Group Fairness with Uncertain Sensitive Attributes


Abhin Shah


Maohao Shen


Jongha Jon Ryu


Subhro Das

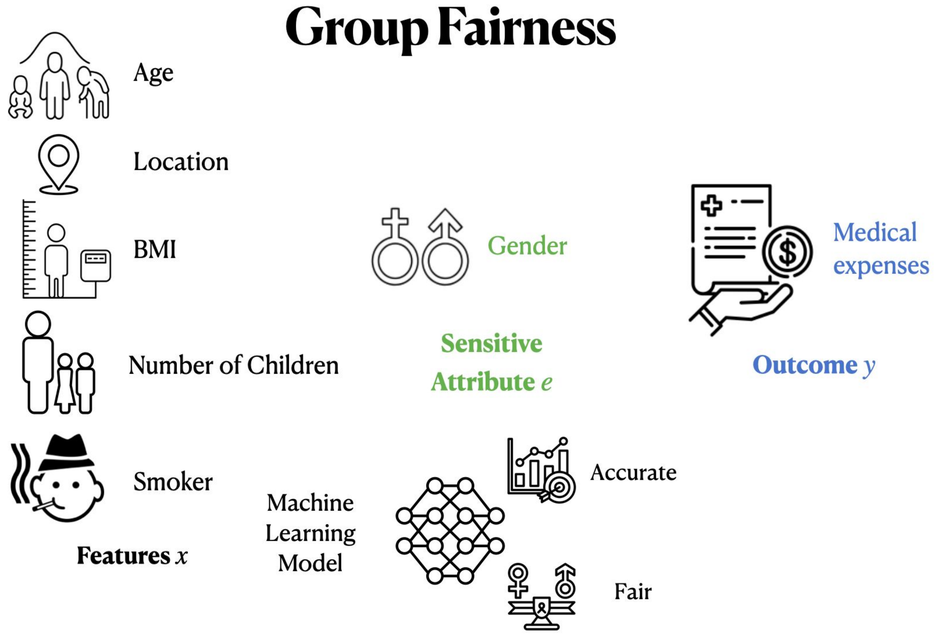

Prasanna Sattegiri


Yuheng Bu

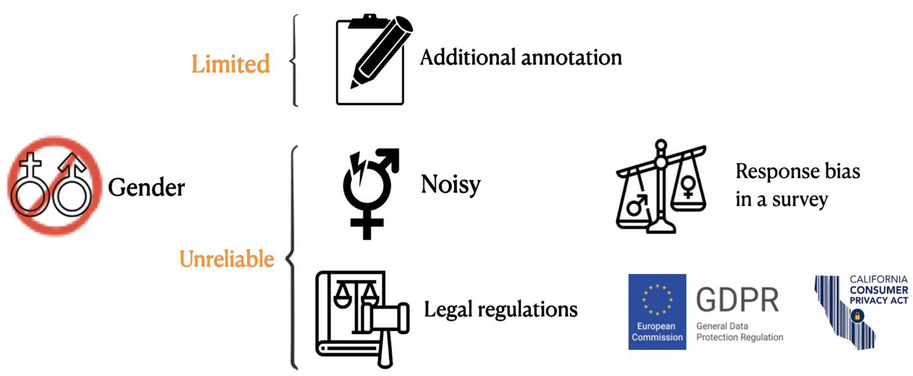

Greg Wornell

arXiv Link

Group Fairness



Uncertainty in Sensitive Attribute



Age	Location	BMI	Number of children	Smoker	Medical expenses	Gender	Limited Gender	Unreliable Gender
19	Southwest	27.9	0	Yes	16884	Female	?	Male
28	Southeast	33	3	No	4449	Male	Male	Male
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
62	Southeast	26.29	0	Yes	27808	Female	?	Female

Goal — Learn a fair model despite uncertain sensitive attribute data

Limitations of Existing Work

A. Proxy variables — effectiveness depends on the degree of correlation between the sensitive attribute and the proxy variables

B. Perturbed sensitive attribute — focus on specific perturbation models

Problem Formulation

Design a randomized mapping $p_{u|x}$ such that the representation u is maximally informative about y while limiting how informative it is about e

$$\min_{p_{u|x}} \min_{\hat{y}(u)} MSE\big(y, \hat{y}(u)\big) \quad \text{s.t} \quad \text{Dependence}(e; u) \le \epsilon$$

Gaussian Data

Model the distribution of (x, y, e, u) as Gaussian

$$\max_{a \in \mathcal{B}(0,1)} \langle a, b_{yx} \rangle^2 \quad \text{s.t} \quad \langle a, b_{ex} \rangle^2 \le \epsilon \quad \text{where} \quad a = b_{ux} \text{ and } b_{vw} \triangleq \Sigma_v^{-1/2} \Sigma_{vw} \Sigma_w^{-1/2}$$

Quadratically Constrained Quadratic Program (QCQP)

An optimal solution $a^\star$ of the above QCQP lies in the subspace spanned by the vectors b_{yx} and b_{ex}

Baseline

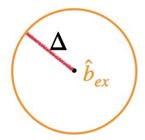
$$\max_{a \in \mathcal{B}(0,1)} \langle a, b_{yx} \rangle^2 \quad \text{s.t} \quad \langle a, \hat{b}_{ex} \rangle^2 \le \epsilon$$

This does not guarantee fairness

Robust QCQP

Uncertainty in sensitive attributes

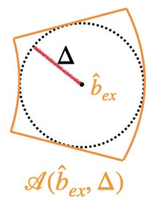
$$\max_{a \in \mathcal{B}(0,1)} \langle a, b_{yx} \rangle^2 \quad \text{s.t} \quad \langle a, b \rangle^2 \le \epsilon \quad \text{for all} \quad b \in \mathcal{B}(\hat{b}_{ex}, \Delta)$$

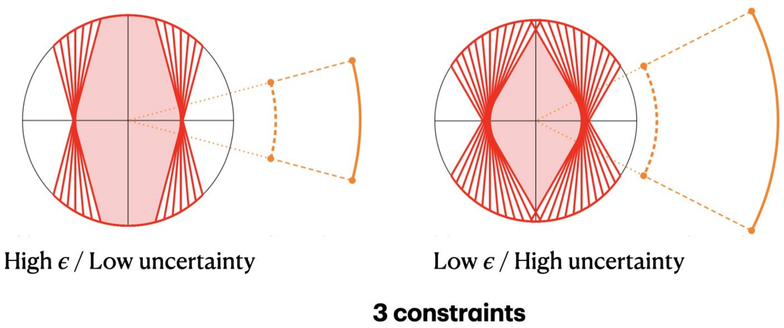


An optimal solution $a^\star$ of the above QCQP lies in the subspace spanned by the vectors b_{yx} and $\hat{b}_{ex}$

Relaxing the uncertainty

$$\max_{a \in \mathcal{B}(0,1)} \langle a, b_{yx} \rangle^2 \quad \text{s.t} \quad \langle a, b \rangle^2 \le \epsilon \quad \text{for all} \quad b \in \mathcal{A}(\hat{b}_{ex}, \Delta)$$



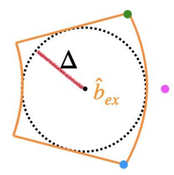


High ϵ / Low uncertainty

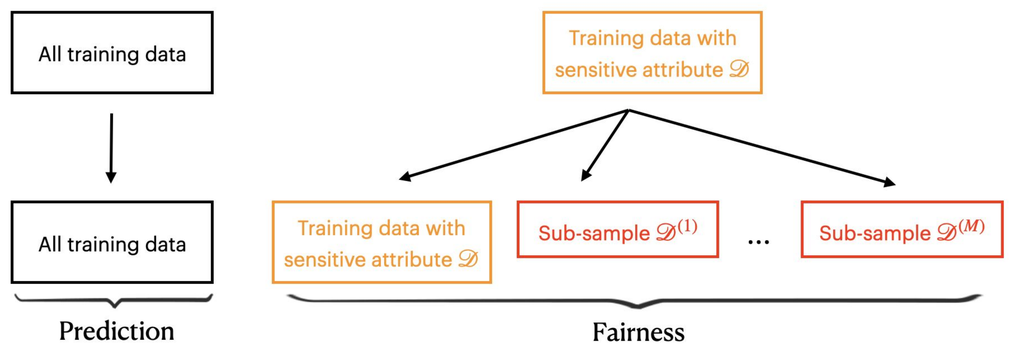
Low ϵ / High uncertainty

3 constraints

$$\max_{a \in \mathcal{B}(0,1)} \langle a, b_{yx} \rangle^2 \quad \text{s.t} \quad \langle a, b^{(i)} \rangle^2 \le \epsilon \quad \text{for all} \quad i \in [3]$$



Algorithm



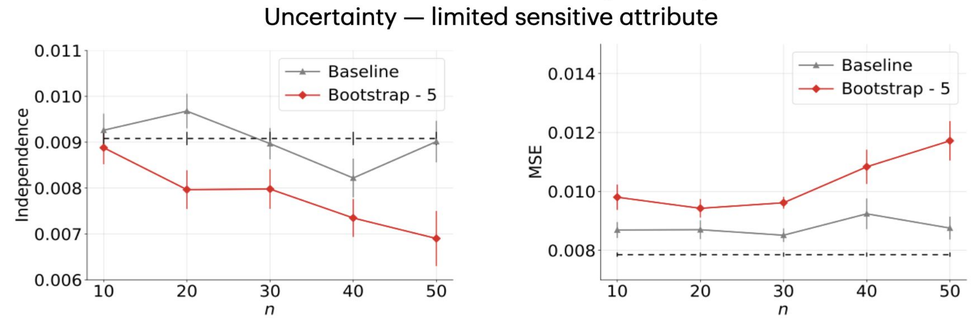
For some $k \in [n]$ and $M \ge 1$, uniformly draw $\mathcal{D}^{(1)}, \cdots, \mathcal{D}^{(M)}$ each of size k from $\mathcal{D}$ with replacement

Baseline :— $\min$ Prediction Loss $+ \lambda$ Fairness Loss$(\mathcal{D})$

Bootstrap-M :— $\min$ Prediction Loss $+ \lambda$ Fairness Loss$(\mathcal{D}) + \dfrac{\lambda}{M} \sum_{i \in [M]}$ Fairness Loss$(\mathcal{D}^{(i)})$

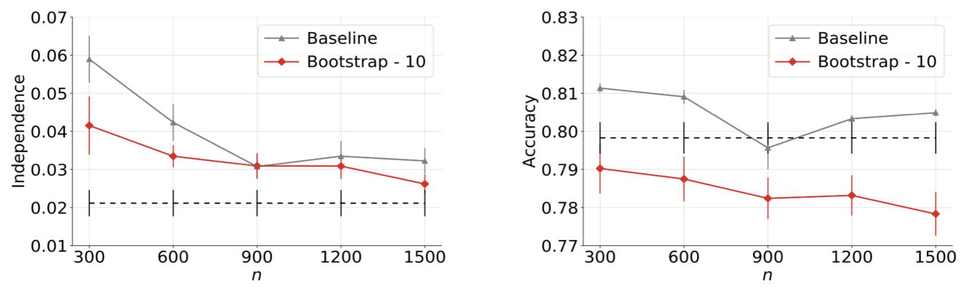
Insurance Dataset

Task — predict an individual's medical expenses (Regression)
Sensitive attribute — Gender (Discrete)
Uncertainty — limited sensitive attribute



Adult Dataset

Task — predict whether an individual's income > $50k (Classification)
Sensitive attribute — Gender (Discrete)
Uncertainty — limited sensitive attribute



Crime Dataset

Task— predict density of violent crimes (Regression)
Sensitive attribute — Race (Continuous)
Uncertainty — unreliable sensitive attribute