Statistical discrimination in selection

Patrick Loiseau
Inria Saclay, FairPlay team

London School of Economics, March 2023

Joint work with

- PhD students
  - Vitalii Emelianov
  - Rémi Castera
- Collaborators
  - Nicolas Gast, Inria
  - Bary Pradelski, CNRS
  - Krishna Gummadi, MPI-SWS

Discriminatory outcomes in selection problems

Hiring Discrimination Against Black Americans Hasn't Declined in 25 Years

by Lincoln Quillian , Devah Pager , Arnfinn H. Midtbøen and Ole Hexel

October 11, 2017

"Brise vocation", "Koh Lanta de l'orientation" : faut-il supprimer Parcoursup ?

par Sonia Princet publié le 18 janvier 2022 à 13h24

THE MANY SINS OF COLLEGE ADMISSIONS

By Jeannie Suk Gersen

October 7, 2019

China: Gender Discrimination in Hiring Persists

11 Percent of Civil Service Ads Specify 'Men Only'

New Evidence of Age Bias in Hiring, and a Push to Fight It

Discriminatory outcomes in selection problems

Hiring Discrimination Against Black Americans Hasn't Declined in 25 Years

by Lincoln Quillian , Devah Pager , Arnfinn H. Midtbøen and Ole Hexel

October 11, 2017

THE MANY SINS OF COLLEGE ADMISSIONS

By Jeannie Suk Gersen

October 7, 2019

"Brise vocation", "Koh Lanta de l'orientation" : faut-il supprimer Parcoursup ?

par Sonia Princet publié le 18 janvier 2022 à 13h24

China: Gender Discrimination in Hiring Persists

11 Percent of Civil Service Ads Specify 'Men Only'

New Evidence of Age Bias in Hiring, and a Push to Fight It

All listed above problems are **selection** problems

What is a Selection Problem?

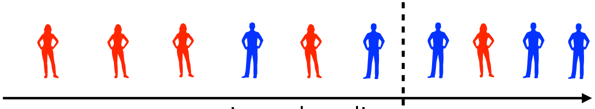
- Candidates are sorted based on their quality estimates
- Best candidates are selected

50% of candidates are women

25% of selected candidates are women

What is a Selection Problem?

- Candidates are sorted based on their quality estimates
- Best candidates are selected
- Decision-makers can use different fairness mechanisms:
  - Rooney Rule (v)
  - 80%-Rule (x)
  - Demographic Parity (x)



50% of candidates are women
25% of selected candidates are women

What is a Selection Problem?

- Candidates are sorted based on their quality estimates
- Best candidates are selected
- Decision-makers can use different fairness mechanisms:
  - Rooney Rule (v)
  - 80%-Rule (x)
  - Demographic Parity (x)

50% of candidates are women

25% of selected candidates are women

Main questions:
What causes discriminatory outcome? What is the effect of fairness mechanisms?

Algorithmic Fairness Literature

- Machine learning systems can lead to discrimination Larson et al., 2016; Speicher et al., 2018...
- Different notions of fairness are proposed

  X – feature representation, $G \in \{A, B\}$ – demographic group
  Y – binary quality (0=bad, 1=good), $\hat{Y}$ – binary prediction (0=reject,1=accept)

Algorithmic Fairness Literature

- Machine learning systems can lead to discrimination Larson et al., 2016; Speicher et al., 2018...
- Different notions of fairness are proposed

  X – feature representation, $G \in \{A, B\}$ – demographic group
  Y – binary quality (0=bad, 1=good), $\hat{Y}$ – binary prediction (0=reject,1=accept)

Individual Fairness

Dwork et al., 2012:

$|\mathbb{P}(\hat{Y} = 1|X) - \mathbb{P}(\hat{Y} = 1|X')| \leq \lambda d(X, X')$

Kearns et al., 2017:

$\mathbb{P}(Y = 1|X) \geq \mathbb{P}(Y = 1|X') \implies$

$\mathbb{P}(\hat{Y} = 1|X) \geq \mathbb{P}(\hat{Y} = 1|X')$

Algorithmic Fairness Literature

- Machine learning systems can lead to discrimination — Larson et al., 2016; Speicher et al., 2018...
- Different notions of fairness are proposed

  X – feature representation, $G \in \{A, B\}$ – demographic group
  Y – binary quality (0=bad, 1=good), $\hat{Y}$ – binary prediction (0=reject,1=accept)

Individual Fairness

Dwork et al., 2012:

$|\mathbb{P}(\hat{Y}=1|X) - \mathbb{P}(\hat{Y}=1|X')| \leq \lambda d(X, X')$

Kearns et al., 2017:

$\mathbb{P}(Y=1|X) \geq \mathbb{P}(Y=1|X') \implies$
$\mathbb{P}(\hat{Y}=1|X) \geq \mathbb{P}(\hat{Y}=1|X')$

Group Fairness

Demographic Parity

$\mathbb{P}(\hat{Y}=1|A) = \mathbb{P}(\hat{Y}=1|B)$

80%-Rule

$\mathbb{P}(\hat{Y}=1|A)/\mathbb{P}(\hat{Y}=1|B) \geq 0.8$

Equal Opportunity

$\mathbb{P}(\hat{Y}=1|Y=1, A) = \mathbb{P}(\hat{Y}=1|Y=1, B)$

Algorithmic Fairness Literature

- Machine learning systems can lead to discrimination — Larson et al., 2016; Speicher et al., 2018...
- Different notions of fairness are proposed

  X – feature representation, $G \in \{A, B\}$ – demographic group

  Y – binary quality (0=bad, 1=good), $\hat{Y}$ – binary prediction (0=reject,1=accept)

Individual Fairness

Dwork et al., 2012:

$$|\mathbb{P}(\hat{Y}=1|X) - \mathbb{P}(\hat{Y}=1|X')| \leq \lambda d(X, X')$$

Kearns et al., 2017:

$$\mathbb{P}(Y=1|X) \geq \mathbb{P}(Y=1|X') \implies$$
$$\mathbb{P}(\hat{Y}=1|X) \geq \mathbb{P}(\hat{Y}=1|X')$$

Group Fairness

Demographic Parity

$$\mathbb{P}(\hat{Y}=1|A) = \mathbb{P}(\hat{Y}=1|B)$$

80%-Rule

$$\mathbb{P}(\hat{Y}=1|A)/\mathbb{P}(\hat{Y}=1|B) \geq 0.8$$

Equal Opportunity

$$\mathbb{P}(\hat{Y}=1|Y=1, A) = \mathbb{P}(\hat{Y}=1|Y=1, B)$$

Other notions

Counterfactual Fairness (Kusner et al., 2017)

Envy-Freeness (Balcan et al., 2019)

Algorithmic Fairness Literature

- Algorithmic solutions to ensure fairness

  Preprocessing learning fair representations — Zemel et al., 2013; Gordaliza et al., 2019...

  Inprocessing fairness as a constraint in the learning procedure — Zafar et al., 2017...

  Postprocessing resampling predictions to ensure fairness — Hardt et al., 2016; Petersen et al., 2021

- Most literature studies fairness in **classification problems**
- The **causes of discrimination** are rarely taken into account

Algorithmic Fairness Literature

- Algorithmic solutions to ensure fairness

  Preprocessing learning fair representations — Zemel et al., 2013; Gordaliza et al., 2019...

  Inprocessing fairness as a constraint in the learning procedure — Zafar et al., 2017...

  Postprocessing resampling predictions to ensure fairness — Hardt et al., 2016; Petersen et al., 2021

- Most literature studies fairness in **classification problems**
- The **causes of discrimination** are rarely taken into account

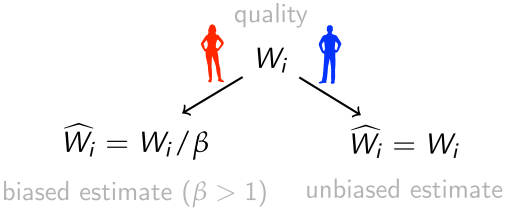
- **A few works on fairness in selection problems** — Kleinberg et al., 2018; Hu et al., 2019...

- Discrimination usually explained by **bias**
- This talk: **second-order statistics create discrimination**

Contents

Causes of Discrimination

Implicit Bias



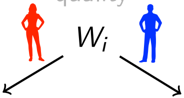
(Kleinberg et al., 2018; Celis et al., 2021):

- Implicit bias naturally leads to discrimination (overrepresentation of a group)
- Fairness mechanisms (Rooney rule) can improve selection utility

Causes of Discrimination

Implicit Bias

quality

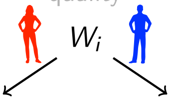


$\widehat{W_i} = W_i / \beta$ $\qquad$ $\widehat{W_i} = W_i$

biased estimate ($\beta > 1$) $\qquad$ unbiased estimate

Differential Variance (our work)

quality



$\widehat{W_i} = W_i + \varepsilon_i \cdot \sigma_H$ $\qquad$ $\widehat{W_i} = W_i + \varepsilon_i \cdot \sigma_L$

noisy estimates, $\varepsilon_i \sim \mathcal{N}(0, 1)$

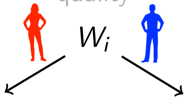
(Kleinberg et al., 2018; Celis et al., 2021):

- Implicit bias naturally leads to discrimination (overrepresentation of a group)
- Fairness mechanisms (Rooney rule) can improve selection utility

Causes of Discrimination

Implicit Bias

quality



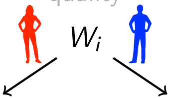
W_i

$\widehat{W_i} = W_i / \beta$ $\quad$ $\widehat{W_i} = W_i$

biased estimate ($\beta > 1$) $\quad$ unbiased estimate

Differential Variance (our work)

quality



W_i

$\widehat{W_i} = W_i + \varepsilon_i \cdot \sigma_H$ $\quad$ $\widehat{W_i} = W_i + \varepsilon_i \cdot \sigma_L$

noisy estimates, $\varepsilon_i \sim \mathcal{N}(0, 1)$

(Kleinberg et al., 2018; Celis et al., 2021):

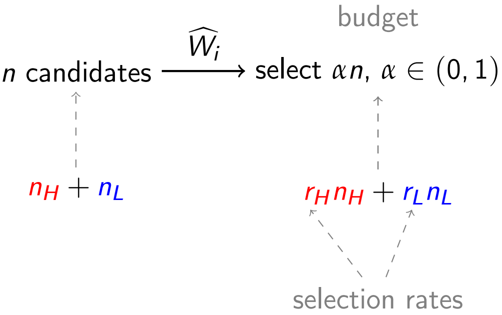
- Implicit bias naturally leads to discrimination (overrepresentation of a group)
- Fairness mechanisms (Rooney rule) can improve selection utility

Main questions:

- What is the impact of differential variance on the selection outcome?
- What is the effect of fairness mechanisms on the selection utility?

Note: (Phelps, 1972; Lundberg et al., 1983) model differential variance in wages allocation

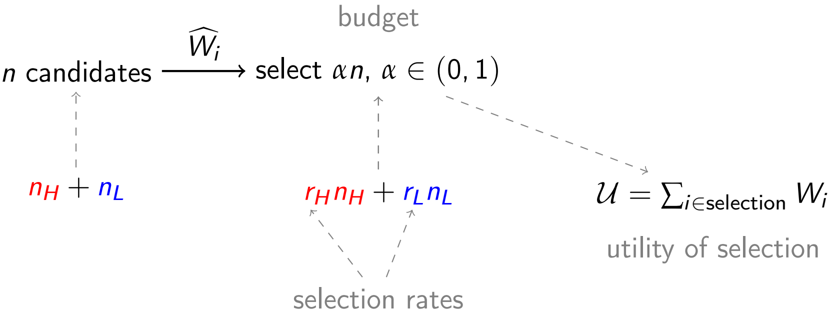
Selection Problem Setup



Assume (for simplicity) that the latent quality $W \sim \mathcal{N}(\mu, \sigma^2)$ (group-independent)

Technically: we assume that n is large and denote p_H, p_L the fractions of candidates for each group.

Selection Problem Setup

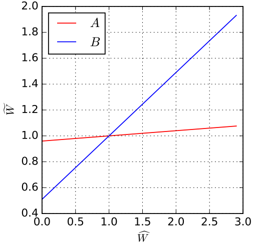


Assume (for simplicity) that the latent quality $W \sim \mathcal{N}(\mu, \sigma^2)$ (group-independent)

Technically: we assume that n is large and denote p_H, p_L the fractions of candidates for each group.

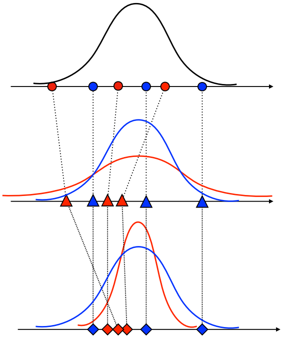
Baseline decision makers

- Group-oblivious: Sort candidates by decreasing estimate $\widehat{W}_i$ and keep the best
  - Does not look at group membership
- Bayesian: Computes posterior $\widetilde{W}_i = \mathbb{E}(W_i | \widehat{W}_i)$ and keep the best



$$\widetilde{W}_G = \widehat{W}_G \rho_G^2 + (1 - \rho_G^2)\mu,$$

$$\text{where } \rho_G^2 = \frac{\sigma^2}{\sigma^2 + \sigma_G^2}$$

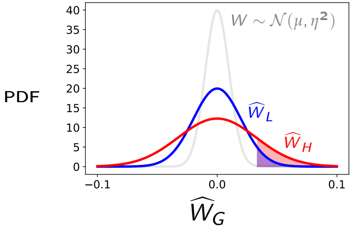


  - Looks at group membership
  - Inverts the variance orders (now group H has lower variance)

Discrimination Caused by Differential Variance

Group-Oblivious DM

sort by $\widehat{W}_i$ and keep best αn

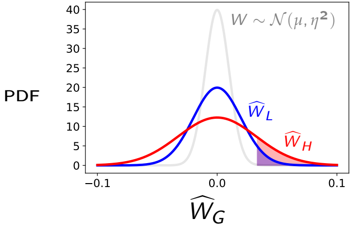


Theorem ([Emelianov, Gast, Loiseau, Gummadi, EC '20])

Discrimination Caused by Differential Variance

Group-Oblivious DM

sort by $\widehat{W}_i$ and keep best αn



Theorem ([Emelianov, Gast, Loiseau, Gummadi, EC '20])

- *If the decision-maker is* ***group-oblivious****, then* $r_H^{\text{obl}} > r_L^{\text{obl}} \iff \alpha < 0.5$

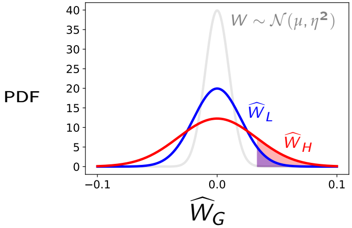
Discrimination Caused by Differential Variance

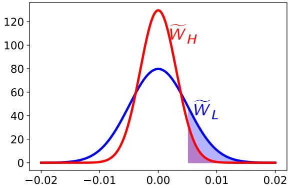
Group-Oblivious DM

sort by $\widehat{W}_i$ and keep best αn

Bayesian DM

sort by $\widetilde{W}_i = \mathbb{E}(W_i | \widehat{W}_i)$ and keep best αn





$$\widetilde{W}_G = \widehat{W}_G \rho_G^2 + (1 - \rho_G^2)\mu$$

Theorem ([Emelianov, Gast, Loiseau, Gummadi, EC '20])

- *If the decision-maker is* ***group-oblivious****, then* $r_H^{\text{obl}} > r_L^{\text{obl}} \iff \alpha < 0.5$
- *If the decision-maker is* ***Bayesian****, then* $r_H^{\text{bayes}} < r_L^{\text{bayes}} \iff \alpha < 0.5$

Discrimination Caused by Differential Variance

Group-Oblivious DM

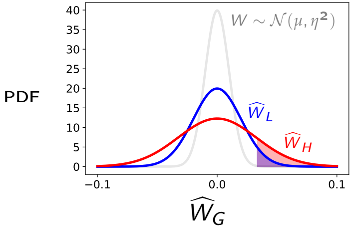
sort by $\widehat{W}_i$ and keep best αn

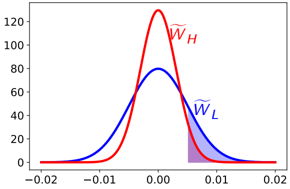
Bayesian DM

sort by $\widetilde{W}_i = \mathbb{E}(W_i | \widehat{W}_i)$ and keep best αn

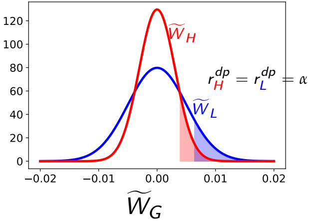
Demographic Parity DM

sort by $\widetilde{W}_i$ (or $\widehat{W}_i$) and keep best αn_G from each group G





$$\widetilde{W}_G = \widehat{W}_G \rho_G^2 + (1 - \rho_G^2)\mu$$



Theorem ([Emelianov, Gast, Loiseau, Gummadi, EC '20])

- *If the decision-maker is* ***group-oblivious****, then* $r_H^{\text{obl}} > r_L^{\text{obl}} \iff \alpha < 0.5$
- *If the decision-maker is* ***Bayesian****, then* $r_H^{\text{bayes}} < r_L^{\text{bayes}} \iff \alpha < 0.5$

Demographic Parity Can Improve Selection Utility

Theorem ([Emelianov, Gast, Loiseau, Gummadi, AIJ '22])

- *If the decision-maker is* ***Bayesian****, then for all selection rates* $\alpha \in (0,1)$*:*

$$1 \leq \frac{\mathcal{U}^{\text{bayes}}}{\mathcal{U}^{\text{dp}}} \leq 1 + \frac{p_H(\nu - 1)}{p_H + p_L \nu} \text{ where } \nu = \frac{\sqrt{\sigma_H + \eta^2}}{\sqrt{\sigma_L + \eta^2}}$$

Demographic Parity Can Improve Selection Utility

Theorem ([Emelianov, Gast, Loiseau, Gummadi, AIJ '22])

- *If the decision-maker is* ***Bayesian****, then for all selection rates* $\alpha \in (0,1)$*:*

$$1 \le \frac{\mathcal{U}^{\text{bayes}}}{\mathcal{U}^{\text{dp}}} \le 1 + \frac{p_H(\nu - 1)}{p_H + p_L \nu} \text{ where } \nu = \frac{\sqrt{\sigma_H + \eta^2}}{\sqrt{\sigma_L + \eta^2}}$$

- *If the decision-maker is* ***Group-Oblivious****, then for all selection rates* $\alpha \in (0,1)$*:*

$$\mathcal{U}^{\text{dp}} \ge \mathcal{U}^{\text{obl}}$$

Demographic Parity Can Improve Selection Utility

Theorem ([Emelianov, Gast, Loiseau, Gummadi, AIJ '22])

- *If the decision-maker is* ***Bayesian****, then for all selection rates* $\alpha \in (0,1)$*:*

$$1 \leq \frac{\mathcal{U}^{\text{bayes}}}{\mathcal{U}^{\text{dp}}} \leq 1 + \frac{p_H(\nu - 1)}{p_H + p_L \nu} \quad \text{where } \nu = \frac{\sqrt{\sigma_H + \eta^2}}{\sqrt{\sigma_L + \eta^2}}$$

- *If the decision-maker is* ***Group-Oblivious****, then for all selection rates* $\alpha \in (0,1)$*:*

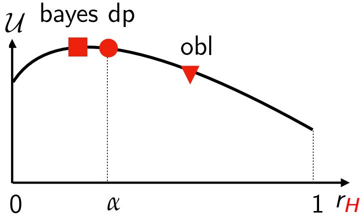
$$\mathcal{U}^{\text{dp}} \geq \mathcal{U}^{\text{obl}}$$

Proof Idea:

- Utility $\mathcal{U}$ is a concave function of selection rate r_H
- From the previous slide, we know that

$$r_H^{bayes} < r_H^{dp} = \alpha < r_H^{obl}$$

Using concavity of $\mathcal{U}$, can extend the result for the γ-rule



Summary and Discussion

- Second-order statistical differences between groups (differential variance) leads to discrimination
- Demographic parity (and γ-rule) fairness mechanism can increase the selection quality

Summary and Discussion

- Second-order statistical differences between groups (differential variance) leads to discrimination
- Demographic parity (and γ-rule) fairness mechanism can increase the selection quality

Extensions

- Generalize to group-dependent quality distribution and/or presence of implicit bias
  $\Longrightarrow$ more nuanced results, typically for small selection budget (α)
- Candidates can be strategic, i.e., they can adapt to the selection rule
  $\Longrightarrow$ results contrast with the non-strategic case
  $\Longrightarrow$ demographic parity can sometimes improve quality even over Bayesian

Contents

Rémi Castera, Patrick Loiseau, and Bary S.R. Pradelski.
Statistical Discrimination in Stable Matchings.
EC '22

Example of a matching problem: college admission

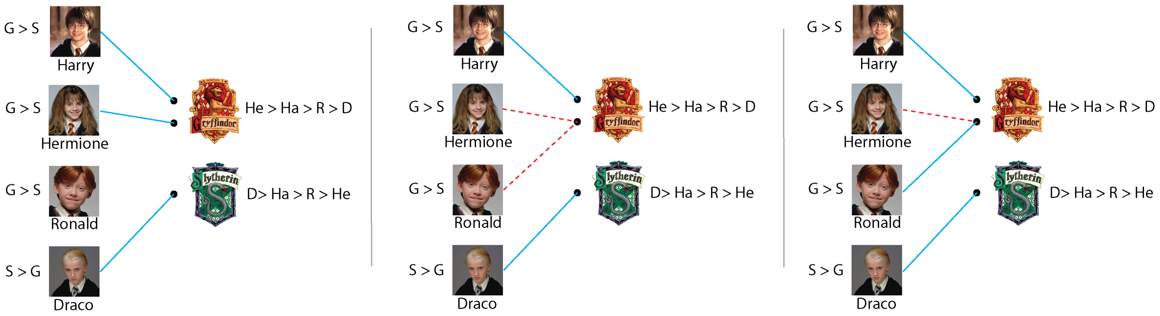


Figure: Example of a college admission problem

Left: **stable**
Middle: **waste** - Hermione and Ronald could go to Gryffindor
Right: **justified envy** - Hermione should replace Ronald in Gryffindor

Second-order correlation: motivating example

Colleges A and B have noisy estimates of applicants' qualities. Each applicant s has a **latent quality** $W^s \sim \mathcal{N}(0, \sigma^2)$; and her grade at each college is:

$$\widehat{W}_A^s = W^s + \varepsilon_A^s, \ \widehat{W}_B^s = W^s + \varepsilon_B^s$$

Second-order correlation: motivating example

Colleges A and B have noisy estimates of applicants' qualities. Each applicant s has a **latent quality** $W^s \sim \mathcal{N}(0, \sigma^2)$; and her grade at each college is:

$$\widehat{W}_A^s = W^s + \varepsilon_A^s, \ \widehat{W}_B^s = W^s + \varepsilon_B^s$$

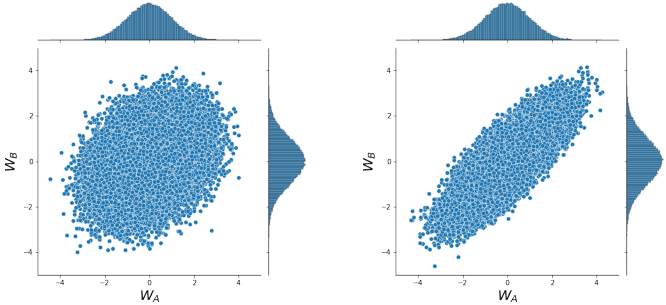
Two groups of applicants: **local and foreign**. Evaluation of local applicants is more precise than for foreign applicants. For a local applicant s, $\varepsilon^s \sim \mathcal{N}(0, \sigma_{loc}^2)$ and for a foreign applicant $\varepsilon^s \sim \mathcal{N}(0, \sigma_{for}^2)$, with $\sigma_{loc} < \sigma_{for}$.

Second-order correlation: motivating example (continued)

For fairness purposes, colleges decide to standardize the grade distributions: grades of local students are fitted into $\mathcal{N}(0,1)$, and so are grades of foreign students:

$$\text{for any local student } s,\ \widetilde{W}_A^s = \widehat{W}_A^s / \sqrt{\sigma^2 + \sigma_{loc}^2},\ \widetilde{W}_B^s = \widehat{W}_B^s / \sqrt{\sigma^2 + \sigma_{loc}^2}$$

$$\text{for any foreign student } s,\ \widetilde{W}_A^s = \widehat{W}_A^s / \sqrt{\sigma^2 + \sigma_{for}^2},\ \widetilde{W}_B^s = \widehat{W}_B^s / \sqrt{\sigma^2 + \sigma_{for}^2}$$



Motivating example 2: different criteria

Two colleges, A and B, with **different criteria**. Suppose college A is interested in the level of applicants in maths, and college B in physics. Applicants come from two high schools:

Motivating example 2: different criteria

Two colleges, A and B, with **different criteria**. Suppose college A is interested in the level of applicants in maths, and college B in physics. Applicants come from two high schools:

High school 1: Physics is taught in a very theoretical way
$\rightarrow$ grades in maths and physics are **highly correlated**.

Motivating example 2: different criteria

Two colleges, A and B, with **different criteria**. Suppose college A is interested in the level of applicants in maths, and college B in physics. Applicants come from two high schools:

High school 1: Physics is taught in a very theoretical way
$\rightarrow$ grades in maths and physics are **highly correlated**.

High school 2: Physics is taught through experiments
$\rightarrow$ grades in maths and physics are more **independent**.

Motivating example 2: different criteria

Two colleges, A and B, with **different criteria**. Suppose college A is interested in the level of applicants in maths, and college B in physics. Applicants come from two high schools:

High school 1: Physics is taught in a very theoretical way
$\rightarrow$ grades in maths and physics are **highly correlated**.

High school 2: Physics is taught through experiments
$\rightarrow$ grades in maths and physics are more **independent**.

Figure: Example of distributions. Left: correlation 0.8, right: correlation 0.3

Main questions

- How does the matching outcome depend on the correlation structure?
- If correlation depends on group, which group is "better-off"?

The model

- We consider a **continuum of students** S. Masses of students are measured with $\eta : S \to [0, 1]$, $\eta(S) = 1$.

The model

- We consider a **continuum of students** S. Masses of students are measured with $\eta : S \to [0,1]$, $\eta(S) = 1$.
- S is divided into **2 groups** G_1 and G_2. Respective masses: $\gamma \in [0,1]$ and $1 - \gamma$.

The model

- We consider a **continuum of students** S. Masses of students are measured with $\eta : S \to [0, 1]$, $\eta(S) = 1$.
- S is divided into **2 groups** G_1 and G_2. Respective masses: $\gamma \in [0, 1]$ and $1 - \gamma$.
- We consider **two colleges**, A and B, with respective capacities: α_A and $\alpha_B \in [0, 1]$, where $\alpha_A + \alpha_B < 1$ (capacity shortage).

The model

- We consider a **continuum of students** S. Masses of students are measured with $\eta : S \to [0, 1]$, $\eta(S) = 1$.

- S is divided into **2 groups** G_1 and G_2. Respective masses: $\gamma \in [0, 1]$ and $1 - \gamma$.

- We consider **two colleges**, A and B, with respective capacities: α_A and $\alpha_B \in [0, 1]$, where $\alpha_A + \alpha_B < 1$ (capacity shortage).

- For G_1: proportion β_1 prefer A, $1 - \beta_1$ prefer B. Same for G_2 with β_2.

The model

- We consider a **continuum of students** S. Masses of students are measured with $\eta : S \to [0, 1]$, $\eta(S) = 1$.
- S is divided into **2 groups** G_1 and G_2. Respective masses: $\gamma \in [0, 1]$ and $1 - \gamma$.
- We consider **two colleges**, A and B, with respective capacities: α_A and $\alpha_B \in [0, 1]$, where $\alpha_A + \alpha_B < 1$ (capacity shortage).
- For G_1: proportion β_1 prefer A, $1 - \beta_1$ prefer B. Same for G_2 with β_2.
- Each college produces a ranking by giving a **grade** to each student.

Differential correlation

For each student s, their grades are $(W_A^s, W_B^s) \sim \mathcal{N}((0,0), C_s)$ with $C_s = \begin{pmatrix} 1 & \rho_{G(s)} \\ \rho_{G(s)} & 1 \end{pmatrix}$.

Differential correlation

For each student s, their grades are $(W_A^s, W_B^s) \sim \mathcal{N}((0,0), C_s)$ with

$$C_s = \begin{pmatrix} 1 & \rho_{G(s)} \\ \rho_{G(s)} & 1 \end{pmatrix}.$$

Groups have different correlation levels, but the same marginals (e.g., normalization).

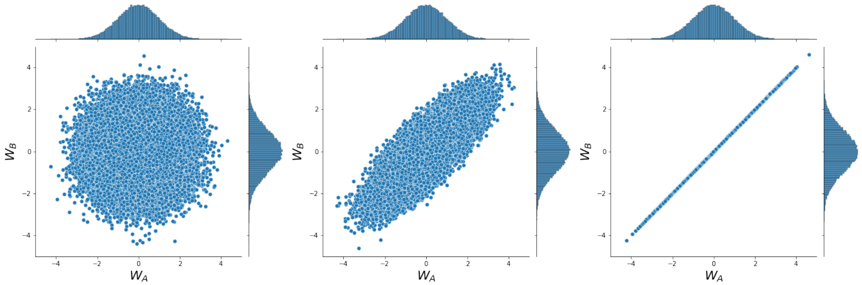


Figure: Grades distributions for different correlation levels, left to right: $\rho_s = 0, 0.8, 1$.

Stable Matching

Definition (Stable matching)

For each student s, for each college c such that s prefers c to the college they are matched with, all students matched to c were ranked better than s at c.

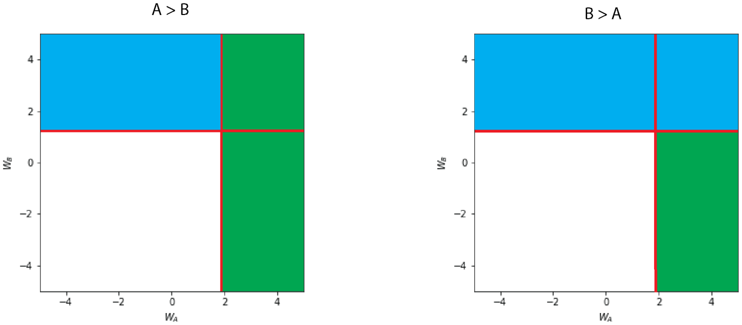


Figure: Green: matched to A, blue: matched to B, white: unmatched

Market clearing equations (solution of stable matching)

Let $P_A, P_B \in \mathbb{R}$ be **cutoffs**, i.e., the grade of the 'worst' admitted student in resp. A and B.

Market clearing equations (solution of stable matching)

Let $P_A, P_B \in \mathbb{R}$ be **cutoffs**, i.e., the grade of the 'worst' admitted student in resp. A and B. Define the **demand** at each college:

$$
\begin{aligned}
D_A(P_A, P_B) = \; & \eta(\{s \in S | W_A^s \geq P_A \text{ and } s \text{ prefers } A \text{ to } B\}) \\
& + \eta(\{s \in S | W_A^s \geq P_A \text{ , } W_B^s < P_B \text{ and } s \text{ prefers } B \text{ to } A\})
\end{aligned}
$$

$$
\begin{aligned}
D_B(P_A, P_B) = \; & \eta(\{s \in S | W_B^s > P_B \text{ and } s \text{ prefers } B \text{ to } A\}) \\
& + \eta(\{s \in S | W_B^s \geq P_B \text{ , } W_A^s < P_A \text{ and } s \text{ prefers } A \text{ to } B\})
\end{aligned}
$$

Market clearing equations (solution of stable matching)

Let $P_A, P_B \in \mathbb{R}$ be **cutoffs**, i.e., the grade of the 'worst' admitted student in resp. A and B. Define the **demand** at each college:

$$\begin{aligned} D_A(P_A, P_B) = \;& \eta(\{s \in S | W_A^s \geq P_A \text{ and } s \text{ prefers } A \text{ to } B\}) \\ & + \eta(\{s \in S | W_A^s \geq P_A \;,\; W_B^s < P_B \text{ and } s \text{ prefers } B \text{ to } A\}) \\ D_B(P_A, P_B) = \;& \eta(\{s \in S | W_B^s > P_B \text{ and } s \text{ prefers } B \text{ to } A\}) \\ & + \eta(\{s \in S | W_B^s \geq P_B \;,\; W_A^s < P_A \text{ and } s \text{ prefers } A \text{ to } B\}) \end{aligned}$$

We say that the cutoffs P_A and P_B are **market clearing** if

$$D_A(P_A, P_B) = \alpha_A \text{ and } D_B(P_A, P_B) = \alpha_B$$

Market clearing equations (solution of stable matching)

Let $P_A, P_B \in \mathbb{R}$ be **cutoffs**, i.e., the grade of the 'worst' admitted student in resp. A and B.
Define the **demand** at each college:

$$\begin{aligned} D_A(P_A, P_B) = \;& \eta(\{s \in S | W_A^s \geq P_A \text{ and } s \text{ prefers } A \text{ to } B\}) \\ & + \eta(\{s \in S | W_A^s \geq P_A \;,\; W_B^s < P_B \text{ and } s \text{ prefers } B \text{ to } A\}) \end{aligned}$$

$$\begin{aligned} D_B(P_A, P_B) = \;& \eta(\{s \in S | W_B^s > P_B \text{ and } s \text{ prefers } B \text{ to } A\}) \\ & + \eta(\{s \in S | W_B^s \geq P_B \;,\; W_A^s < P_A \text{ and } s \text{ prefers } A \text{ to } B\}) \end{aligned}$$

We say that the cutoffs P_A and P_B are **market clearing** if

$$D_A(P_A, P_B) = \alpha_A \text{ and } D_B(P_A, P_B) = \alpha_B$$

Theorem (Azevedo et al., 2016)

There is a unique stable matching, and it is given by the unique pair of market clearing cutoffs.

Matching outcomes

Each student will get either their first choice, second choice, or stay unmatched. The masses of students in each of these cases can be expressed using the cutoffs:

$$\begin{aligned} V_1^{G_1,A} &= \mathbb{P}(\text{s is matched to A} \mid \text{s prefers A and } s \in G_1) \\ &= \mathbb{P}_{\rho_1}(W_A^s \geq P_A) \end{aligned}$$

Define analogously $V_1^{G_1,B}$, $V_1^{G_2,A}$ and $V_1^{G_2,B}$.

Matching outcomes

Each student will get either their first choice, second choice, or stay unmatched. The masses of students in each of these cases can be expressed using the cutoffs:

$$\begin{aligned} V_1^{G_1,A} &= \mathbb{P}(\text{s is matched to A} \mid \text{s prefers A and } s \in G_1) \\ &= \mathbb{P}_{\rho_1}(W_A^s \geq P_A) \end{aligned}$$

Define analogously $V_1^{G_1,B}$, $V_1^{G_2,A}$ and $V_1^{G_2,B}$.

Similarly, the probability of getting their second choice or to stay unmatched is given by:

$$\begin{aligned} V_2^{G_1,A} &= \mathbb{P}(\text{s is matched to B} \mid \text{s prefers A and } s \in G_1) \\ &= \mathbb{P}_{\rho_1}(W_A^s < P_A, W_B^s \geq P_B) \\ V_\emptyset^{G_1,A} &= \mathbb{P}(\text{s is unmatched} \mid \text{s prefers A and } s \in G_1) \\ &= \mathbb{P}_{\rho_1}(W_A^s < P_A, W_B^s < P_B) \end{aligned}$$

Solving the market clearing equation

To compute the matching, it suffices to find the market clearing cutoffs P_A and P_B, i.e., to solve the equations:

$$\begin{cases} \gamma_1 \beta_1^A V_1^{G_1,A} + \gamma_1 \beta_1^B V_2^{G_1,B} + \gamma_2 \beta_2^A V_1^{G_2,A} + \gamma_2 \beta_2^B V_2^{G_2,B} & = \alpha_A, \\ \gamma_1 \beta_1^A V_2^{G_1,A} + \gamma_1 \beta_1^B V_1^{G_1,B} + \gamma_2 \beta_2^A V_2^{G_2,A} + \gamma_2 \beta_2^B V_1^{G_2,B} & = \alpha_B. \end{cases}$$

This equation generally has no analytic solution, thus the matching cannot be computed directly. **However, it can be used to derive qualitative results.**

Statistical discrimination

First question: **is one group advantaged compared to the other?**

Statistical discrimination

First question: **is one group advantaged compared to the other?**

Theorem ([Castera, Loiseau, Pradelski, EC '22])

i) ***The probability for a student to get her first choice is independent of the group she belongs to.*** *($V_1^{G_1,A} = V_1^{G_2,A}$, $V_1^{G_1,B} = V_1^{G_2,B}$)*

ii) ***Students from the high correlation group have a lower probability to get their second choice, and therefore a higher probability of staying unmatched.*** *(If $\rho_1 < \rho_2$, then $V_2^{G_1,A} > V_2^{G_2,A}$ and $V_\emptyset^{G_1,A} < V_\emptyset^{G_2,A}$; same for B.)*

Therefore, **belonging to the high correlation group leads to a worse outcome**.

Proof idea: Probabilities of the type $\mathbb{P}_\rho(W_A^s < P_A, W_B^s \geq P_B)$ decrease with ρ (property of the Gaussian distribution).

Illustration

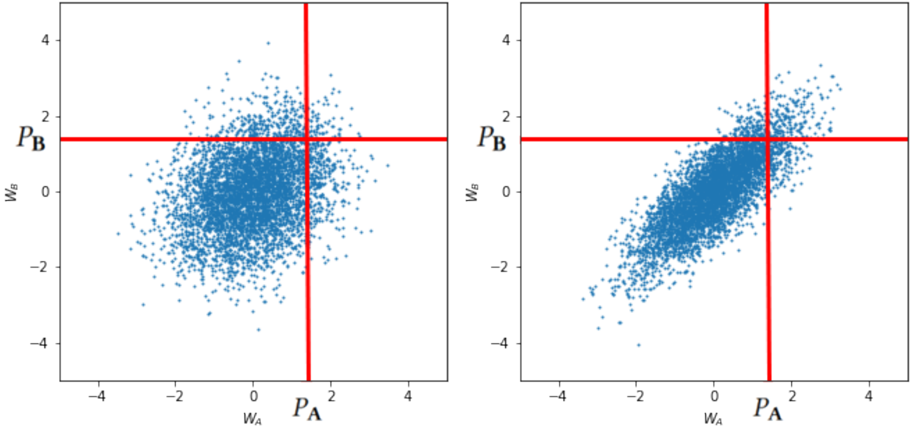


Figure: Illustration of statistical discrimination

Comparative statics

Second question: **what happens when correlation levels change?**

Comparative statics

Second question: **what happens when correlation levels change?**

Theorem ([Castera, Loiseau, Pradelski, EC '22])

i) ***The probability of a student getting their first choice is increasing in both groups' correlation levels.*** *(For $G \in \{G_1, G_2\}$, for $C \in \{A, B\}$, $\frac{\partial V_1^{G,C}}{\partial \rho_G} > 0$.)*

ii) ***The probability of a student remaining unmatched is decreasing in the other group's correlation level and increasing in her own.*** *(For $G \in \{G_1, G_2\}$, for $C \in \{A, B\}$, $\frac{\partial V_\emptyset^{G,C}}{\partial \rho_G} > 0$ and $\frac{\partial V_\emptyset^{G,C}}{\partial \rho_{\bar{G}}} < 0$.)*

Students benefit from an increase in the other group's correlation level, but may suffer from an increase of their own correlation level.

Proof idea + Illustration

Proof idea: implicit function theorem to compute variation of the thresholds wrt the correlation coefficients (+ Gaussian assumption).

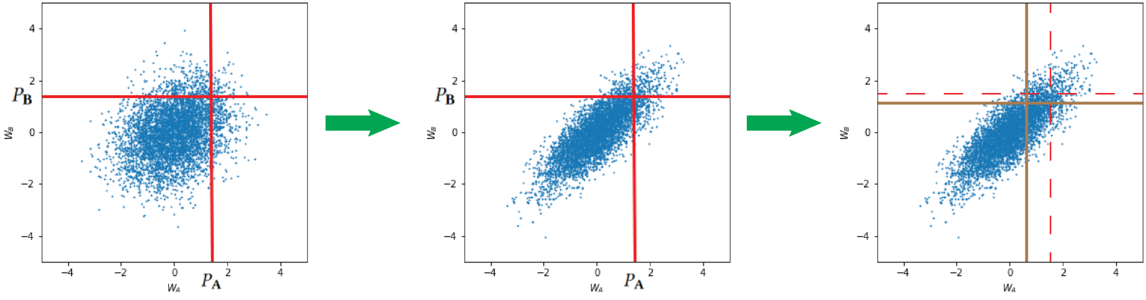


Figure: Illustration of the consequence of a correlation increase

Extension to non-Gaussian distributions

Important thing: the probabilities of type $\mathbb{P}_\rho(W_A^s < P_A, W_B^s \geq P_B)$ decrease with ρ.

It is enough to assume that (W_A^s, W_B^s) follows a joint distribution of densitiy f_θ with parameter θ (dependent on the group) such that:

- The marginals (at each college) are the same for both groups (i.e., independent of θ)
- The family f_θ is coherent: the cumulative $F_\theta(x, y)$ is increasing in θ for all x, y

Note: if f_θ is coherent, then standard correlation coefficients are increasing in θ (Scarsini, 1984).

This assumption is satisfied by natural copulas (gaussian, Archimedean).

This lets the marginal be completely free (no need for Gaussian marginals).

Conclusion

- Differential correlation alone leads to discrimination
  $\implies$ imposing "fair rankings" ($\sim$ normalization) does not implies fair matchings

- Some qualitatively counter-intuitive outcomes:
  - Differential correlation has no effect on good students.
  - Intermediate students are better off in the low correlation group.
  - An increase in the correlation level of one group will benefit good students from both groups.
  - At the same time, it will hurt intermediate students of this group and benefit those from the other group.

Open questions

- What is the effect of fixing discrimination? (and how to do it in the first place?)

- What happens with a mix of differential variance (i.e., different marginals) and differential correlation?

Thank you!

Papers:

Vitalii Emelianov, Nicolas Gast, Krishna P. Gummadi, and Patrick Loiseau.
On Fair Selection in the Presence of Implicit and Differential Variance.
EC '20 and Artificial Intelligence Journal '22

Vitalii Emelianov, Nicolas Gast, and Patrick Loiseau.
Fairness in Selection Problems with Strategic Candidates.
EC '22

Rémi Castera, Patrick Loiseau, and Bary S.R. Pradelski.
Statistical Discrimination in Stable Matchings.
EC '22

Contact:

patrick.loiseau@inria.fr