Pushing the limits of fairness impossibility: Who's the fairest of them all?

Kyungseon Lee

January 30, 2023

Seoul National University

Outline

1. Introduction

2. Multiple Fairness Optimization

3. Finding Optimal Solutions via Mixed Integer Programming (MIP)

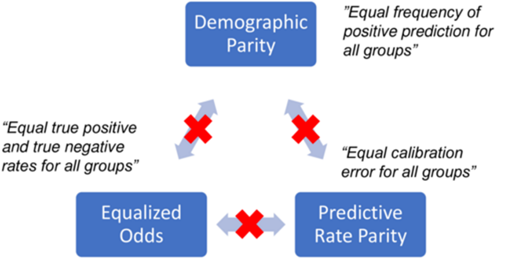
4. Applications

Introduction

Introduction(1/3)

Theorem (Impossibility Theorem in fairness)

Three common definitions of algorithmic fairness - demographic parity, equalized odds, and predictive parity, cannot be simultaneously satisfied outside of pathological situations.


Demographic Parity
"Equal frequency of positive prediction for all groups"
"Equal true positive and true negative rates for all groups"
"Equal calibration error for all groups"
Equalized Odds
Predictive Rate Parity

Introduction(2/3)

Theorem (Impossibility Theorem in fairness)

Three common definitions of algorithmic fairness - demographic parity, equalized odds, and predictive parity, cannot be simultaneously satisfied outside of pathological situations.

The trade-offs among multiple fairness criteria and model performance.

- A constrained optimization problem.
- The paper propose a post-processing methodology for simultaneously achieving approximate fairness in the conflicting definitions simultaneously.

Introduction(3/3)-Related Work

Overall, we make three main contributions in this work:

1. We design a flexible optimization framework that returns a **post-processing score transformation function**.
   - It can **make scores group-wise ϵ-fair** along three definitions simultaneously.
2. We present a novel reformulation of this **non-convex optimization problem** as a **Mixed Integer Linear Program (MILP)**.
3. We discuss and extend our framework from a post-processing mechanism to a tool.
   - The tool can aid practitioners in better understanding **their data and models' empirical fairness characteristics and trade-offs**.
   - It can **compare** these traits across models.

Multiple Fairness Optimization

Multiple Fairness Optimization(1/6)

We discretize the **scores** into **nonempty bins** $b \in \mathcal{B} := \{1, \ldots, |\mathcal{B}|\}$ by using, for example, a quantile transformation.

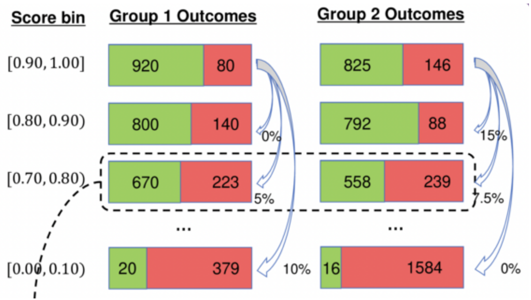
- $N_{b+}^{[g]}$: the number of group g positive class ($y_i = 1$) instances in bin b.
- $N_{b}^{[g]}$: the total number of instances of group g instances in bin b.
- $N_{+}^{[g]} \left(N_{-}^{[g]} \right)$: the total number of group g positive (negative) instances.
- $N^{[g]}$: the total number of group g instances.
- $N_{b}^{[g]}$: the number of observations from group g in bin b (before transformation).

- i : i-th observation.
- $y_i \in \{0, 1\}$: label of binary classification.
- $g_i \in \mathcal{G}$: group membership.
- $s_i \in [0, 1]$ for $i = 1, \ldots, N$: model predicted probability.
- Y, G, S : the corresponding random variables.

Multiple Fairness Optimization(2/6)

Definition

We define variable $x_{bb'}^{[g]}$ as the probability of moving an instance of group attribute g and score in bin b into a new bin b'



Multiple Fairness Optimization(3/6)

Fairness Metrics

1. Demographic Parity (DP)

$$P(S = s \mid G = 1) = P(S = s \mid G = 2).$$

2. Equalized Odds (EOdds)

$$P(S = s \mid Y = y, G = 1) = P(S = s \mid Y = y, G = 2) \text{ for } y \in \{0, 1\}.$$

3. Predictive Rate Parity (PRP)

$$P(Y = 1 \mid S = s, G = 1) = P(Y = 1 \mid S = s, G = 2).$$

Multiple Fairness Optimization(4/6) - Fairness Constraints

After transforming the scores using $\left\{x_{bb'}^{[g]}\right\}_{b,b'}$ the ϵ-approximate :

1. Demographic Parity (DP)

$$\left| \frac{1}{N^{[1]}} \sum_{b \in \mathcal{B}} x_{bb'}^{[1]} N_b^{[1]} - \frac{1}{N^{[2]}} \sum_{b \in \mathcal{B}} x_{bb'}^{[2]} N_b^{[2]} \right| \leq \epsilon_{DP} \quad \forall b' \in \mathcal{B}$$

2. Equalized Odds (EOdds)

$$\left| \frac{1}{N_{\pm}^{[1]}} \sum_{b \in \mathcal{B}} x_{bb'}^{[1]} N_{b\pm}^{[1]} - \frac{1}{N_{\pm}^{[2]}} \sum_{b \in \mathcal{B}} x_{bb'}^{[2]} N_{b\pm}^{[2]} \right| \leq \epsilon_{EOdds} \forall b' \in \mathcal{B}$$

3. Predictive Rate Parity (PRP) : non-convex constraint.

$$\left| \frac{\sum_{b \in \mathcal{B}} x_{bb'}^{[1]} N_{b+}^{[1]}}{\sum_{b \in \mathcal{B}} x_{bb'}^{[1]} N_b^{[1]}} - \frac{\sum_{b \in \mathcal{B}} x_{bb'}^{[2]} N_{b+}^{[2]}}{\sum_{b \in \mathcal{B}} x_{bb'}^{[2]} N_b^{[2]}} \right| \leq \epsilon_{PRP} \quad \forall b' \in \mathcal{B}.$$

- We want to make the "best" probabilistic moves such that the fairness constraints will be satisfied in expectation.

Multiple Fairness Optimization(5/6)

MFOpt: Multiple Fairness Optimization Framework

$$\text{minimize}_{\left\{x_{bb'}^{[g]}\right\}_{b,b',g}} \sum_{g\in\mathcal{G}} \sum_{b\in\mathcal{B}} \sum_{b'\in\mathcal{B}} \left| \frac{N_b^{[g]}}{N} \left(\bar{s}_b - \bar{s}_{b'}\right) x_{bb'}^{[g]} \right|$$

s.t. $\sum_{b\in\mathcal{B}} x_{bb'}^{[g]} = 1 \quad \forall b' \in \mathcal{B}, g \in \mathcal{G}$

$x_{bb}^{[g]} \geq 1 - \xi \quad \forall b \in b' \in \mathcal{B}, g \in \mathcal{G}$

$x_{bb'}^{[g]} = 0 \quad \forall b' \text{ s.t. } |b' - b| \geq w, \forall g \in \mathcal{G}$

Fairness Constraints: $(1), (2), (3)$

$$\frac{\sum_{b\in\mathcal{B}} x_{bb'}^{[g]} N_{b+}^{[g]}}{\sum_{b\in\mathcal{B}} x_{bb'}^{[g]} N_b^{[g]}} \leq \frac{\sum_{b\in\mathcal{B}} x_{b\left(b'+1\right)}^{[g]} N_{b+}^{[g]}}{\sum_{b\in\mathcal{B}} x_{b\left(b'+1\right)}^{[g]} N_b^{[g]}} \quad \forall b' \in \{1, \ldots, B-1\}, g \in \mathcal{G}$$

$0 \leq x_{bb'}^{[g]} \leq 1, \quad \forall b, b', g$

Multiple Fairness Optimization(6/6)

MFOpt: Multiple Fairness Optimization Framework

$$\text{minimize}_{\left\{x_{bb'}^{[g]}\right\}_{b,b',g}} \sum_{g\in\mathcal{G}} \sum_{b\in\mathcal{B}} \sum_{b'\in\mathcal{B}} \left| \frac{N_b^{[g]}}{N} \left(\bar{s}_b - \bar{s}_{b'}\right) x_{bb'}^{[g]} \right|$$

- Optimization vairable : $x_{bb'}^{[g]}$.
- $\bar{s}_b$: the midpoint score in the bin b.
- The objective : the product of the movement distance weighed by the fraction of total samples moved and the amount of movement.
- Overlap assumption: Each bin contains at least one member from each group.

Finding Optimal Solutions via Mixed Integer Programming (MIP)

Finding Optimal Solutions via Mixed Integer Programming (MIP)(1/6)

Non-convex quadratic constraints

- $\left|\frac{\sum_{b\in\mathcal{B}} x_{bb'}^{[1]} N_{b+}^{[1]}}{\sum_{b\in\mathcal{B}} x_{bb'}^{[1]} N_{b}^{[1]}} - \frac{\sum_{b\in\mathcal{B}} x_{bb'}^{[2]} N_{b+}^{[2]}}{\sum_{b\in\mathcal{B}} x_{bb'}^{[2]} N_{b}^{[2]}}\right| \leq \epsilon_{PRP} \quad \forall b' \in \mathcal{B}.$
- $\frac{\sum_{b\in\mathcal{B}} x_{bb'}^{[g]} N_{b+}^{[g]}}{\sum_{b\in\mathcal{B}} x_{bb'}^{[g]} N_{b}^{[g]}} \leq \frac{\sum_{b\in\mathcal{B}} x_{b\left(b'+1\right)}^{[g]} N_{b+}^{[g]}}{\sum_{b\in\mathcal{B}} x_{b\left(b'+1\right)}^{[g]} N_{b}^{[g]}} \quad \forall b' \in \{1, \ldots, B-1\}, g \in \mathcal{G}$

We overcome this with three tricks and MIP framework.

❶ We substitute the numerator and denominator to make the fraction a product of two continuous variables.

❷ We apply the normalized multiparametric disaggregation technique to convert our problem to a MILP.

❸ We enhance our MIP solution by tightening bounds through solving fractional linear programs.

▶ Our reformulation is necessary to convert this into a MILP.

Finding Optimal Solutions via Mixed Integer Programming (MIP)(2/6)

Step 1: Reducing the number of bilinear terms.

New optimization variables : $v_b^{[g]} \geq 0$ & $t_b^{[g]} \geq 0$

Let,

$$v_{b'}^{[g]} = \sum_{b \in \mathcal{B}} x_{bb'}^{[g]} N_b^{[g]} \quad \text{and} \quad t_{b'}^{[g]} v_{b'}^{[g]} = \sum_{b \in \mathcal{B}} x_{bb'}^{[g]} N_{b+}^{[g]} \quad \forall b' \in \mathcal{B}$$

Then we have the following:

$$\text{Fairness Constraint (3)} \iff \left| t_{b'}^{[1]} - t_{b'}^{[2]} \right| \leq \epsilon_{PRP} \quad \forall b' \in \mathcal{B},$$

$$\text{Non-convex Constraint (2)} \iff t_{b'}^{[g]} \leq t_{b'+1}^{[g]} \quad \forall b' \in \{1, \ldots, B-1\}$$

Finding Optimal Solutions via Mixed Integer Programming (MIP)(3/6)

Step 2-1 : NMDT(Normalized Multiparametric Disaggregation Technique)

We show how we can model each bilinear term $t_b^{[g]} v_b^{[g]}$.

▶ We make use of the **NMDT transformation**.

- any bounded optimization variable $x \in [x_L, x_U]$
- precision factor p (a negative integer)
- We can represent this variable exactly as $x = (x_U - x_L)\,\lambda + x_L$

$$\lambda = \sum_{l \in \{-p, \ldots, -1\}} 2^l z_l + \Delta\lambda$$

- $0 \leq \Delta\lambda \leq 2^p$ is a remainder term.
- $z_l \in \{0, 1\}$ are binary optimization variables.

▶ Dropping the remainder term $\Delta\lambda$, it can be effectively handled via modern MIP solvers.

Finding Optimal Solutions via Mixed Integer Programming (MIP)(4/6)

Step 2-2 : Bound tightening through fractional LP subproblems.

To apply NMDT, $v_b^{[g]}$ & $t_b^{[g]}$ must be bounded.

❶ bounds on $v_b^{[g]}$: a simple Linear Programming.

❷ bounds on $t_b^{[g]}$: nonlinear problem ▶ simple LP.
by Charnes-Cooper transformation

Charnes-Cooper transformation

$$\xi_{bb'}^{[g]} = \frac{x_{bb'}^{[g]}}{\sum_{b\in\mathcal{B}} x_{b\bar{b}}^{[\bar{g}]} N_b^{[\bar{g}]}} \quad \phi_{\bar{b}}^{[\bar{g}]} = \frac{1}{\sum_{b\in\mathcal{B}} x_{b\bar{b}}^{[\bar{g}]} N_b^{[\bar{g}]}}$$

We can express the min / max problem for $t_{\bar{b}}^{[\bar{g}]}$ as

$$\text{Min} \underset{\xi_{bb'}^{[g]}}{\text{or}} \text{Max} \quad t_{\bar{b}}^{[\bar{g}]} = \sum_{b\in\mathcal{B}} N_{b+}^{[\bar{g}]} \xi_{b\bar{b}}^{[\bar{g}]}$$

Finding Optimal Solutions via Mixed Integer Programming (MIP)((5/6))

The benefits of our reformulation from a non-convex QCQP to an MILP.

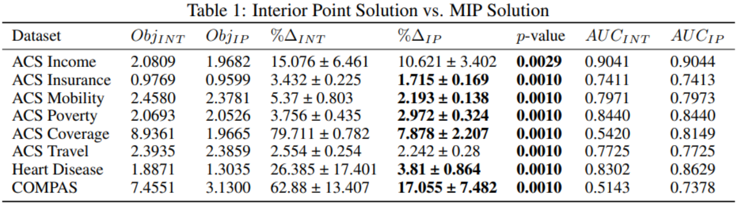


Table 1: Interior Point Solution vs. MIP Solution

Dataset	Obj_{INT}	Obj_{IP}	$\%\Delta_{INT}$	$\%\Delta_{IP}$	p-value	AUC_{INT}	AUC_{IP}
ACS Income	2.0809	1.9682	15.076 ± 6.461	10.621 ± 3.402	**0.0029**	0.9041	0.9044
ACS Insurance	0.9769	0.9599	3.432 ± 0.225	**1.715 ± 0.169**	**0.0010**	0.7411	0.7413
ACS Mobility	2.4580	2.3781	5.37 ± 0.803	**2.193 ± 0.138**	**0.0010**	0.7971	0.7973
ACS Poverty	2.0693	2.0526	3.756 ± 0.435	**2.972 ± 0.324**	**0.0010**	0.8440	0.8440
ACS Coverage	8.9361	1.9665	79.711 ± 0.782	**7.878 ± 2.207**	**0.0010**	0.5420	0.8149
ACS Travel	2.3935	2.3859	2.554 ± 0.254	2.242 ± 0.28	**0.0010**	0.7725	0.7725
Heart Disease	1.8871	1.3035	26.385 ± 17.401	**3.81 ± 0.864**	**0.0010**	0.8302	0.8629
COMPAS	7.4551	3.1300	62.88 ± 13.407	**17.055 ± 7.482**	**0.0010**	0.5143	0.7378

- AUC_{INT} & AUC_{IP} : the average result of applying the interior-point (INT) or integer programming (IP) method.
- The optimality gap : $\%\Delta = \frac{\text{Upper Bound-Lower Bound}}{\text{Upper Bound}}$
- Bold figures : the statistical significance of the improvement based the p-value from the Wilcoxon signed-rank test to determine if $\%\Delta_{IP} \leq \%_{INT}$ is a consistent result.

Finding Optimal Solutions via Mixed Integer Programming (MIP)(6/6)

The benefits of our reformulation from a non-convex QCQP to an MILP.

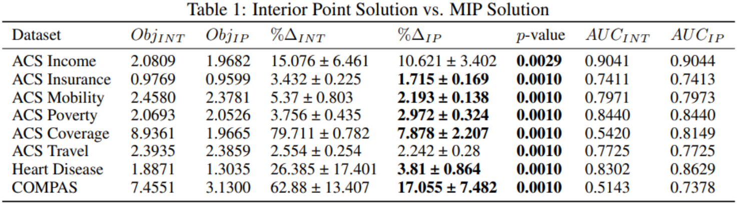


Table 1: Interior Point Solution vs. MIP Solution

Dataset	Obj_{INT}	Obj_{IP}	$\%\Delta_{INT}$	$\%\Delta_{IP}$	p-value	AUC_{INT}	AUC_{IP}
ACS Income	2.0809	1.9682	15.076 ± 6.461	10.621 ± 3.402	**0.0029**	0.9041	0.9044
ACS Insurance	0.9769	0.9599	3.432 ± 0.225	**1.715 ± 0.169**	**0.0010**	0.7411	0.7413
ACS Mobility	2.4580	2.3781	5.37 ± 0.803	**2.193 ± 0.138**	**0.0010**	0.7971	0.7973
ACS Poverty	2.0693	2.0526	3.756 ± 0.435	**2.972 ± 0.324**	**0.0010**	0.8440	0.8440
ACS Coverage	8.9361	1.9665	79.711 ± 0.782	**7.878 ± 2.207**	**0.0010**	0.5420	0.8149
ACS Travel	2.3935	2.3859	2.554 ± 0.254	2.242 ± 0.28	**0.0010**	0.7725	0.7725
Heart Disease	1.8871	1.3035	26.385 ± 17.401	**3.81 ± 0.864**	**0.0010**	0.8302	0.8629
COMPAS	7.4551	3.1300	62.88 ± 13.407	**17.055 ± 7.482**	**0.0010**	0.5143	0.7378

1. Solving a MIP method yields lower bounds.
2. We can always theoretically continue improving the solution to optimality based on the acceptable time limits.

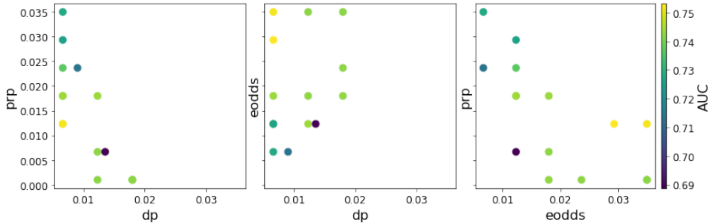
Applications

Applications(1/5)

To apply our framework

We develop an efficient frontier of fairness solutions over a grid of parameter $\epsilon_{DP}, \epsilon_{EOdds}, \epsilon_{PRP}$.

Figure 1: Efficient frontier of solutions for ACS West Insurance data



We show the 2-d profile shots of our 4-d fairness surface in Figure 1.

- We could obtain a true Pareto-optimal frontier using IPOPT.

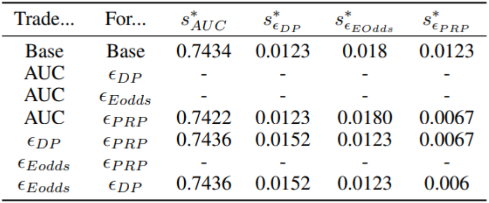
Applications(2/5)

Understanding fairness tradeoffs

Pick an point s based on the desired AUC & tolerable fairness violations ϵ.

▶ If we want to trade AUC for ϵ_{EOdds}, then we would find a point s' with at least as good ϵ_{DP} , ϵ_{PRP} but worse AUC and better ϵ_{EOdds}.

Table 2: Performance Fairness Trade-off Analysis



Trade...	For...	s^*_{AUC}	$s^*_{\epsilon_{DP}}$	$s^*_{\epsilon_{EOdds}}$	$s^*_{\epsilon_{PRP}}$
Base	Base	0.7434	0.0123	0.018	0.0123
AUC	ϵ_{DP}	-	-	-	-
AUC	ϵ_{Eodds}	-	-	-	-
AUC	ϵ_{PRP}	0.7422	0.0123	0.0180	0.0067
ϵ_{DP}	ϵ_{PRP}	0.7436	0.0152	0.0123	0.0067
ϵ_{Eodds}	ϵ_{PRP}	-	-	-	-
ϵ_{Eodds}	ϵ_{DP}	0.7436	0.0152	0.0123	0.006

- First row : hypothetical opertaing point.
- The following row : point after trading.
- Blank row : no such point.

Applications(3/5)

Performance comparison

We compare our framework against Rezaei & Pleiss methods.

1. Get the base scores $\hat{y}_0$.
2. Compare $\rightarrow$ get method scores $\hat{y}_m$.
3. Bin the base output.
4. Compute the AUC with (ϵ_0, ϵ_m).
5. Solve the constrained optimization problem.
6. $\epsilon = \frac{1}{2}\min(\epsilon_0, \epsilon_m)$.
7. Apply the optimal solutions $x_{bb'}^{[g]}$.
8. Assign a group g instance with score $s \in b$.
9. Compute the resulting AUC and fairness metrics on remapped bins.

- These figures are shown in Table 3.

Applications(4/5)

Performance comparison

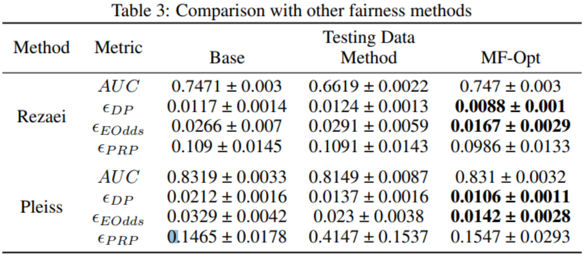


Table 3: Comparison with other fairness methods

Method	Metric	Testing Data		
		Base	Method	MF-Opt
Rezaei	AUC	0.7471 ± 0.003	0.6619 ± 0.0022	0.747 ± 0.003
	ϵ_{DP}	0.0117 ± 0.0014	0.0124 ± 0.0013	**0.0088 ± 0.001**
	ϵ_{EOdds}	0.0266 ± 0.007	0.0291 ± 0.0059	**0.0167 ± 0.0029**
	ϵ_{PRP}	0.109 ± 0.0145	0.1091 ± 0.0143	0.0986 ± 0.0133
Pleiss	AUC	0.8319 ± 0.0033	0.8149 ± 0.0087	0.831 ± 0.0032
	ϵ_{DP}	0.0212 ± 0.0016	0.0137 ± 0.0016	**0.0106 ± 0.0011**
	ϵ_{EOdds}	0.0329 ± 0.0042	0.023 ± 0.0038	**0.0142 ± 0.0028**
	ϵ_{PRP}	0.1465 ± 0.0178	0.4147 ± 0.1537	0.1547 ± 0.0293

- Adventages of our framework over in-processing framework :
  - ▶ MFOpt can be applied on top of any model class.
- Disadventages of post-processing framework :
  - ▶ Pleiss method results in large violations of bin-wise PRP.

Applications(5/5)

Who's the fairest of them all?

Describe a method of gauging a model's efficiency in trading between difference fairness definitions.

1. Construct the frontier and then filter all points on the efficient frontier with tolerable performance. $AUC \geq AUC_{min}$
2. Find the point on the respective frontiers with minimum Euclidean distance to the origin.

The model with the shorter distance can then be declared as the model that has better tradeoff properties.