Semester 6 – Jan'25–Apr'25

UGP REPORT

AlgoBias: Algorithmic Matching in Biased Settings

Mentee: Dhruv Rai (Roll no.-220365)

Mentor: Prof. Sayak Ray Chowdhury

Department of Computer Science and Engineering IIT Kanpur

1 Centralized Selection in the Presence of Biases[1]

1.1 Introduction

This section is focused on a centralized selection problem which has p institutions, each with capacities $k_1, k_2, \ldots, k_p$, and n candidates, and $\sum_i k_i$ is much less than n. A central entity evaluates the utility of each candidate to the institutions (same for each institution) and the candidates declare preferences for the institutions. Given utilities and preferences, the high-level goal is to select a subset of at most k_l candidates for the l-th institution while trying to maximize total utility and ensuring the preferences of the candidates are taken into account. Candidates' utilities are denoted as $u_1, u_2, \ldots, u_p$ and they denote candidates' test scores, performance in interviews, etc. Candidate i's preferences are given by a ranking σ_i of the p institutions, ordered from most to least preferred.

The most interesting part of the setting is the fact that the true utilities are not observed and central institution try to find estimates for the true utilities of candidates through some kind of evaluation process like exams or interviews etc. These evaluation methods are known to yield biased estimates against underprivileged groups. We first try to evaluate the affect of this bias on the fairness and utility matrix and then learn about the algorithm that can minimize the loss in these matrices due to bias.

1.2 Model of bias in utilities

We consider that the candidates are divided into two groups: the advantaged group G_1 and the disadvantaged group G_2. The estimated utility u_i for a disadvantaged candidate is β times their latent utility, i.e., $\hat{u}_i = \beta * u_i$, and that of an advantaged candidate j is the same as their latent utility, i.e., $\hat{u}_j = u_j$, where β is an fixed but unknown bias parameter $0 < \beta \leq 1$

1.3 Distributional assumptions

It is assumed that latent utility u_i of each candidate i is drawn from some distribution $\mathcal{D}$ independent of all other candidates. I.I.D. utilities encode the fact that there are no systematic differences in utilities across groups.

It also assumes that preference list σ_i of each candidate i is drawn independently from some distribution $\mathcal{L}$ of the set of all preference lists of p institutions. I.I.D. preference lists encode the assumption that candidates in either group have the same distribution of preferences over institutions.

1.4 Evaluation metrices

To evaluate the algorithm used for matching, following evaluation matrices are considered-

- **Utility ratio:** For a fixed utility vector v and a preference vector σ, given a selection $M_{v,\sigma} : [n] \to [p]$, we define the utility of $M_{v,\sigma}$ as the sum of utilities of all candidates that are assigned to any institution.

$$\mathcal{U}(M_{v,\sigma}) := \sum_{i \in M^{-1}([p])} v_i.$$

Clearly, the maximum value $\mathcal{U}(\cdot)$ can take is the sum of top K entries in the true utility vector $\mathbf{u}$; we denote this by $\mathcal{U}^\star(\mathbf{u})$.

Consider an algorithm $\mathcal{A}$ that, given preferences σ and utilities $\hat{\mathbf{u}}$, outputs a selection $M_{\hat{\mathbf{u}},\sigma}$. The utility-ratio of $\mathcal{A}$ is defined as:

$$\mathscr{U}_{\mathcal{D},\mathcal{L}}(\mathcal{A}) := \mathbb{E}_{\mathbf{u} \sim \mathcal{D}, \sigma \sim \mathcal{L}} \frac{\mathcal{U}(M_{\hat{\mathbf{u}},\sigma})}{\mathcal{U}^\star(\mathbf{u})}.$$

Note that the algorithm's assignment relies on the estimated utilities $\hat{\mathbf{u}}$ of the candidates. However, when measuring the utility ratio, we use the true utilities of the selected candidates.

- **Fairness metrics:** The two fairness matrices considered are-

  i. **Representational fairness:** It considers how many candidates from each group are assigned to at least one institution. For group $j \in \{1, 2\}$, let ρ_j denote the fraction of candidates in G_j that are selected by $M_{\hat{\mathbf{u}},\sigma}$. For an algorithm $\mathcal{A}$, its representational fairness is defined as:

$$\mathcal{R}_{\mathcal{D},\mathcal{L}}(\mathcal{A}) := \mathbb{E}_{\mathbf{u} \sim \mathcal{D}, \sigma \sim \mathcal{L}} \frac{\min_{j \in \{1,2\}} \rho_j}{\max_{j' \in \{1,2\}} \rho_{j'}}.$$

  By definition, $\mathcal{R}_{\mathcal{D},\mathcal{L}}(\mathcal{A})$ is a value between 0 and 1. $\mathcal{R}_{\mathcal{D},\mathcal{L}}(\mathcal{A})$ is close to 1 if the $\mathcal{A}$ ensures that a proportional number of candidates from each group are assigned to at least one institution. The larger the value of $\mathcal{R}_{\mathcal{D},\mathcal{L}}(\mathcal{A})$, the more "fair" $\mathcal{A}$ is.

  ii. **Preference-based fairness:** It captures the disparity in the fraction of candidates in each group that are assigned to their "top" preferences. Let π_j denote the fraction of candidates in group G_j that get their first preference. For an algorithm $\mathcal{A}$, its preference-based fairness is defined as:

$$\mathscr{P}_{\mathcal{D},\mathcal{L}}(\mathcal{A}) := \mathbb{E}_{\mathbf{u} \sim \mathcal{D}, \sigma \sim \mathcal{L}} \frac{\min_{j \in \{1,2\}} \pi_j}{\max_{j' \in \{1,2\}} \pi_{j'}}.$$

  $\mathscr{P}_{\mathcal{D},\mathcal{L}}(\mathcal{A})$ is a value between 0 and 1 and large value is preferable.

1.5 Algorithms

- **Unconstrained algorithm:** It is a special case of the Gale-Shapley algorithm for stable matching where all the institutions agree on a common central ranking of the candidates. The algorithm considers candidates in decreasing order of observed utility for doing the assignment and whenever a candidate is considered, it assigns the candidate to its highest preferred institution that still has an available slot. The detailed description is as follows-

Algorithm 1 Unconstrained Algorithm (A_{st})

Input: S (set of candidates), $k = (k_1, k_2, \ldots, k_p)$ (institutional capacities), $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ (preferences for $i \in S$), $\hat{\mathbf{u}} = (\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_n)$ (estimated utilities for $i \in S$)

$r_{\text{unmatched}} \leftarrow$ Candidates in S sorted in order of decreasing utility
$M \leftarrow$ assignment filled with -1(no institution)
for each candidate a in $r_{\text{unmatched}}$ **do**
    $M_a \leftarrow$top choice institution n such that $k_n > 0$
    $k_n \leftarrow k_n - 1$
    until $k_n = 0$ for all $n \in k$
end for
return M

- **Institution-wise algorithm:** It creates two independent instances, one for each group. In the instance corresponding to group G_j, the capacity at an institution i is set to $\frac{|G_j|}{n} \cdot k_i$. Then A_{st} is independently executed on each of these two instances. This ensures that every institute has a proportional number of students from both the groups. The detailed description is as follows-

Algorithm 2 Institution-wise Algorithm ($A_{\text{inst-wise}}$)

Input: G_1 (set of candidates in group 1); G_2 (set of candidates in group 2);
$k = (k_1, k_2, \ldots, k_p)$ (institutional capacities);
$\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ (preferences for each $i \in S$);
$\hat{\mathbf{u}} = (\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_n)$ (estimated utilities for each $i \in S$)

$A \leftarrow \{k_i \cdot |G_1|/n : i \in [p]\}$ // Define capacity vector A for the p institutions for candidates in G_1
$B \leftarrow \{k_i \cdot |G_2|/n : i \in [p]\}$ // Define capacity vector B for the p institutions for candidates in G_2
$M_1 \leftarrow A_{\text{st}}(G_1, A, \{\sigma_i : i \in A\}, \{\hat{u}_i : i \in G_1\})$
$M_2 \leftarrow A_{\text{st}}(G_2, B, \{\sigma_i : i \in B\}, \{\hat{u}_i : i \in G_2\})$
return $M_1 \cup M_2$

1.6 Theoretical results

1. **Inefficiency of unconstrained algorithm in presence of bias:**

   Consider an instance where the utilities of the candidates are drawn from the uniform distribution on $[0,1]$ and the preference lists are drawn from an arbitrary distribution. Assume there are constants η_1, η_2 such that $n_j \geq \eta_j n$ for each $j \in \{1,2\}$, $K \geq \eta_2 n$. For any algorithm $\mathcal{A}$ that, given $\hat{\mathbf{u}}$ and σ, outputs an assignment maximizing the estimated utility,

   $$\begin{aligned}\mathscr{R}(\mathcal{A}) &= \max\left\{\frac{K - n_1(1-\beta)}{K\beta + n_2(1-\beta)}, 0\right\} \pm O\left(\frac{\sqrt{\log n}}{\eta_1\eta_2\sqrt{n}}\right)\\ \mathcal{U}(\mathcal{A}) &= \frac{\sum_{j=1}^{2} f(\alpha_j, n_j)}{f(K,n)} \pm O\left(\frac{\sqrt{\log n}}{\eta_2\sqrt{n}}\right)\end{aligned}$$

   Here $f(x,y) := x - \frac{x(x+1)}{2(y+1)}$, $\alpha_2 := K - \alpha_1$, $\alpha_1 := K - \frac{n_2}{\beta n_1 + n_2} \cdot \max\{K - (1-\beta)n_1, 0\}$. Assuming $\mathcal{A}$ is the stable assignment algorithm $\mathcal{A}_{st}$, $\mathscr{P}(\mathcal{A})$ is at most

   $$\max\left\{\frac{K - n_1(1-\beta)}{K\beta + n_2(1-\beta)}, 0\right\} + O\left(\frac{\sqrt{\log n}}{\eta_1\eta_2\sqrt{n}}\right).$$

   Since $\mathcal{A}_{st}$ is also utility maximizing and hence, first two parts also applies to it.

   Special case: If $n_1 = n_2 = K$. Then,

   $$\mathscr{P}(\mathcal{A}_{st}) \leq \beta + O\left(\frac{p\sqrt{\log n}}{\sqrt{n}}\right)$$

   $$\mathscr{R}(\mathcal{A}_{st}) = \beta \pm O\left(\frac{\sqrt{\log n}}{\sqrt{n}}\right)$$

   $$\mathcal{U}(\mathcal{A}_{st}) = \frac{2}{3} + \frac{4\beta}{3(\beta+1)^2} \pm O\left(\frac{\sqrt{\log n}}{\sqrt{n}}\right)$$

   The above result shows that when the bias parameter β is close to zero (the case of high bias), then the fairness metrics $\mathscr{P}(\mathcal{A}_{st})$, $\mathscr{R}(\mathcal{A}_{st})$ and $\mathcal{U}(\mathcal{A}_{st})$ deteriorate substantially, and there is a significant drop in the utility of the selected candidates as well. Also note that the error terms in the theorem decay as $p/\sqrt{n}$ and are negligible for reasonable values of n and p.

2. **Efficiency of institute-wise algorithm in presence of bias:**

   Let $\eta_1, \eta_2, \eta_3 > 0$ be parameters such that $|G_j| \geq \eta_j n$ for each $j \in \{1,2\}$, $K \geq \eta_2 n$, and $k_\ell \geq \eta_3 K$ for each $\ell \in [p]$. The allocation algorithm $\mathcal{A}_{\text{inst-wise}}$ will be more efficient compared to $\mathcal{A}_{st}$, for any distribution of utilities and preference lists, and bias parameter β. The theoretical values of evaluation matrices for $\mathcal{A}_{\text{inst-wise}}$ are as follows-

$$\mathcal{U}(\mathcal{A}_{\text{inst-wise}}) \geq 1 - O\left(\frac{\sqrt{\log n}}{\sqrt{\eta_2 n}}\right)$$

$$\mathscr{P}(\mathcal{A}_{\text{inst-wise}}) \geq 1 - O\left(\frac{p\sqrt{\log K}}{\eta_1 \eta_3 \sqrt{K}}\right)$$

$$\mathscr{R}(\mathcal{A}_{\text{inst-wise}}) = 1$$

Clearly $\mathscr{R}(\mathcal{A}_{\text{inst-wise}})$ is exactly equal to 1 by construction of the algorithm and the other two matrices are also closed to 1 for appropriate n and p. Compared to the results for $\mathcal{A}_{\text{st}}$ it is clearly more efficient.

2 Statistical Discrimination in Stable Matchings[2]

2.1 Introduction

This section considers a matching problem in which the colleges observe a noisy estimate of quality of each candidate and noise depend on the group to which candidate belong. Here if noise is larger for a group then the correlation between the estimates of different colleges is lower compared to other group.

In this setting when we apply a matching algorithm known as Deferred Acceptance algorithm we get some interesting results that are as follows:

- Lower correlation for one group worsens the estimates for all the groups.
- Probability that a candidate is assigned to their first choice is independent of their group.
- Probability that a candidate is assigned to a college at all depends on their group, revealing the presence of discrimination coming from the correlation effect alone. Group having more noise is better off in it.

2.2 Model Specification

The study aims isolate the effect of difference in correlation between the estimates of quality of student done by different colleges across the groups, from the effects of variance and bias. This is achieved by equalizing the marginal distributions of estimations for each group at each college, leaving only a difference in correlation. The detailed specification of the model are as follows:

- The model consider two colleges **A** and **B** and a continuum unit mass of students **S**
- **Capacity** of college A and B is given as $\boldsymbol{\alpha_A}$ and $\boldsymbol{\alpha_B}$, both in $(0, 1]$, respectively. It is also a continuum mass.
- The students are divided into **two groups**, denoted $\boldsymbol{G_1}$ and $\boldsymbol{G_2}$ having proportion $\boldsymbol{\gamma}$ and $\boldsymbol{1 - \gamma}$ respectively, where $\gamma \in [0, 1]$
- **Individual preference-** Among group G_1, a proportion $\boldsymbol{\beta_{G_1}} \in [0, 1]$ prefers college A to college B, the remaining $\boldsymbol{1-\beta_{G_1}}$ prefers B, similarly in G_2 proportion $\boldsymbol{\beta_{G_2}} \in [0, 1]$ prefers A and $\boldsymbol{1 - \beta_{G_2}}$ prefer B. All students prefer attending any college rather than remaining unmatched.
- **Colleges preferences** over students are based on grades and each college assigns grades according to a standard normal distribution. Colleges have variable accuracies in evaluating students, depending on whether they belong to group G_1 or group G_2. The correlation between the grades assigned to a student by

two colleges is $\boldsymbol{\rho_{G_1}}$ if he belongs to G_1 and is $\boldsymbol{\rho_{G_1}}$ if he belongs to G_2. This is called differential correlation.

Formally, for each student $s \in S$, we assume that his grades at college A and B form a vector (W_s^A, W_s^B) and are drawn randomly from the bi-variate normal distribution with mean $(0,0)$, variance $(1,1)$ and the covariance dependent on group.

$$f_{W^A,W^B}(W_s^A, W_s^B) = \frac{1}{2\pi\sqrt{1-\rho_G^2}} \exp\left(-\frac{1}{2(1-\rho_G^2)}\left[(W_s^A)^2 - 2\rho_G W_s^A W_s^B + (W_s^B)^2\right]\right)$$

$$f_{W^A}(W_s^A) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(W_s^A)^2}{2}\right)$$

$$f_{W^B}(W_s^B) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(W_s^B)^2}{2}\right)$$

2.3 Some Definitions

- **Taste-based discrimination** refers to actual preferences being based on demographic groups, i.e. the discrimination that results from the individual bias.

- **Statistical discrimination** refers to the discrimination that occur from the imperfect information that a decision maker has about individuals' qualities, i.e. the discrimination that results from the limitation of estimation method.

- **Blocking** The pair (s, C) blocks a matching μ if s would prefer C to her current match, and either C has remaining capacity or it has admitted a student with a lower score than s. Formally,

$$\mu(s) \prec_s C \quad \text{and} \quad \left(\eta(\mu(C)) < \alpha_C \quad \text{or} \quad \exists s' \in \mu(C) \ \text{ such } \ \text{ that } \ W_{s'}^C < W_s^C\right).$$

- **Stability** A matching is stable if it is not blocked by any student-college pair.

- **Cutoffs** The cutoff of a college represents the grade above which a student who applies gets admitted. If μ is a stable matching and A and B are the colleges then the cutoffs for college A and B are $P_A$and P_B respectively, which are given as

$$P_A := \inf\{W_s^A : \mu(s) = A\} \quad \text{and} \quad P_B := \inf\{W_s^B : \mu(s) = B\}.$$

- **Individual student demand** Given the cutoffs P_A and P_B, we say that student s demand $D_s(P_A, P_B)$ which is the college they prefer among those where they pass the cutoff, or themselves if they do not pass the cutoff at any college.

- **Aggregate demand** The aggregate demand at college $C \in \{A, B\}$ is the mass of students demanding it:

$$D_C(P_A, P_B) = \eta(\{s : D_s(P_A, P_B) = C\}), \text{ where } C \in \{A, B\}$$

- **Aggregate supply** For a given college C it is simply the capacity of the college C.

- **Market clearing** A cutoff vector is market clearing if it induces a demand that is equal to supply at each college.

  Formally when $P_A, P_B \neq -\infty$, the system of equations

$$D_A(P_A, P_B) = \alpha_A$$
$$D_B(P_A, P_B) = \alpha_B$$

  is called the market clearing equations, and the cutoffs P_A and P_B that satisfy these equations are called market clearing cutoffs.

2.4 Matching algorithm

Since this is a continuum mass setting so the classic Deferred Acceptance algorithm is used which is a extension of Gale and Shapely algorithm to the continuum model which is given as-

---

Algorithm 3 Deferred Acceptance Algorithm

1: **First step:** All students apply to their favorite college; they are temporarily accepted. If the mass of students applying to college C is greater than its capacity α_C, then C only keeps the α_C best.
2: **while** A positive mass of students are unmatched and have not yet been rejected from every college **do**
3: Each student who has been rejected at the previous step proposes to her preferred college among those which have not rejected him yet.
4: Each college C keeps the best α_C mass of students among those it had temporarily accepted and those who just applied, and rejects the others.
5: **end while**
6: **End:** If the mass of students that are either matched or rejected from every college is 1, the algorithm stops. However, it could happen that it takes an infinite number of steps to converge.

It was theoretically proven that the above algorithm always gives a stable matching once it stops.

2.5 Welfare metrics

For correlation levels ρ_{G_1} and ρ_{G_2} associated to groups G_1 and G_2, we define $V_1^{G_1,\mathbf{A}}(\rho_{G_1},\rho_{G_2})$, $V_1^{G_1,\mathbf{B}}(\rho_{G_1},\rho_{G_2})$, $V_1^{G_2,\mathbf{A}}(\rho_{G_1},\rho_{G_2})$ and $V_1^{G_2,\mathbf{B}}(\rho_{G_1},\rho_{G_2})$ as the proportion of students from each group-preference profile who get their first choice. Equivalently, it is the probability of a student to get their first choice conditionally on their profile. Formally,

$$
\begin{aligned}
V_1^{G_1,\mathbf{A}}(\rho_{G_1},\rho_{G_2}) &:= \frac{1}{\gamma\beta_{G_1}}\eta(\{s\in G_1 : \mathbf{A}\succ_s \mathbf{B}, \mu(s)=\mathbf{A}\})\\
V_1^{G_1,\mathbf{B}}(\rho_{G_1},\rho_{G_2}) &:= \frac{1}{\gamma(1-\beta_{G_1})}\eta(\{s\in G_1 : \mathbf{B}\succ_s \mathbf{A}, \mu(s)=\mathbf{B}\})\\
V_1^{G_2,\mathbf{A}}(\rho_{G_1},\rho_{G_2}) &:= \frac{1}{(1-\gamma)\beta_{G_2}}\eta(\{s\in G_2 : \mathbf{A}\succ_s \mathbf{B}, \mu(s)=\mathbf{A}\})\\
V_1^{G_2,\mathbf{B}}(\rho_{G_1},\rho_{G_2}) &:= \frac{1}{(1-\gamma)(1-\beta_{G_2})}\eta(\{s\in G_2 : \mathbf{B}\succ_s \mathbf{A}, \mu(s)=\mathbf{B}\})
\end{aligned}
$$

We also define the following aggregated quantities:

$$
\begin{aligned}
V_1^{G_1}(\rho_{G_1},\rho_{G_2}) &:= \beta_{G_1}V_1^{G_1,\mathbf{A}}(\rho_{G_1},\rho_{G_2}) + (1-\beta_{G_1})V_1^{G_1,\mathbf{B}},\\
V_1^{G_2}(\rho_{G_1},\rho_{G_2}) &:= \beta_{G_2}V_1^{G_2,\mathbf{A}}(\rho_{G_1},\rho_{G_2}) + (1-\beta_{G_2})V_1^{G_1,\mathbf{B}},
\end{aligned}
$$

which represent the probability of a student getting their first choice conditional on belonging to group G_1 or G_2 respectively. We define similarly the proportions of students getting their second choice or staying unmatched, using respectively V_2 and $V_\emptyset$ instead of V_1

2.6 Main Results

1. **First choice probability: group independence**

   The probability that a student gets its first choice is independent from the group they belong to. Formally, for any $\gamma, \beta_{G1}, \beta_{G2}, \alpha_A, \alpha_B, \rho_{G1}, \rho_{G2} \in [0,1]$,

   $$V_1^{G1,A}(\rho_{G1},\rho_{G2}) = V_1^{G2,A}(\rho_{G1},\rho_{G2})$$

   and

   $$V_1^{G1,B}(\rho_{G1},\rho_{G2}) = V_1^{G2,B}(\rho_{G1},\rho_{G2})$$

   If $\beta_{G1} = \beta_{G2}$, then

   $$V_1^{G1}(\rho_{G1},\rho_{G2}) = V_1^{G2}(\rho_{G1},\rho_{G2})$$

2. **Second choice & unmatched probability: group dependence**

   The proportion of students getting their second choice and remaining unmatched is not the same across both groups: students from the group with higher correlation coefficient have a lower probability of getting their second choice and a higher probability of staying unmatched. Formally, if $\rho_{G1} < \rho_{G2}$, then

$$V_2^{G1,A} > V_2^{G2,A}, \quad V_2^{G1,B} > V_2^{G2,B}, \quad V_\emptyset^{G1,A} < V_\emptyset^{G2,A}, \quad \text{and} \quad V_\emptyset^{G1,B} < V_\emptyset^{G2,B}$$

3. **Impact of correlation on first choice probability**

   The probability that a student of either group gets their first choice is increasing in both ρ_{G1} and ρ_{G2}. Formally, let $\gamma, \beta_{G1}, \beta_{G2} \in [0,1]$, $\alpha_A, \alpha_B \in (0,1)$ such that $\alpha_A + \alpha_B < 1$, and $\rho_{G1}, \rho_{G2} \in [0,1)$. Then for $C \in \{A, B\}$ and $G \in \{G1, G2\}$:

$$\frac{\partial V_1^{G,C}}{\partial \rho_{G1}} > 0 \quad \text{and} \quad \frac{\partial V_1^{G,C}}{\partial \rho_{G2}} > 0$$

4. **Impact of correlation on unmatched probability**

   The proportion of students from a given group remaining unmatched is increasing in its own correlation level and decreasing in the correlation level of the other group. Formally, for $C \in \{A, B\}$ and $G \in \{G1, G2\}$:

$$\frac{\partial V_\emptyset^{G,C}}{\partial \rho_G} > 0 \quad \text{and} \quad \frac{\partial V_\emptyset^{G,C}}{\partial \rho_{\overline{G}}} < 0$$

   where $\overline{G}$ represents the other group.

2.7 Empirical Validation[3]

This is the new sub-section that we add to this study which has not been analyzed before. For verifying the results found empirically we have used synthetic data for scores generated from the bi-variate normal distribution with mean $(70, 70)$ and standard deviation $(15, 15)$. Clearly, we have taken same mean and standard deviation, we just scale up these terms from $(0, 0)$ and $(1, 1)$ to avoid negative scores. Although without scaling also we get same results. Since the paper talks about a continuum setting so we have take a large n close to 10000, as for large population discrete model behave same as the continuum model. After generating the data we run the matching algorithm and find all the 16 different welfare metrics. And then we have plotted the appropriate plots to validate the results. The plots for different results are as follows-

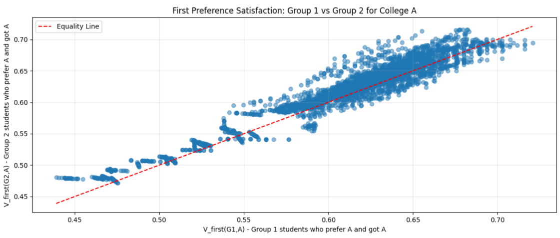
1. **First choice probability: group independence** This results suggest the following

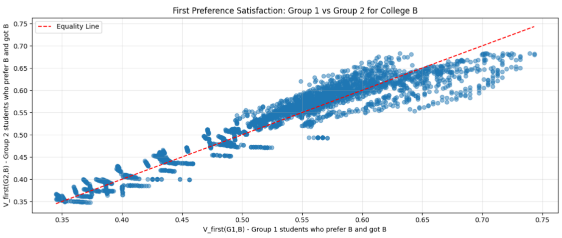
$$V_1^{G1,A}(\rho_{G1}, \rho_{G2}) = V_1^{G2,A}(\rho_{G1}, \rho_{G2})$$

and

$$V_1^{G1,B}(\rho_{G1}, \rho_{G2}) = V_1^{G2,B}(\rho_{G1}, \rho_{G2})$$

We have calculated these matrices for different parameter values and taken over 3000 instances and the plots clearly indicate that on average they lie on the line of equality which is the suggested theoretically in the result.

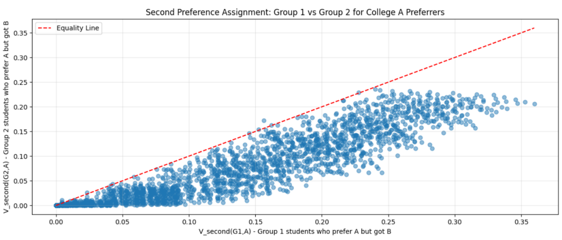


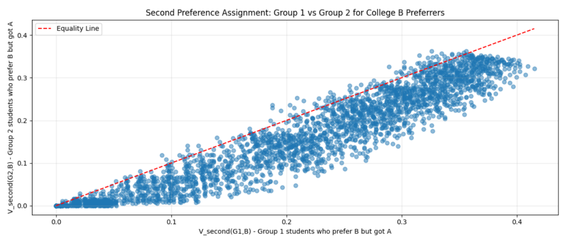


2. **Second choice & unmatched probability: group dependence** When the $\rho_{G1} < \rho_{G2}$, this result suggest that the probability of getting second choice is such that-

$$V_2^{G1,A} > V_2^{G2,A} \quad and \quad V_2^{G1,B} > V_2^{G2,B}$$

which is evident in the empirical result given below

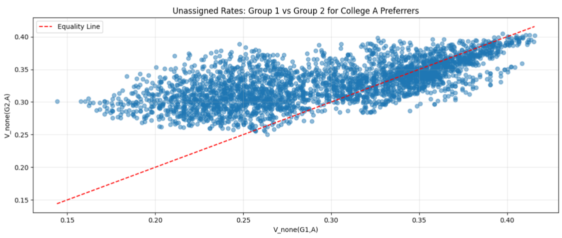


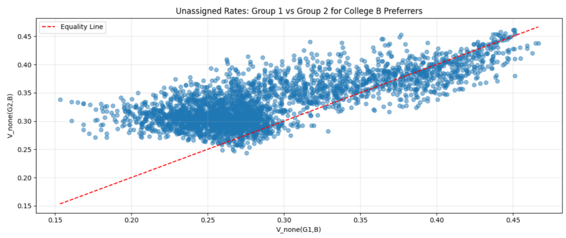


For probability of getting no choice the result suggest the following-

$$V_\emptyset^{G1,A} < V_\emptyset^{G2,A} \quad \text{and} \quad V_\emptyset^{G1,B} < V_\emptyset^{G2,B}$$

this is also evident from the empirical result-

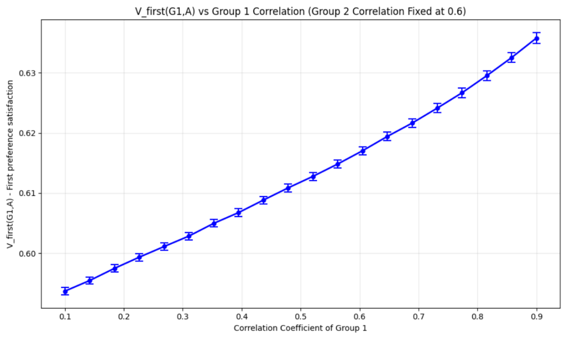


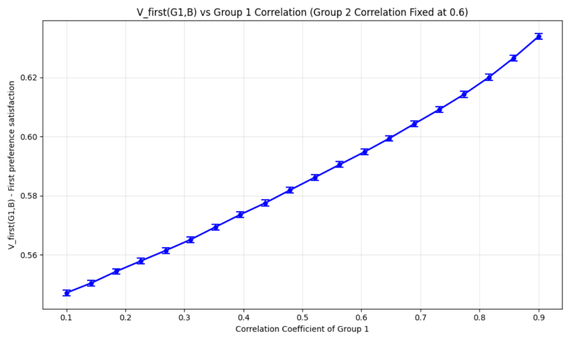


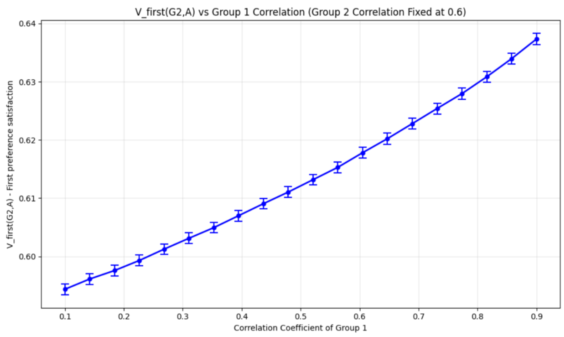
3. **Impact of correlation on first choice probability** This result suggest the following theoretically-

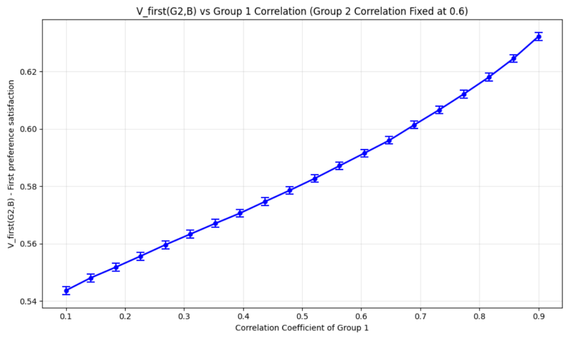
$$\frac{\partial V_1^{G,C}}{\partial \rho_{G1}} > 0 \quad \text{and} \quad \frac{\partial V_1^{G,C}}{\partial \rho_{G2}} > 0$$

And clearly this also holds on an average for the empirical observation-





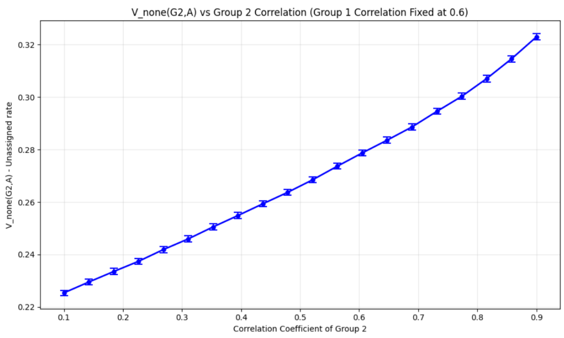


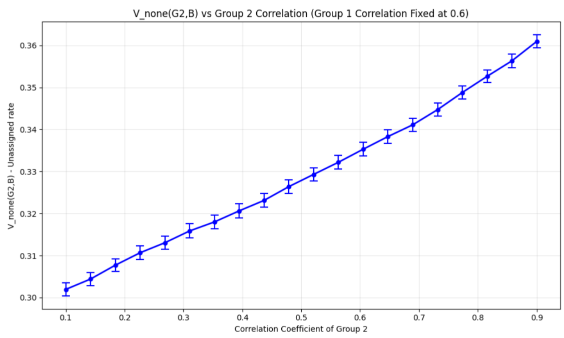


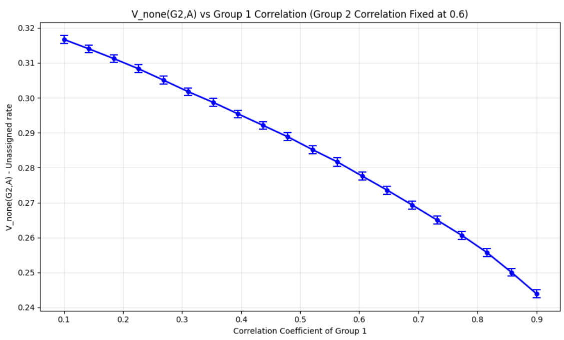
Similar results also hold when we keep the correlation of group 1 constant and vary group 2 correlation coefficient.

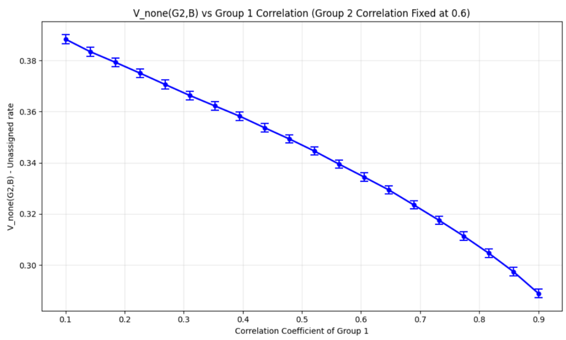
4. **Impact of correlation on unmatched probability** This suggest the following theoretical result-
$$\frac{\partial V_{\emptyset}^{G,C}}{\partial \rho_G} > 0 \quad \text{and} \quad \frac{\partial V_{\emptyset}^{G,C}}{\partial \rho_{\overline{G}}} < 0$$
where $\overline{G}$ represents the other group.

This result is consistent with observations obtained empirically.









Similar results are also observed for matrices evaluated for group 1.

3 New problems for Research

This is a completely new section added by us in the course of UGP. In this section we have discussed some of the new matching problems that are not studied before and have a very high theoretical and industrial relevance. The detailed description is given in the following subsections.

3.1 Decentralized Matching with Group-Dependent Bias and Noise

3.1.1 Introduction

It is clear that the first section deals with the scenario having centralized institution for estimating the quality of student and there is a group dependent bias. And the second section deals with a scenario having decentralized system of estimating quality of each candidate but do not consider any group dependent bias. The most natural extension of these scenarios is a setting having both group dependent bias and a decentralized system of evaluating quality i.e. group dependent distribution of estimation noise. This kind of model is really useful for modeling real world situations, one of such situation is given below:

Consider there are two companies A and B and the set of candidates N is divided into two groups G_1 and G_2. And consider G_1 is a advantaged group compared to G_2. Then in most real life scenarios candidates from group G_1 have much access to education compared to candidates from G_2 so they are better prepared for selection exams taken by companies and this gives a group dependent bias for the group G_2. At the same time G_1's candidates have better certificates to showcase their quality which lower down the estimation noise for group G_1. Clearly this situation can be modeled best by our Decentralized matching model with group-dependent bias and noise.

3.1.2 Model Specification

Consider there are n candidates in set of candidates N which are divided into two groups G_1 and G_2 such that

- G_1 is the advantaged group and its candidates face no bias in utility estimation and also has low estimation noise.
- G_2 is the disadvantaged group and its candidates experience both *biased estimation of utility* and *higher estimation noise.*

Now consider U_i, true utility of each candidate i irrespective of its group, is drawn from a normal distribution i.e.

$$U_i \sim \mathcal{N}(\mu, \sigma^2).$$

The fact that distribution of true utility is same for both groups stress on the fact that there is no systematic bias in the true utilities of candidates. But now there is a group dependent bias and noise distribution that add up during estimation of utility. If $\hat{U}_i^C$ denotes the estimated utility of candidate i by institution C, where $C \in \{A, B\}$, and $G(i)$ is the group of candidate i then

$$\hat{U}_i^C = \beta_{G(i)} U_i + \epsilon_{G(i)}^C$$

Here

- $\beta_{G(i)} = 1$ if $G(i) = G_1$ and $\beta_{G(i)} = \beta < 1$ if $G(i) = G_2$.
- $\epsilon_{G(i)}^c \sim \mathcal{N}(0, \sigma_{G(i)}^2)$ is estimation noise, where $\sigma_{G_2}^2 > \sigma_{G_1}^2$.

3.1.3 Joint Distribution of Observed Utilities

Clearly $\hat{U}_i^A$ and $\hat{U}_i^B$ denotes the estimated utility of candidate i by company $A\&B$ respectively. Since both follow normal marginal distribution so their joint distribution is also normal and given as-

$$\begin{bmatrix} \hat{U}_i^A \\ \hat{U}_i^B \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \beta_{G(i)}\mu \\ \beta_{G(i)}\mu \end{bmatrix}, \begin{bmatrix} \beta_{G(i)}^2 \sigma^2 + \sigma_{G(i)}^2 & \beta_{G(i)}^2 \sigma^2 \\ \beta_{G(i)}^2 \sigma^2 & \beta_{G(i)}^2 \sigma^2 + \sigma_{G(i)}^2 \end{bmatrix} \right).$$

and then correlation coefficient will be

$$\rho(\hat{U}_i^A, \hat{U}_i^B) = \frac{\beta_{G(i)}^2 \sigma^2}{\beta_{G(i)}^2 \sigma^2 + \sigma_{G(i)}^2}$$

Clearly as $\sigma_{G_2}^2 > \sigma_{G_1}^2$ so keeping all other same we have $\rho_{G_1} > \rho_{G_2}$ i.e. advantaged group has higher correlation in its estimates compared to the disadvantaged group. **Remark:** If we put $\beta = 1$, $\mu = 0$, $\sigma^2 = \frac{\rho}{\beta^2}$ and $\rho_{G_1} = \rho_{G_2} = 1 - \rho$ we reach back to the setting of second section.

3.1.4 Problem Statement

Given the set of candidates N and companies C, the objective is to design a fair matching mechanism $M : N \to C$ that:

1. **Maximizes true Utility:** It is a fundamental desired quality from any matching algorithm as it maximizes the social surplus.
2. **Corrects for scaling bias:** Ensures candidates from disadvantaged groups are not unfairly underrepresented.
3. **Stabilizes noisy estimates:** Prevents high noise from affecting disadvantaged candidates disproportionately.
4. **Ensures Representation Fairness:** Introduces constraints to balance selection across groups.

Mathematically, we maximize total utility:

$$U(M) = \sum_{i \in N} U_i \mathbb{1}(i \text{ is matched}).$$

subject to fairness constraints:

$$\mathbb{E}[P(G_2, c)] \geq \lambda \mathbb{E}[P(G_1, c)], \quad \forall c \in C.$$

where $P(G_j, c)$ is the fraction of group G_j candidates matched to institution c and $\lambda \leq 1$ is the fairness parameter.

3.1.5 Why This Is New and Useful

- **Extends Section 1:** Unlike past model that only scale down disadvantaged groups' utility, we introduce noise in estimation along with it.
- **Extends Section 2:** Instead of assuming all candidates have the same mean utility, we incorporate β, which creates a dual-bias problem.
- **Accounts for Decentralized Estimation:** Unlike centralized models where all institutions agree on rankings, our model allows different institutions to have different, noisy estimates.
- **Model Real world Situations:** Similar to given example there are many real world situations in which this double bias model can work more efficiently then the previous models.

3.2 Dynamic Bias Mitigation in Gig Economy Platforms

3.2.1 Introduction

Gig economy platforms (e.g., Uber, TaskRabbit) match workers to tasks based on two key factors:

1. **Historical ratings** (utilities), which are often biased against certain demographic groups (e.g., gender, race).
2. **Worker preferences** (e.g., location, task type).

Existing matching algorithms (e.g., Gale-Shapley) assume utilities(derived from ratings) and fail to account for:

- **Dynamic biases**: Bias parameters are not static but depend on temporal factors (e.g., demand surges, changing evaluator pools).
- **Feedback loops**: Biased ratings perpetuate inequitable opportunities, as marginalized workers receive fewer tasks and thus fewer chances to improve their ratings.

3.2.2 Research Gap

Prior work treats each matching instance as an independent static problem and this do not account for the compounding of bias due to feedback loop and there is no study that models how bias and thus the utilities evolve across multiple matching rounds which is highly relevant for gig platforms.

3.2.3 Proposed direction for solution

A two-phase online algorithm can solve this problem:

1. **Bias Estimation**:
   - Use **online learning models** (e.g., Online Gradient Descent, Thompson Sampling) to estimate time-varying bias parameters from streaming rating data.
   - Example: For worker group G_j at time t, model observed ratings as $\widehat{u}_i^t = \beta_j^t \cdot u_i^t + \epsilon$, where β_j^t is the dynamic bias parameter.
2. **Matching with Dynamic Fairness Constraints**:
   - Extend the Gale-Shapley algorithm to incorporate real-time bias estimates, ensuring:
     (a) *Proportional opportunities* for workers from all groups (ensure fairness).
     (b) *Utility maximization* for the platform (matching high-quality workers to important tasks).

3.2.4 Novelty & Impact

- **Dynamic Bias Adaptation**: Unlike static approaches, our new model will adjust to dynamic biases (e.g., seasonal trends in ratings).
- **Industry Relevance**: Directly applicable to gig platforms seeking to improve fairness without sacrificing efficiency.

3.3 Job Market Matching with Evolving Capacities

3.3.1 Introduction & Problem Statement

Again consider a job market where candidates belong to two demographic groups:

- G_1 (Advantaged): Typically have higher qualifications and more documented proofs of skills.
- G_2 (Disadvantaged): Have lower average qualifications and fewer proofs, leading to systematically different hiring outcomes.

Each candidate i has a **utility** U_i, representing their expected productivity in a job. However, due to structural inequalities, utility follows **group-dependent distribution**:

$$U_i|G_{G(i)} \sim \mathcal{N}(\mu_{G(i)}, \sigma^2_{G(i)})$$

where:

- $G(i)$ is the group to which i belongs.
- $\mu_{G_1} > \mu_{G_2}$ (Advantaged group has a higher mean utility due to better access to education and opportunities).
- $\sigma_{G_1} < \sigma_{G_2}$ (Disadvantaged group has a higher variance due to more uncertainty in skill assessment).

There is a set of companies C where each company $c \in C$ has a fixed capacity q_c. Job seekers and companies both have preferences over whom they match with. Traditionally, the Gale-Shapley is used for job-market matching, which prioritizes utility maximization.

3.3.2 Issues with Traditional Matching

1. **Feedback Loop of Inequality:** Since advantaged group has higher average utilities, they are consistently hired by top companies. And due to this disadvantaged group is left with fewer opportunities, further reducing their future utility which increase with experience. This creates a reinforcing cycle, where the utility gap between groups remains or even worsens.

2. **Evolving Utility Over Time:** It is a well known fact that with training and work experience, underrepresented groups can also improve their utility. Let U_i^t denote candidate i's utility at time t then by evolving utilities we mean:

   $$U_i^{t+1} = U_i^t + \gamma \cdot \mathbb{1}(i \text{ was hired}) - \delta \cdot \mathbb{1}(i \text{ was unemployed}).$$

   where $\gamma > 0$ & $\delta > 0$.

3. **Dynamic Market Capacity:** If candidates from advantaged group are not allocated to top jobs, they may turn to startups, increasing market capacity. Thus we can model total market size at time t as:

   $$Q^t = Q^{t-1} + \beta \sum_{i \in G_1} \mathbb{1}(i \text{ remains unemployed}),$$

   where β represents the rate at which unemployed males contribute to new job creation.

3.3.3 Proposed Solution: Fair and Dynamic Job Matching

To ensure long-term societal benefits, we propose that a **fair allocation algorithm** should be created such that:

1. **Incorporates Representation Fairness Constraints:** Matching algorithm should be modified to ensure that each company has at least a fraction λ of candidates from the disadvantaged group:

$$P(G_2, c) \geq \lambda P(G_1, c), \quad \forall c \in C.$$

2. **Implements a Temporal Utility Update Model:** Candidate utilities evolve based on past job experience-

$$U_i^{t+1} = U_i^t + \gamma \cdot \mathbb{1}(i \text{ was hired}) - \delta \cdot \mathbb{1}(i \text{ was unemployed}).$$

   If a candidate remains unemployed, its utility decreases and if candidate gets employment, its utility increases.

3. **Incorporates dynamic market capacity:** As mentioned above the market size would change according to the equation-

$$Q^t = Q^{t-1} + \beta \sum_{i \in G_1} \mathbb{1}(i \text{ remains unemployed}),$$

   So matching algorithm should also take into account which set of candidate will help to increase market size and thus the final social surplus.

4. **Optimizes for Long-Term Societal Utility:** Instead of static utility maximization, the new algorith should maximize total welfare of T time i.e.

$$W^T = \sum_{t=1}^{T} \sum_i U_i^t,$$

   ensuring that utility growth is considered.

3.3.4 Expected Outcomes

- More balanced initial job allocations → Proportional number of candidates from disadvantaged group gain access to high-paying jobs.
- Increase in long-term societal utility → Groups that were previously underrepresented improve their productivity.
- Higher job market capacity → More startups are created as top candidates explore entrepreneurship and this potentially increase the total social surplus accumulated over a period of time.

References

[1] L. et al. (2024). *Centralized Selection with Preferences in the Presence of Biases*. arXiv preprint.
https://arxiv.org/abs/2409.04897

[2] Castera, R. et al. (2022). *Statistical Discrimination in Stable Matchings*. Stony Brook University.
https://gtcenter.org/wp-content/uploads/2022/04/Stony_Brooks_Submission_R_Castera.pdf

[3] Dhruv Rai (2025). *AlgoBias GitHub Repository*.
https://github.com/Dhruv-Rai365/AlgoBias