# Is It Legitimate Statistics or Is It Sexism: Why Discrimination Is Not Rational

Martha Osegueda Escobar[1], Vladik Kreinovich[1],
and Thach N. Nguyen[2]

[1]Department of Computer Science, University of Texas at El Paso, USA
mcoseguedaescobar@miners.utep.edu, vladik@utep.edu
[2]Banking University of Ho Chi Minh City, 56 Hoang Dieu 2
Quan Thu Duc, Thu Duc, Ho Chí Minh City
Vietnam,Thachnn@buh.edu.vn

# 1. There Are Statistical Differences

- People of different gender and/or ethnicity have different success rates in different disciplines.
- For example, there are many highly successful female computer scientists.
- However, in the US, in the US,
  - the percentage of female computer science students who get a PhD is lower
  - then the percentage of male students.
- In some other disciplines and in other countries, the difference is reverse.
- Similarly with ethnicity; for example,
  - the corresponding percentage is higher among Asian-American students
  - than among white students.

# 2. An Important and Difficult Challenge

- The percentage of successful females varies from country to country.
- This is true even for countries with similar ethnicity.
- This shows that the reasons for the statistical differences are not biological.
- We thus need to learn from the success of other countries and other disciplines.
- We need to make sure that everyone has an equal chance to succeed.
- This idea may sound straightforward.
- However, in reality, how to do it is an important and difficult challenge, way beyond the scope of this paper.

## 3. The Problem That We Deal With in This Talk

- In this ta;l, we deal with a more mundane problem: what is the best strategy in the current situation?
- The situation is very simple and straightforward.
- We want to graduate a certain number of PhDs.
- We have limited resources.
- So, at first glance, it seems that a rational strategy is:
  - to concentrate on undergraduate students for whom the probability of success is higher,
  - i.e., on male students,
  - and ignore the female students, since for them, the probability of success is lower.

## 4. The Problem (cont-d)

- This argument has nothing to do with prejudice against females:
  - if in a few years, the situation reverses, and the probability of a female student succeeding becomes higher than for male ones,
  - a person following this rational will start concentrating on promising female students only and ignore male students completely.
- A similar argument can be applied to hiring.
- Female applicants tend to have a higher probability of retiring early because of their family obligations.
- So should we stop hiring them?
- Should we just ignore resumes coming from female applicants and only hire males?

# 5. The Resulting Discriminatory Strategy Strategy May Sound Rational, But Is It Moral?

- The usual argument against the above hypothetical strategy is that:
  - while it may sound rational,
  - it goes against the basic moral principles.
- Everyone should get a chance to succeed.
- We should judge every person based on his/her individuality, not based on their gender, race, ethnicity.
- This is an explanation many people give.
- In this talk, we show that discriminatory strategies are not just immoral, they are actually *not rational*.

# 6. Let Us Start Analyzing the Problem

- Without losing generality, let us consider the problem of hiring.
- The same argument can be used for selecting the most promising students to "groom" them for PhD.
- For simplicity, let us assume that the candidates belong to two possible groups.
- We have a group for which the probability of success $p$ is higher.
- For simplicity, we will call this group *majority*,
- We say "for simplicity", since, e.g., Asian-Americans are not a majority.
- We also have a group for which the probability of success $p'$ is somewhat lower: $p' < p$.
- For simplicity, we will call this group *minority*.

# 7. We Will Compare Two Strategies

- Let us consider two possible strategies:
  - a *discriminatory* strategy, when we ignore all minority applicants, and
  - an *inclusive* strategy, when we consider all applicants.
- We analyze these strategies from a purely economic viewpoint: which one brings more benefit to the company.
- From this viewpoint, each of these two strategies has its gains and its losses.
- In the discriminatory strategy:
  - we save some money on analyzing minority applicants,
  - but we miss potential gains that we could have if we hired good female employees.

# 8. We Will Compare Two Strategies (cont-d)

- In the inclusive strategy:
  - we lose some money on checking the applications of all minority applicants,
  - but we may gain by hiring good female employees.
- If we combine these gains and losses, which of the two strategies will turn out to be the most beneficial?

# 9. Let Us Prepare to Evaluate Gains and Losses

- The cost of analyzing an application is approximately the same for all candidates.
- Let us denote this cost by $a$.
- There is also a cost of training a person and supporting this person through the probation period.
- Let us denote this cost by $t$.
- What can be drastically different is the gain.

# 10. Zipf Law

- Like many other things, potential gains are distributed according the *Zipf law*.
- If we denote the lifetime gain from hiring the best possible candidate by $G$, then:
  - the gain from hiring the 2nd best candidate is $\dfrac{G}{2}$,
  - the gain from hiring the 3rd best candidate is $\dfrac{G}{3}$,
  - and, in general, the gain from hiring the $i$-th best candidate is $\dfrac{G}{i}$.

# 11. Case of Inclusive Strategy

- Let us first consider the profit in the case of the inclusive strategy.
- Let us first count expenses.
- The easiest to evaluate are the expenses related to reviewing applications.
- In the inclusive strategy, we review all $N + N'$ applications.
- Reviewing each application requires amount $a$.
- So overall, we spend the amount $a \cdot (N + N')$ on these reviews.
- The next expense item is training.
- Let us assume that we have $k$ positions that we want to be eventually filled.

## 12. Case of Inclusive Strategy (cont-d)

- E.g., in the case of a university, we have $k$ tenured positions.
- Some of the people we hire will not succeed after a probation period.
- So, we hire more people to make sure that at the end, we have $k$ successful folks.
- In general, from $N$ majority candidates, $p \cdot N$ will succeed if hired.
- From $N'$ minority candidates, $p' \cdot N'$ will succeed if hired.
- Overall, if we could hire all of them, we would end up with $p \cdot N + p' \cdot N'$ successful folks.
- Out of these folks, we select $k$ best.

## 13. Case of Inclusive Strategy (cont-d)

- Out of successful folks, the probability of being among the $k$ best is the same:
  - whether it is a successful majority
  - or a successful minority.
- Thus, out of $k$ best, we will have proportionally many majority and minority folks:
  - $k_0 \stackrel{\text{def}}{=} k \cdot \dfrac{p \cdot N}{p \cdot N + p' \cdot N'}$ majority folks and
  - $k_0' \stackrel{\text{def}}{=} k \cdot \dfrac{p' \cdot N'}{p \cdot N + p' \cdot N'}$ minority folks.

## 14. Case of Inclusive Strategy (cont-d)

- For a majority applicant, the probability of success is $p$; thus:
  - to make sure that at the end, we have $k \cdot \dfrac{p \cdot N}{p \cdot N + p' \cdot N'}$ majority employees,
  - we need to hire $n_0 \stackrel{\text{def}}{=} \dfrac{k_0}{p} = k \cdot \dfrac{N}{p \cdot N + p' \cdot N'}$ majority applicants.
- For a minority applicant, the probability of success is $p'$; thus:
  - to make sure that at the end, we have $k \cdot \dfrac{p' \cdot N'}{p \cdot N + p' \cdot N'}$ minority employees,
  - we need to hire $n'_0 \stackrel{\text{def}}{=} \dfrac{k'_0}{p'} = k \cdot \dfrac{N'}{p \cdot N + p' \cdot N'}$ minority applicants.

## 15. Case of Inclusive Strategy (cont-d)

- Overall, we need to hire $n_0 + n_0'$ applicants.
- Training one hire costs the amount $t$.
- So the overall expenses on training are equal to

$$t \cdot (n + n_0) = t \cdot \frac{k \cdot (N + N')}{p \cdot N + p' \cdot N'}.$$

- Let us now count the gains.
- We considered all the applicants.
- So, we are sure that the $k$ folks that remain after the probation period are the $k$ best ones:
  - the best of these folks brings the gain $G$,
  - the second best brings the gain $\dfrac{G}{2}$, etc.,
  - the $k$-th person contributes the gain $\dfrac{G}{k}$.

# 16. Case of Inclusive Strategy (cont-d)

- The overall gain from all these folks is

$$G + \frac{G}{2} + \frac{G}{3} + \ldots + \frac{G}{k} = G \cdot \left(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{k}\right).$$

- This sum is an integral sum for the interval

$$\int_0^k \frac{1}{x}\,dx = \ln(x)|_1^k = \ln(k).$$

- So, the above sum is approximately equal to this integral $G \cdot \ln(k)$.
- Subtracting the expenses from this gain, we conclude that for the inclusive strategy, the profit is:

$$G \cdot \ln(k) - a \cdot (N + N') - t \cdot \frac{k \cdot (N + N')}{p \cdot N + p' \cdot N'}.$$

## 17. Case of Discriminatory Strategy

- In this case, we only screen $N$ majority candidates.
- So, the amount we spend on screening is $a \cdot N$ (smaller amount that for the inclusive strategy).
- We want to end up with $k$ candidates.
- We only hire majority folks, for whom the probability of success is $p$.
- Thus, to end with $k$ employees after the probation period, we need to hire $\dfrac{k}{p}$ folks.
- The cost of training all these hires is equal to $t \cdot \dfrac{k}{p}$.
- What is the gain of all these hires?
- Out of all $p \cdot N + p' \cdot N'$ potentially successful folks, we hired only the majority persons.

## 18. Case of Discriminatory Strategy (cont-d)

- So, our hiring pool consisted of $p \cdot N$ folks out of
$$p \cdot N + p' \cdot N'.$$
- The probability $p_b$ that the best of the $p \cdot N + p' \cdot N'$ folks is a majority is: $p_b = \dfrac{p \cdot N}{p \cdot N + p \cdot N'}$.
- So, in the formula for the expected gain:
  - the contribution of the best person is not $G$ (as in the case of the inclusive strategy),
  - but rather the product $p_b \cdot G = G \cdot \dfrac{p \cdot N}{p \cdot N + p \cdot N'}$.
- Similarly, the probability that the second best person is in the majority is also $p_b$.
- Thus, the contribution of this second best person into the formula for the expected gain is not $\dfrac{G}{2}$, but $p_b \cdot \dfrac{G}{2}$.

## 19. Case of Discriminatory Strategy (cont-d)

- Same with the 3rd best person, etc.
- We need to be careful now as we count further.
- We end up with $k$ employees, but they are not $k$ best folks, they are $k$ best out of *majority* folks.
- Overall, there are $p \cdot N$ potentially successful majority folks out of $p \cdot N + p' \cdot N'$ successful folks.
- Thus, when we select $k$ top majority top, there are overall $K \stackrel{\text{def}}{=} k \cdot \dfrac{p \cdot N + p' \cdot N'}{p \cdot N}$ folks of similar quality.
- Here, $K = 1 + \dfrac{p' \cdot N'}{p \cdot N}$.
- So, in counting down in quality, we have to go down to the $K$-th person.

## 20. Case of Discriminatory Strategy (cont-d)

- As a result, the overall gain for this strategy is:

$$p_b \cdot G + p_b \cdot \frac{G}{2} + p_b \cdot \frac{G}{3} + \ldots + p_b \cdot \frac{G}{K} = p_b \cdot G \cdot \left(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{K}\right).$$

- Here, as before, $1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{K} \approx \ln(K)$, where

$$\ln(K) = \ln\left(k \cdot \left(1 + \frac{p' \cdot N'}{p \cdot N}\right)\right) = \ln(k) + \ln\left(1 + \frac{p' \cdot N'}{p \cdot N}\right).$$

- Thus, for the discriminatory strategy, the gain is:

$$\frac{p \cdot N}{p \cdot N + p \cdot N'} \cdot G \cdot \left(\ln(k) + \ln\left(1 + \frac{p' \cdot N'}{p \cdot N}\right)\right).$$

- By subtracting the expenses from this gain, we conclude that the profit of using this strategy is:

$$\frac{p \cdot N}{p \cdot N + p \cdot N'} \cdot G \cdot \left(\ln(k) + \ln\left(1 + \frac{p' \cdot N'}{p \cdot N}\right)\right) - a \cdot N - t \cdot \frac{k}{p}.$$

## 21. So Which of the Two Strategies Is the Most Profitable?

- In several realistic numerical examples that we tried:
  - the profit from the inclusive strategy
  - exceeds the profit from the discriminatory strategy.
- To have a general result, let us consider the case:
  - when what we called "minority" is really a minority,
  - i.e., when the ratio $m \stackrel{\text{def}}{=} \dfrac{N'}{N}$ is small,
  - so that we can ignore terms which are quadratic or of higher order in terms of $m$.
- Let us denote $r \stackrel{\text{def}}{=} \dfrac{p'}{p} < 1$.

## 22. Which Strategy Is More Profitable (cont-d)

- Dividing both numerator and denominator of the training-expenses term in the inclusive formula, we get:

$$t \cdot \frac{k \cdot (N + N')}{p \cdot N + p' \cdot N'} = t \cdot \frac{k(1 + m)}{p + p' \cdot m}.$$

- Here, $p + p' \cdot m = p \cdot (1 + r \cdot m)$, and

$$\frac{1}{1 + r \cdot m} = 1 - r \cdot m + o(m).$$

- Thus, the training-expenses term takes the form

$$t \cdot \frac{k}{p} + t \cdot \frac{k}{p} \cdot (1 - r) \cdot m.$$

- So, the profit from using the inclusive strategy is:

$$G \cdot \ln(k) - a \cdot N - t \cdot \frac{k}{p} - a \cdot N \cdot m - t \cdot \frac{k}{p} \cdot (1 - r) \cdot m.$$

## 23. Which Strategy Is More Profitable (cont-d)

- Similarly, terms in the discriminatory formula take the following form:

$$\frac{p \cdot N}{p \cdot N + p' \cdot N'} = \frac{1}{1 + m \cdot r} \approx 1 - m \cdot r;$$

$$\ln\left(1 + \frac{p' \cdot N'}{p \cdot N}\right) = \ln(1 + r \cdot m) \approx r \cdot m.$$

- Thus, this formula takes the following form:

$$G \cdot \ln(k) - a \cdot N - t \cdot \frac{k}{p} - G \cdot \ln(k) \cdot r \cdot m + G \cdot r \cdot m.$$

## 24. Which Strategy Is More Profitable (cont-d)

- So, in comparison with case $m = 0$, we lose the following amounts proportional to $m$:
  - in the discriminatory case, we lose the amount proportional to $G \cdot (\ln(k) - 1)$, while
  - in the inclusive case, we lose the amount proportional to $a \cdot N + t \cdot \dfrac{k}{p} \cdot (1 - r)$.
- Let us take into account that:
  - even for the weakest of the $k$ hires, for whom the gain is equal to $\dfrac{G}{k}$,
  - this gain is still much larger than all the expenses on selection and training,
  - otherwise, the company would not be hiring this person in the first place.

## 25. Which Strategy Is More Profitable (cont-d)

- The expenses of selecting a person are equal to $a$.
- The expenses of training $\dfrac{1}{p}$ persons (needed for one person to succeed) are $t \cdot \dfrac{1}{p}$; thus:

$$\frac{G}{k} \gg a+t\cdot\frac{1}{p},\ \text{hence } G \gg a\cdot k+t\cdot\frac{k}{p} \text{ and } G \gg a\cdot k+t\cdot\frac{k}{p}\cdot(1-r).$$

- So, the discriminatory-strategy loss is larger than the inclusive-strategy loss if

$$G\cdot((\ln(k)-1)\cdot r-1) \geq a\cdot(N-k).$$

- Then, by adding the last inequality, we would get the desired one.

## 26. Which Strategy Is More Profitable (cont-d)

- This last inequality is definitely true:
  - even the gain $\dfrac{G}{k}$ of the least productive hire
  - is of the same order as this person's lifetime salary,
  - i.e., in the US, several million dollars,
  - while the cost of scanning all $N$ candidates is much smaller.
- Thus, *the inclusive strategy is indeed economically preferable.*

## 27. Comment

- From the purely mathematical viewpoint, the discriminatory strategy can be more profitable.
- For example, when the probability $p'$ of the minority hire's success is close to 0:
  - there is no gain in hiring them,
  - only additional expenses in screening and training.
- However, in practice, the ratio $p'/p$ is not 0:
  - it can be 0.5, even somewhat less
  - but still sufficiently positive to make sure that the inclusive strategy is economically preferable.

## 28. Acknowledgments

This work was supported in part by the National Science Foundation grant HRD-1242122 (Cyber-ShARE Center).