●	x	●	●	●	x	●	●	●	●	●	●	●	●
Housing	x	x	●	●	●	x	●	●	●	x	x	●	●	●
Address/ZIP code	●	●	●	●	●	●	●	●	●	x	x	●	●	●

Table 1: A factor is "permitted" (●) if state or provincial laws don't forbid insurers from using it; otherwise, it's "prohibited" (x). * We have given some simplifications. **Source**: in the United States, The Zebra (2022) and in Canada, Insurance Bureau of Canada (2021).

In particular, one of the common goals of **Algorithmic Fairness** is to make any prediction $\hat{Y}$ independent of a sensible feature S.

$$\hat{Y} \perp\!\!\!\perp S$$

This notion of fairness is called **Demographic Parity (DP)**.

Quantification of Bias

Consider two probability measures, ν_1 and ν_2. We define **distance function** between ν_1 and ν_2:

Definition (Wasserstein distance)

The squared Wasserstein distance between ν_1 and ν_2 is defined as

$$\mathcal{W}_2^2(\nu_1, \nu_2) = \inf_{\pi \in \Pi(\nu_1, \nu_2)} \mathbb{E}_{(Z_1, Z_2) \sim \pi} \left(Z_2 - Z_1\right)^2 \quad ,$$

where $\Pi(\nu_1, \nu_2)$ is the set of distributions on $\mathcal{Y} \times \mathcal{Y}$ having ν_1 and ν_2 as marginals.

The Wasserstein barycenter finds a **representative distribution** that lies between multiple given distributions in the Wasserstein space. It is defined for a family of K measures $(\nu_1, \ldots, \nu_K)$ in $\mathcal{V}$ and some positive weights $(w_1, \ldots, w_K) \in \mathbb{R}_+^K$.

Definition (Wasserstein Barycenters)

The Wasserstein barycenter, denoted as $\mathrm{Bar}\left\{(w_k, \nu_k)_{k=1}^K\right\}$ is the minimiser

$$\mathrm{Bar}(w_k, \nu_k)_{k=1}^K = \underset{\nu}{\mathrm{argmin}} \sum_{k=1}^{K} w_k \cdot \mathcal{W}_2^2\left(\nu_k, \nu\right) \quad .$$

The barycenter exists and is unique if one of ν_k admits a density wrt the Lebesgue measure [AC11].

Mitigation of the Bias

To provide mitigation, we make use of the **Wasserstein Barycenter**.

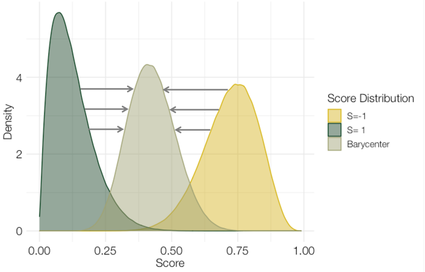


Figure 2: Two distributions, induced by differences in the sensitive value *s* and their Barycenter.

The Wasserstein Barycenter **minimises the Risk under DP-fairness constraint**.

We use a **model-agnostic**, **closed form solution** to obtain our **Barycenters** [HRC23].

Case Study on Motor Insurance: Gender-free prediction

- **Data:** Publicly available data set from **motor insurance**.
- **Task:** *Gender-free* prediction.

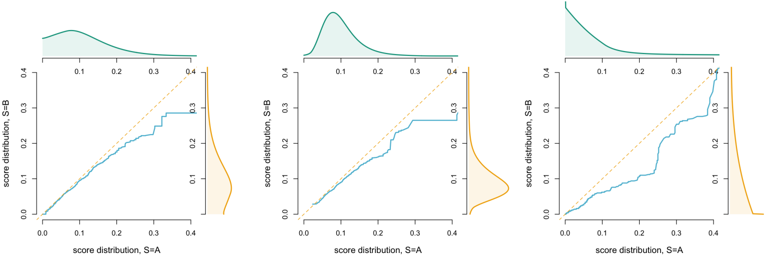


Figure 3: Matching btw $m(\boldsymbol{x}, s = A)$ and $m(\boldsymbol{x}, s = B)$, where m is (left to right) GLM, GBM and RF.

Gender-free prediction

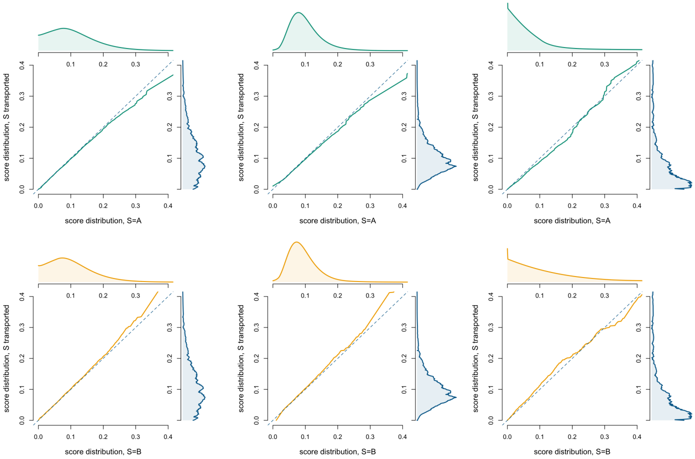


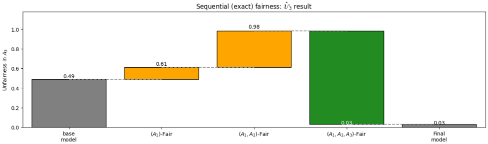
Figure 4: Matching between $m(\boldsymbol{x}, s = A)$ and $m^\star(\boldsymbol{x}, s = A)$, on top, and between $m(\boldsymbol{x}, s = B)$ and $m^\star(\boldsymbol{x}, s = B)$, below, on the probability to claim a loss when s is the gender of the driver.

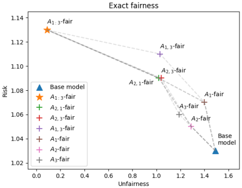
Conclusion

1. Discrimination emerges in risk prediction models tied to sensitive attributes.
2. Wasserstein distance and optimal transport enable fair predictions by assessing **complete prediction distributions**.
3. Empirical results affirm the approach's value in **fostering fairness in insurance**.

Extension: Leverage optimal transport theory to address **multiple sensitive features**.

See our recent paper: *"Sequentially Fair Mechanism for Multiple Sensitive Attributes"*





Conclusion

Thank you !

References I

[AC11] M. Agueh and G. Carlier. Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

[Avr17] Ronen Avraham. Discrimination and insurance. In Kasper Lippert-Rasmussen, editor, *Handbook of the Ethics of Discrimination*, pages 335–347. Routledge, 2017.

[HRC23] François Hu, Philipp Ratz, and Arthur Charpentier. Fairness in multi-task learning via wasserstein barycenters. In Danai Koutra, Claudia Plant, Manuel Gomez Rodriguez, Elena Baralis, and Francesco Bonchi, editors, *Machine Learning and Knowledge Discovery in Databases: Research Track*, pages 295–312, Cham, 2023. Springer Nature Switzerland.