Notes on "Delayed Impact of Fair Machine Learning" [1]

Dhruv Medarametla* dhruvm2@stanford.edu

Anav Sood* anavsood@stanford.edu

December 8, 2018

1 Problem Setting

The paper we study considers two *groups*, A and B, which make up $g_{\mathsf{A}} \in [0,1]$ and $g_{\mathsf{B}} := 1 - g_{\mathsf{A}}$ proportion of the population. Throughout our review, we consider A to be disadvantaged compared to B. Each individual has a *score* $\mathcal{X} := [C]$, an abstract quantity summarizing how well an individual is suited to being *accepted* (or not *rejected*) by some *institution*. For group $\mathsf{j} \in \{\mathsf{A}, \mathsf{B}\}$, scores are distributed according to $\boldsymbol{\pi}_{\mathsf{j}} \in \text{Simplex}^{C-1}$. For each group, the institution adopts a *policy* $\boldsymbol{\tau}_{\mathsf{j}} : \mathcal{X} \to [0,1]$, where $\boldsymbol{\tau}_{\mathsf{j}}(x)$ represents the probability of accepting a person in group j with score x. The policy $\boldsymbol{\tau}_{\mathsf{j}}$ can be seen as a vector in $[0,1]^C$ with $\boldsymbol{\tau}_{\mathsf{j}}(x)$ as the xth component. The institution then adopts an overall policy $\boldsymbol{\tau} = (\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}})$ while attempting to maximize its *utility*

$$\mathcal{U}(\boldsymbol{\tau}) := \sum_{\mathsf{j} \in \{\mathsf{A},\mathsf{B}\}} g_{\mathsf{j}} \sum_{x \in \mathcal{X}} \boldsymbol{\tau}_{\mathsf{j}}(x) \boldsymbol{\pi}_{\mathsf{j}}(x) \boldsymbol{u}(x)$$

where the authors have assumed the existence of such a reasonable $\boldsymbol{u} : \mathcal{X} \to \mathbb{R}$ which models the utility contribution of a individual with score x. We also let $\mathcal{U}_{\mathsf{j}}(\boldsymbol{\tau}) := \sum_{x \in \mathcal{X}} \boldsymbol{\tau}_{\mathsf{j}}(x) \boldsymbol{\pi}_{\mathsf{j}}(x) \boldsymbol{u}(x)$.

The crux of the paper's interest is in the socially motivated *outcome* for groups A and B. Namely, the authors assume existence of $\Delta : \mathcal{X} \to \mathbb{R}$, where $\Delta(x)$ represents the abstract change of well-being for an individual of score x should they be accepted. They then define change in the mean outcome of group j due to a policy $\boldsymbol{\tau}$

$$\Delta \boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau}) := \sum_{x \in \mathcal{X}} \boldsymbol{\pi}_{\mathsf{j}}(x) \boldsymbol{\tau}_{\mathsf{j}}(x) \Delta(x)$$

Note that this metric of social impact does not take into account those who have not been accepted; it is possible to modify $\Delta(x)$ in order to do so.

Lastly, the authors define a *success* function $\boldsymbol{\rho} : \mathcal{X} \to [0,1]$ where $\boldsymbol{\rho}(x)$ is the probability that an individual with score x succeeds given that they were accepted. It is often intuitive and useful to define $\Delta(x)$ and $\boldsymbol{u}(x)$ using $\boldsymbol{\rho}$. Also, note that in an initial effort to maintain fairness, $\boldsymbol{u}(x)$, $\Delta(x)$, and $\boldsymbol{\rho}(x)$, all take values independent of the group the individual is from.

1.1 Examples

Middle Level Management. Consider a large tech company looking to promote current employees to middle management. Scores $x \in [C]$ encode a current employee's managerial ability. The company decides to promote or not promote each individual according to their policy $\boldsymbol{\tau}$. For an employee with score x, company and personal utility are naturally dependent on $\boldsymbol{\rho} : x \mapsto \boldsymbol{\rho}(x) \in [0,1]$, the probability the employee will be a successful manager. Given $\boldsymbol{\rho}(x)$, we could define $\boldsymbol{u}(x) = 50\boldsymbol{\rho}(x) - 150(1 - \boldsymbol{\rho}(x))$, representing the idea that a successful manager increases the company's utility by 50, while an unsuccessful one decreases it by 150. The individual impact of acceptance can be directly represented by change in income, in which case we can define $\Delta(x) = 25000$, representing promotion incurring a raise of 25000.

Now let the two groups of people be A, women, and B, men, with $g_{\mathsf{A}} = 0.2$ and $g_{\mathsf{B}} = 0.8$. Note that $\mathbb{E}_{\mathsf{A}}[x]$ might be slightly less than $\mathbb{E}_{\mathsf{B}}[x]$ due to years of unconscious bias and societal pushback, something to consider when determining a fair $\boldsymbol{\tau}$.

American Immigration Policies. Consider America as an institution deciding to accept or reject immigrants. The USA may accept immigrants according to a score $x \in [C]$, encoding an "American preparedness" measure. Immigrants may succeed (finding a job above a certain income and paying taxes) with some probability $\boldsymbol{\rho}(x)$. The country may experience a utility $\boldsymbol{u}(x)$ corresponding to a positive or negative change in GDP depending on the immigrants success. We can consider $\Delta(x)$ to be an immigrant's change in income minus a fixed positive factor accounting for the turmoil and uncertainty of emigration. In this scenario, let the two groups of people be A, refugees, and B, non-refugees, with $g_{\mathsf{A}} = 0.15$ and $g_{\mathsf{B}} = 0.85$. Then, this model effectively represents the immigration policy when selecting for new residents among these two groups; most likely, $\mathbb{E}_{\mathsf{A}}[x] < \mathbb{E}_{\mathsf{B}}[x]$ due to the situations that refugee immigrants come from, something to consider when determining a fair $\boldsymbol{\tau}$.

1.2 Possible Constraints

In an attempt to force the chosen policy $\boldsymbol{\tau} \in [0,1]^{2C}$ to be more fair to the disadvantaged group, humans have historically constrained $\boldsymbol{\tau}$ by demanding $\boldsymbol{\tau} \in \mathcal{C} \subset [0,1]^{2C}$. Below, we present the examples studied in the paper.

1. `MaxUtil`: The `MaxUtil` policy simply corresponds to $\mathcal{C} := [0,1]^{2C}$. The only imposed fairness constraint is that $\boldsymbol{u}(x)$ is independent of group.
2. `DemParity`: The `DemParity` policy demands that an equal proportion of each group be accepted by the institution. Namely, $\mathcal{C} := \{(\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}}) : \sum_{x\in\mathcal{X}} \boldsymbol{\pi}_{\mathsf{A}}(x)\boldsymbol{\tau}_{\mathsf{A}}(x) = \sum_{x\in\mathcal{X}} \boldsymbol{\pi}_{\mathsf{B}}(x)\boldsymbol{\tau}_{\mathsf{B}}(x)\}$
3. `EqOpt`: The `EqOpt` policy demands that and equal proportion of people who are likely to be successful must be chosen from each group. Namely, if we define $\mathrm{TPR}_{\mathsf{j}}(\boldsymbol{\tau}) := \frac{\sum_{x\in\mathcal{X}} \boldsymbol{\pi}_{\mathsf{j}}(x)\boldsymbol{\rho}(x)\boldsymbol{\tau}(x)}{\sum_{x\in\mathcal{X}} \boldsymbol{\pi}_{j}(x)\boldsymbol{\rho}(x)}$, then $\mathcal{C} := \{(\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}}) : \mathrm{TPR}_{\mathsf{A}}(\boldsymbol{\tau}_{\mathsf{A}}) = \mathrm{TPR}_{\mathsf{B}}(\boldsymbol{\tau}_{\mathsf{B}})\}$. To gain a little more intuition for this expression, we note that $\mathrm{TPR}_{\mathsf{A}}(\boldsymbol{\tau})$ is equal to $\mathcal{P}_{a\in\mathsf{A}}[a \text{ selected}|a \text{ successful}]$, and correspondingly for B.

Considering the policy $\boldsymbol{\tau}^{\texttt{MaxUtil}}$ achieved as the optimal solution to the `MaxUtil` constraint, for a policy $\boldsymbol{\tau}$ the authors say that $\boldsymbol{\tau}$ causes *relative harm* to group j if $\Delta\boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) < \Delta\boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau}^{\texttt{MaxUtil}})$ and *relative improvement* if $\Delta\boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) > \Delta\boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau}^{\texttt{MaxUtil}})$. In the same vein, the authors say that $\boldsymbol{\tau}$ causes *active harm* if$\Delta\boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) < 0$, *improvement* if $\Delta\boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) > 0$, and *stagnation* if $\Delta\boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) = 0$.

2 Reduction to Threshold Policies

One of the paper's large technical insights was a methodology for reducing the high parameter problem of picking a policy $\boldsymbol{\tau}$ to a one parameter problem given by our *selection rate function* $r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}) := \sum_{x\in\mathcal{X}} \boldsymbol{\pi}_{\mathsf{j}}(x)\boldsymbol{\tau}_{\mathsf{j}}(x)$. Note when picking a policy $\boldsymbol{\tau}_{\mathsf{j}}$ for group j, the institution only cares about the values $\boldsymbol{\tau}_{\mathsf{j}}$ takes on scores for which $\boldsymbol{\pi}_{\mathsf{j}}$ is non-zero, inducing an equivalence under $\boldsymbol{\pi}_{\mathsf{j}}$. Two equivalent policies provide identical utility for the institution, outcome (so our analysis of a policy is the same up to equivalence), true positive rate, and selection rate.

Now, as a thought experiment, let's consider picking a policy under the `MaxUtil` constraint. It is easy to see from $\mathcal{U}(\boldsymbol{\tau})$ that an optimal policy would accept any individual with score $\boldsymbol{u}(x) > 0$. Then, the reasonable assumption that $\boldsymbol{u}(x)$ is strictly increasing in x would imply that anyone with score above some threshold score should be accepted. Under a `DemParity` constraint, we must select an equal proportion β of each group. Again if $\boldsymbol{u}(x)$ is strictly increasing in x, it makes sense to select the β proportion of individuals in each population with the highest scores, yielding a similar threshold. This motivates the definition of a *threshold policy*, a policy $\boldsymbol{\tau}_{c,\gamma}$ of the form

$$\boldsymbol{\tau}_{c,\gamma} = \begin{cases} 1, & \text{if } x > c \\ \gamma, & \text{if } x = c \\ 0, & x < c \end{cases}$$

for some $c \in \mathcal{X}$ and $\gamma \in [0,1]$. The authors were able to then show these intuitive, but important results:

Result One. (Lemma 5.1) Let $\boldsymbol{\tau}_{\mathsf{j}}$ and $\boldsymbol{\tau}'_{\mathsf{j}}$ be threshold policies. Then $\boldsymbol{\tau}_{\mathsf{j}} \simeq_{\boldsymbol{\pi}_{\mathsf{j}}} \boldsymbol{\tau}'_{\mathsf{j}}$ if and only if $r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}) = r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}'_{\mathsf{j}})$. Further, $r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}})$ is a bijection from $\mathcal{T}_{\text{thresh}}(\boldsymbol{\pi}_{\mathsf{j}})$ to $[0,1]$ where $\mathcal{T}_{\text{thresh}}(\boldsymbol{\pi}_{\mathsf{j}})$ is the set of equivalences classes between threshold policies under $\simeq_{\boldsymbol{\pi}_{\mathsf{j}}}$.
Result Two. (Proposition 5.2) Suppose $\boldsymbol{u}(x)$ is strictly increasing in x. Then all optimal `MaxUtil` and `DemParity` policies $(\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}})$ satisfy $\boldsymbol{\tau}_{\mathsf{j}} \simeq_{\boldsymbol{\pi}_{\mathsf{j}}} r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}))$ for $\mathsf{j} \in \{\mathsf{A}, \mathsf{B}\}$. If in addition $\boldsymbol{u}(x)/\boldsymbol{\rho}(x)$ is increasing, the same is true for all optimal `EqOpt` policies.

Thus, all optimal policies are effectively threshold policies. Combining these results, the authors have reduced from optimizing over all policies to optimizing over equivalences classes of threshold policies, which can be uniquely characterized by their selection rate.

3 Concavity of Utility Functions

The authors have re-parameterized the problem from $2C$ parameters to 2: the proportion of each group the institution accepts. The authors now show the optimal parameters β_{j} for utility are unique under the `MaxUtil` constraint, the `DemParity` constraint, the `EqOpt` constraint, and also for the social outcome function by proving that all of these functions are concave in β_{j} or a similar parameter. The first important result is the following:

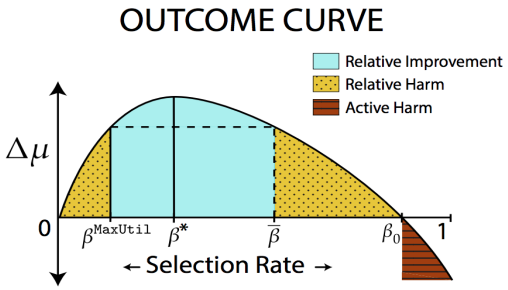
Result Three (Proposition 5.3). Let $\boldsymbol{\pi} \in \text{Simplex}^{C-1}$ be a distribution over C states. Then if $\boldsymbol{w}(x)$ is any non-decreasing map from $\mathcal{X} \to \mathbb{R}$, $\beta \to \langle \boldsymbol{w}, \boldsymbol{\pi} \circ r^{-1}_{\boldsymbol{\pi}}(\beta) \rangle$ is concave.

Setting $\boldsymbol{w}(x) = \Delta(x)$, we get that the function $\beta_{\mathsf{A}} \to \Delta\boldsymbol{\mu}_{\mathsf{A}}(r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\beta_{\mathsf{A}}))$ [social outcome function] is concave. Setting $\boldsymbol{w}(x) = \boldsymbol{u}(x)$, we get that $\beta_{\mathsf{j}} \to \mathcal{U}_{\mathsf{j}}(r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}(\beta_{\mathsf{j}}))$ for a fixed j [`MaxUtil` function] is concave, and $\beta \to g_{\mathsf{A}}\mathcal{U}_{\mathsf{A}}(r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\beta)) + g_{\mathsf{B}}\mathcal{U}_{\mathsf{B}}(r^{-1}_{\boldsymbol{\pi}_{\mathsf{B}}}(\beta)) = \mathcal{U}((r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\beta), r^{-1}_{\boldsymbol{\pi}_{\mathsf{B}}}(\beta)))$ [`DemParity` function] is concave as well using properties of concave functions. The authors then prove a similar result for the `EqOpt` constraint.

Result Four (Lemma 6.1, B.1). For $\mathsf{j} \in \{\mathsf{A}, \mathsf{B}\}$, the function $T_{\mathsf{j}}(\beta) = \langle r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}(\beta), \boldsymbol{\pi}_{\mathsf{j}} \circ \boldsymbol{u} \rangle$ is a bijection from $[0,1]$ to $[0, \langle \boldsymbol{\pi}_{\mathsf{j}} \circ \boldsymbol{u} \rangle]$. Additionally, the function $t \to \mathcal{U}_{\mathsf{j}}(r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}(T^{-1}_{\mathsf{j}}(t)))$ is concave in t.

Recall that the `EqOpt` policy forces $\text{TPR}_{\mathsf{A}}(\boldsymbol{\tau}_{\mathsf{A}}) = \text{TPR}_{\mathsf{B}}(\boldsymbol{\tau}_{\mathsf{B}})$; this is equivalent to having $T_{\mathsf{A}}(\beta_{\mathsf{A}}) = T_{\mathsf{B}}(\beta_{\mathsf{B}})$. Then, the first part of the result 4 implies that it is enough to prove that $t \to g_{\mathsf{A}}\mathcal{U}_{\mathsf{A}}(r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(T^{-1}_{\mathsf{A}}(t))) + g_{\mathsf{B}}\mathcal{U}_{\mathsf{B}}(r^{-1}_{\boldsymbol{\pi}_{\mathsf{B}}}(T^{-1}_{\mathsf{B}}(t))) = \mathcal{U}((r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(T^{-1}_{\mathsf{A}}(t)), r^{-1}_{\boldsymbol{\pi}_{\mathsf{B}}}(T^{-1}_{\mathsf{B}}(t))))$ [`EqOpt` function] is concave, which follows from the second part of result 4. The authors have thus shown that the social outcome function and the three constrained utility functions are concave, either as a function of β_{j} or a bijection to β_{j}. Since each function is concave, we have the existence of a unique interval of optimal β_{j}. Thus we can compare each constraint's social impact by examining $\Delta\mu_{\mathsf{A}}(\beta_{\mathsf{A}})$ for β_{A} in the constraint's respective optimal interval.

Additionally, we can better understand our social outcome function. Under the reasonable assumption that $\boldsymbol{u}(x) > 0 \implies \Delta(x) > 0$ (the institution takes on more risk than the individual) it is clear that $\beta^{\texttt{MaxUtil}} \leq \beta^*$. Then, we can define $\bar{\beta} \geq \beta^*$ such that $\mathcal{U}_{\mathsf{A}}(r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\bar{\beta})) = \mathcal{U}_{\mathsf{A}}(r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\beta^{\texttt{MaxUtil}}))$, and $\beta_0 \geq \bar{\beta}$ such that $\mathcal{U}_{\mathsf{A}}(r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\beta_0)) = 0$. The image on the right (from the paper) summarizes this paragraph pictorially by considering a graph of the social outcome function.



4 Main Results

In Section 6, the authors provide two theorems which give a robust characterization where the optimal selection rate parameter β lies for group A under `DemParity` and `EqOpt` constraints. These Theorems imply the

Corollaries in Section 3, which provide important insight into how such social constraints impact group A. Letting $G^{(\mathsf{A}\to\mathsf{B})}(\beta_{\mathsf{A}}) := T_{\mathsf{B}}^{-1}(T_{\mathsf{A}}(\beta_{\mathsf{A}}))$ be the transfer function as described in the paper, we have that

Result Five (Corollary 3.2). (a) Under the assumption that $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \overline{\beta}$ and $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > \beta_{\mathsf{A}}^{\texttt{MaxUtil}}$, there exist population proportions $g_0 < g_1 < 1$ such that, for all $g_{\mathsf{A}} \in [g_0, g_1]$, $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta_{A}^{\texttt{DemParity}} < \overline{\beta}$. That is, `DemParity` causes relative improvement.
(b)Under the assumption that there exist $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta < \beta' < \overline{\beta}$ such that $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > G^{(\mathsf{A}\to\mathsf{B})}(\beta), G^{(\mathsf{A}\to\mathsf{B})}(\beta')$, there exist population proportions $g_2 < g_3 < 1$ such that, for all $g_{\mathsf{A}} \in [g_2, g_3]$, $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta_{A}^{\texttt{EqOpt}} < \overline{\beta}$. That is, `EqOpt` causes relative improvement.

Result Six (Corollary 3.3, 3.4). (a) Fix a selection rate β. Assume that $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > \beta > \beta_{\mathsf{A}}^{\texttt{MaxUtil}}$. Then there exists a population proportion g_0 such that, for all $g_{\mathsf{A}} \in [0, g_0]$, $\beta_{\mathsf{A}}^{\texttt{DemParity}} > \beta$. In particular, when $\beta = \beta_0$, `DemParity` causes active harm, and when $\beta = \overline{\beta}$, `DemParity` causes relative harm.
(b) Suppose that $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > G^{(\mathsf{A}\to\mathsf{B})}(\beta)$ and $\beta > \beta_{\mathsf{A}}^{\texttt{MaxUtil}}$. Then there exists a population proportion g_0 such that, for all $g_{\mathsf{A}} \in [0, g_0]$, $\beta_{\mathsf{A}}^{\texttt{EqOpt}} > \beta$. In particular, when $\beta = \beta_0$, `EqOpt` causes active harm, and when $\beta = \overline{\beta}$, `EqOpt` causes relative harm.

Corollary 3.5 shows that `DemParity` can cause active harm while `EqOpt` causes improvement. We see this is intuitive as, under `DemParity`, if scores of group B are particularly high in comparison to low scores in group A, maximizing utility still gives large β due to the massive increase in utility given by accepting much of group B. The `DemParity` constraint then forces many members of group A with negative $\Delta(x)$ to also be accepted. The `EqOpt` constraint avoids this by not forcing equal acceptance rates. We instead present and examine Corollary 3.6, which is much more non-intuitive.

Result Seven (Corollary 3.6) Suppose it is the case that $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta_{\mathsf{B}}^{\texttt{MaxUtil}}$ and $\text{TPR}_{\mathsf{A}}(\boldsymbol{\tau}^{\texttt{MaxUtil}}) > \text{TPR}_{\mathsf{B}}(\boldsymbol{\tau}^{\texttt{MaxUtil}})$. Then $\beta_{\mathsf{A}}^{\texttt{EqOpt}} < \beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta_{\mathsf{A}}^{\texttt{DemParity}}$. That is, `EqOpt` causes relative harm.

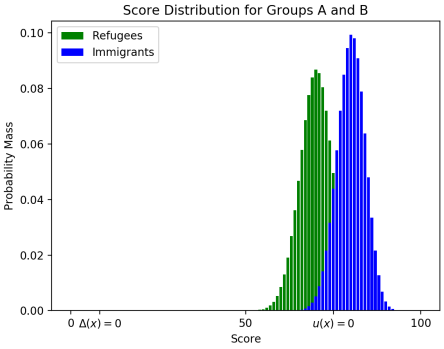
The above Corollary shows it is possible for `EqOpt` to select less of group A than `MaxUtil`, meaning `EqOpt` causes relative harm. This is a surprising result, and has interesting social implications surrounding how inequality in a disadvantaged group can reverse the intended effect of the `EqOpt` constraint. See Application 2 below for a concrete example of this result.

5 Applications

We go over two case studies that study how $\beta_{\mathsf{j}}^{\texttt{MaxUtil}}$, $\beta_{\mathsf{j}}^{\texttt{DemParity}}$, $\beta_{\mathsf{j}}^{\texttt{EqOpt}}$, and β_{j}^{*} differ in concrete examples.

5.1 American Immigration Policies: An Example of Corollary 3.2

Recall again our example of immigrants coming to America. Say that every immigrant can be assigned a score $x \in \mathcal{X}$, where $\mathcal{X} = \{0, 1, \ldots, 99, 100\}$, that measures how prepared they are to become a citizen of America. Assume that we can calculate $\boldsymbol{\rho}(x) = 0.35 + 0.006x$, $\boldsymbol{u}(x) = 20\boldsymbol{\rho}(x) - 80(1 - \boldsymbol{\rho}(x))$, and $\Delta(x) = 30\boldsymbol{\rho}(x) - 20(1 - \boldsymbol{\rho}(x))$.
In other words, we are saying that America gains utility of 20 for every successful immigrant and loses utility of 80 for every unsuccessful immigrant, and that an immigrant gains 30 utility if they are successful and loses 20 utility if they are unsuccessful in America (these parameters are estimated quantities; the true values may differ significantly).



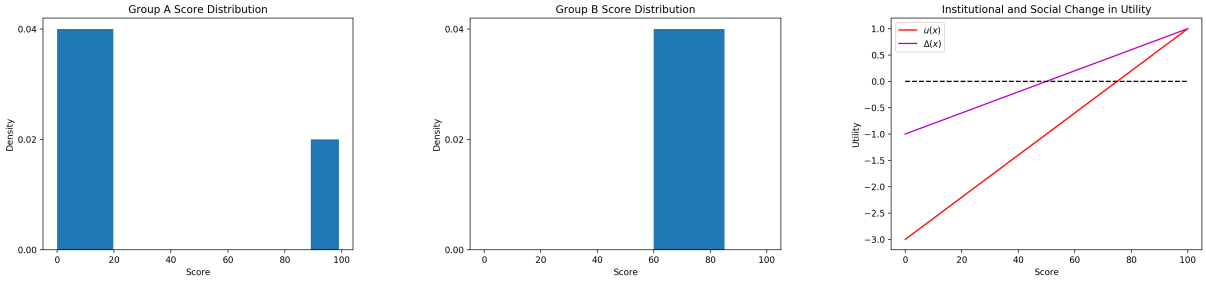
Now, say we have two subpopulations: A, refugees and B,
non-refugees, with $g_{\mathsf{A}} = 0.15$ and $g_{\mathsf{B}} = 0.85$. Say that the scores of the refugees are distributed according to $\mathcal{B}(100, 0.7)$, where $\mathcal{B}(n, p)$ is the discrete binomial distribution with. Similarly, say that the scores of the non-refugees are distributed according to $\mathcal{B}(100, 0.8)$. We display this in the figure above, marking where $\boldsymbol{u}(x)$ and $\Delta(x)$ are 0. Given these set of parameters, we can calculate the optimal β_{j} for each policy, as well as institutional utilization and social outcome. Our results are summarized below:

Application 1: American Immigration Policies				
Parameter	`MaxUtil`	`DemParity`	`EqOpt`	Social Max
β_{A}	0.163	0.811	0.809	> 0.999
β_{B}	0.913	0.811	0.811	> 0.999
$\mathcal{U}(\boldsymbol{\tau})$	2.690	2.381	2.392	2.100
$\Delta\boldsymbol{\mu}_{\mathsf{A}}(\boldsymbol{\tau}_{\mathbf{A}})$	3.35	15.38	15.35	18.50

In this case, we can see that group B dominates due to the high value of g_{B}; as a result, β_{B} is relatively similar for the three constraints, but β_{A} varies noticeably. This application displays that for many reasonable modeling scenarios, parts (a) and (b) of Corollary 3.2 are at work, as both `DemParity` and `EqOpt` cause relative improvement for group A.

5.2 Application 2: Understanding Corollary 3.6

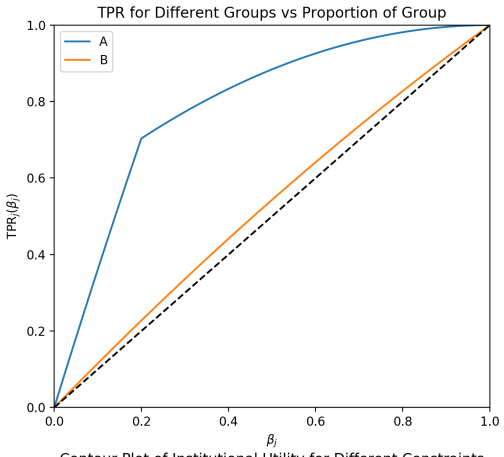
Consider the following abstract example, which is set in the natural continuous extension of the problem statement. [1] We let $g_{\mathsf{A}} = 0.25$, $g_{\mathsf{B}} = 0.75$. We let $\mathcal{X} = [0, 100]$ with the scores of A distributed with 80% of data uniform in $[0, 20]$ and 20% of data uniform in $[90, 100]$ and B distributed with all data uniformly in $[60, 85]$. We let $\boldsymbol{\rho}(x) = x/100$, $\boldsymbol{u}(x) = \boldsymbol{\rho}(x) - 3(1 - \boldsymbol{\rho}(x)) = -3 + x/25$, and $\Delta(x) = \boldsymbol{\rho}(x) - (1 - \boldsymbol{\rho}(x)) = -1 + x/50$. We provide both plots visually depicting this and tabulated results below:

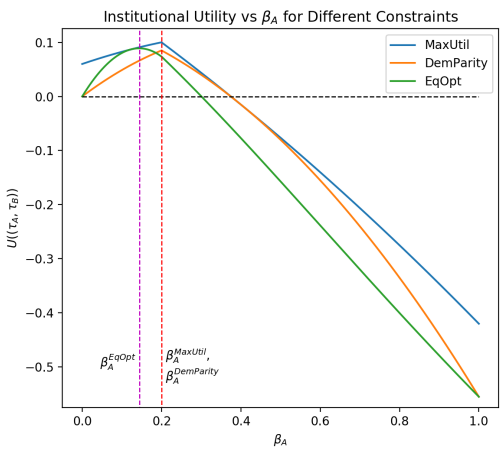


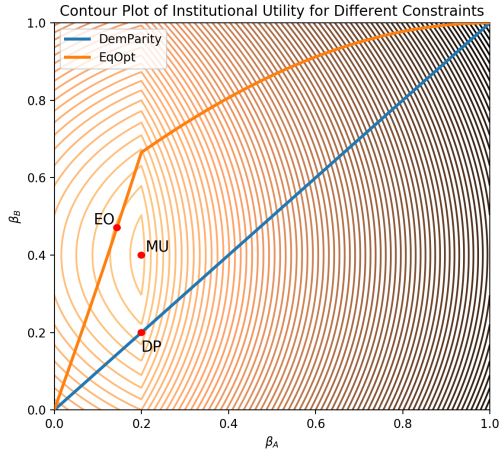
Application 2: High Inequality in Group A				
Parameter	`MaxUtil`	`DemParity`	`EqOpt`	Social Max
β_{A}	0.2	0.2	0.144	0.2
β_{B}	0.4	0.2	0.471	1.0
$\mathcal{U}(\boldsymbol{\tau})$	0.1	0.085	0.089	-0.035
TPR_{A}	0.704	0.704	0.514	0.704
TPR_{B}	0.441	0.228	0.514	1.0
$\Delta\boldsymbol{\mu}_{\mathsf{A}}(\boldsymbol{\tau}_{\mathbf{A}})$	0.18	0.18	0.133	0.18

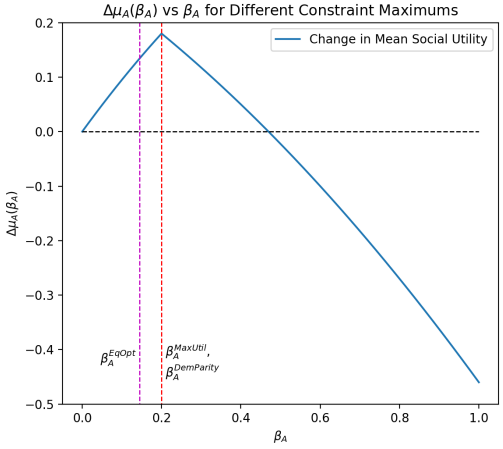
Due to inequality in group A, a large amount of group A's potential success is concentrated among high-achievers. The success of group B, however, is more evenly distributed. Thus, under the constraint that $\text{TPR}_{\mathsf{A}} = \text{TPR}_{\mathsf{B}}$, even a small β_{A} naturally results in a large β_B, causing the policy to accept individuals from group B who cause a decrease in $\mathcal{U}$. Thus we see $\beta_A^{\texttt{EqOpt}} < \beta_{\mathsf{A}}^{\texttt{MaxUtil}}$, meaning that `EqOpt` is causing relative harm to group A. This can be very clearly visually seen in our plots below:

[1] Treating scores and such continuously lends for easier computation, while still allowing us to explain the impact and importance of Corollary 3.6. The intuition transfers to the discrete case directly.









6 Limitations and Future Work

We first discuss some possibilities for future work.

1. Consider Corollary 3.6. In the scenario outlined by our second example, can we prevent `EqOpt` from relatively harming group A by selecting $\beta'_{\mathsf{B}} = \min\{(1+\epsilon)\beta_{\mathsf{A}}, \beta_{\mathsf{B}}\}$ for group B?

2. There is some study of an outcome-based alternative, where the institution solves the optimization problem $\max_{\boldsymbol{\tau}_{\mathsf{A}}} \Delta_{\mathsf{A}}\boldsymbol{\mu}_{\mathsf{A}}(\boldsymbol{\tau}_{\mathsf{A}})$ s.t. $\mathcal{U}^{\texttt{MaxUtil}} - \mathcal{U}(\boldsymbol{\tau}) < \delta$. Essentially, the institution prioritizes the outcome of group A so long as they stay within a δ budget. We wonder if $\max_{\boldsymbol{\tau}} \mathcal{U}(\tau) + \lambda\Delta\boldsymbol{\mu}_{\mathsf{A}}(\boldsymbol{\tau}_{\mathsf{A}})$ is a more natural way to phrase the problem (in fact we believe the best way), where we have given the institution incentive (public image, tax returns etc.) to help the disadvantaged group. How similar are these two problem statements? Do they offer differing results?

3. The topic of non-selection is only touched on in one footnote. The authors let $\Delta_a(x)$ and $\Delta_n(x)$ represent the effect of being accepted and rejected respectively, then let $\Delta\mu_{\mathsf{j}}(\boldsymbol{\tau}) := \sum_{x\in\mathcal{X}} \boldsymbol{\pi}_{\mathsf{j}}(x)(\boldsymbol{\tau}_{\mathsf{j}}(x)\Delta_a(x) + (1-\boldsymbol{\tau}_{\mathsf{j}}(x))\Delta_n(x))$. Our above results hold if $\Delta_a(x) - \Delta_n(x)$ is increasing in x; however, this is a rather nuanced assumption. For example, consider the case of immigrants, where a low-scoring immigrant will likely have a larger value for $\Delta_a(x) - \Delta_n(x)$ due to a very negative $\Delta_n(x)$. We wonder what can be said about non-selection without this assumption.

With regard to the work itself, it is a rather complete study of the problem setup. That being said, it applies to one epoch, or one cycle of being accepted and denied. Helping disadvantaged groups requires the proper constraints to be employed over long periods of time and monitored throughout multiple epochs. This likely requires a more restrictive and focused study per particular issue. Additionally it requires an expert to be incredibly confident in setting $\boldsymbol{u}, \Delta$, and $\boldsymbol{\rho}$ functions, which offers a lot of modeling freedom, much of which we exploited in making our examples. Currently, at best, the results provide very solid intuition to guide more focused and less general work for specific issues.

References

[1] Liu, L. T., Dean, S., Rolf, E., Simchowitz, M., and Hardt, M. (2018). Delayed impact of fair machine learning. In International Conference on Machine Learning.

[2] Bertrand, M., Chugh, D., and Mullainathan, S. (2005). Implicit discrimination. The American Economic Review, 95(2):94–98.