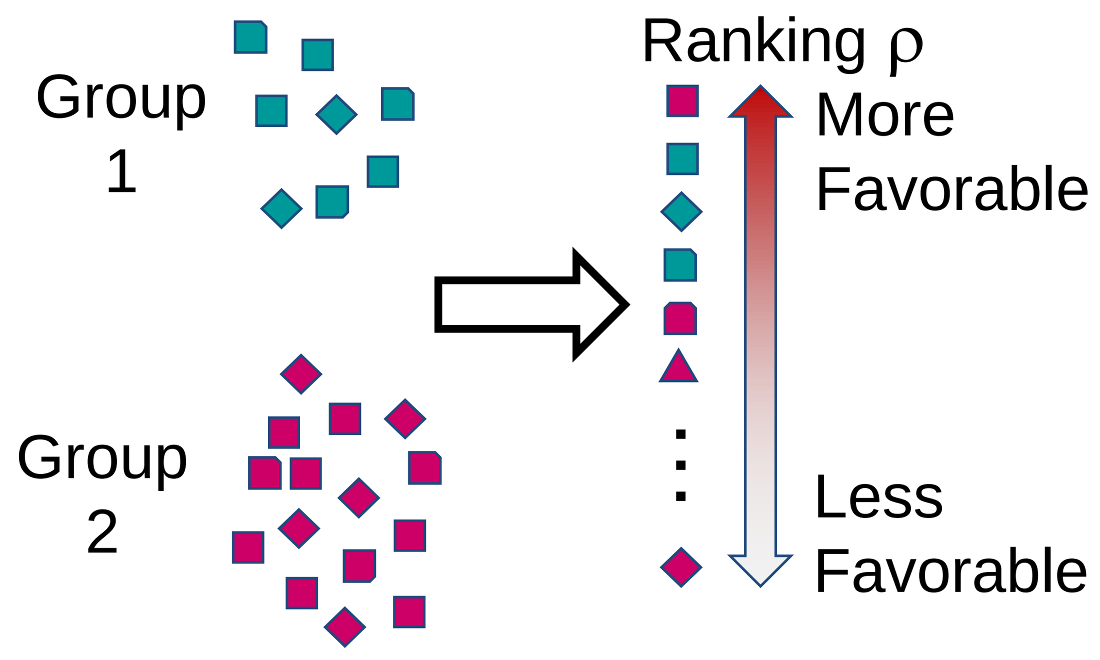
Measuring Group Advantage: A Comparative Study of Fair Ranking Metrics

Caitlin Kuhlman, Walter Gerych, Elke Rundensteiner

{cakuhlman,wgerych,rundenst}@wpi.edu

Worcester Polytechnic Institute Worcester, USA

FAIR RANKINGS



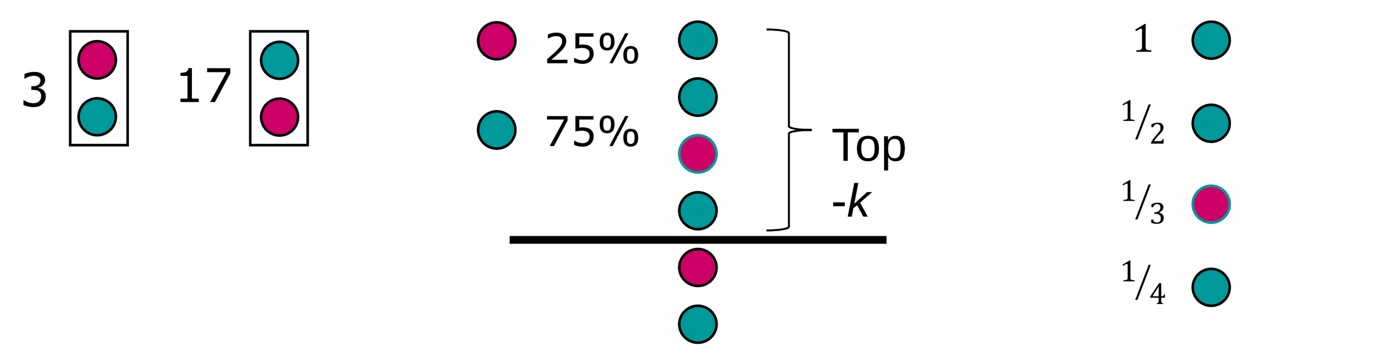
Question:

Are the groups treated equally?

STATISTICAL PARITY METRICS

Three major kinds of statistical parity fairness metrics:

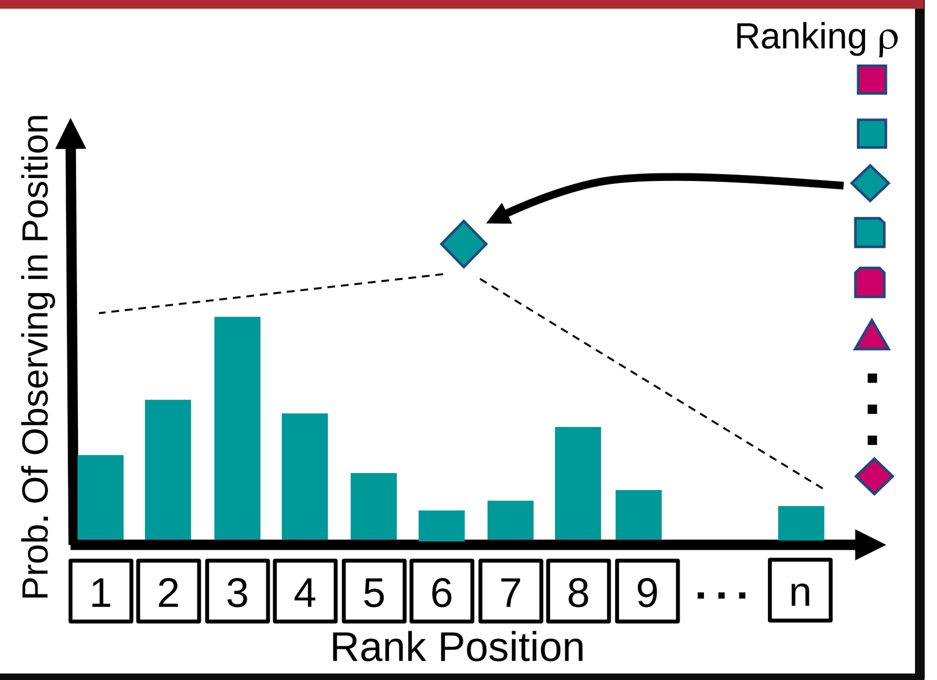
Pairwise | **Top-*k*** | **Exposure**



Our work is the first comprehensive comparison of these three metric types

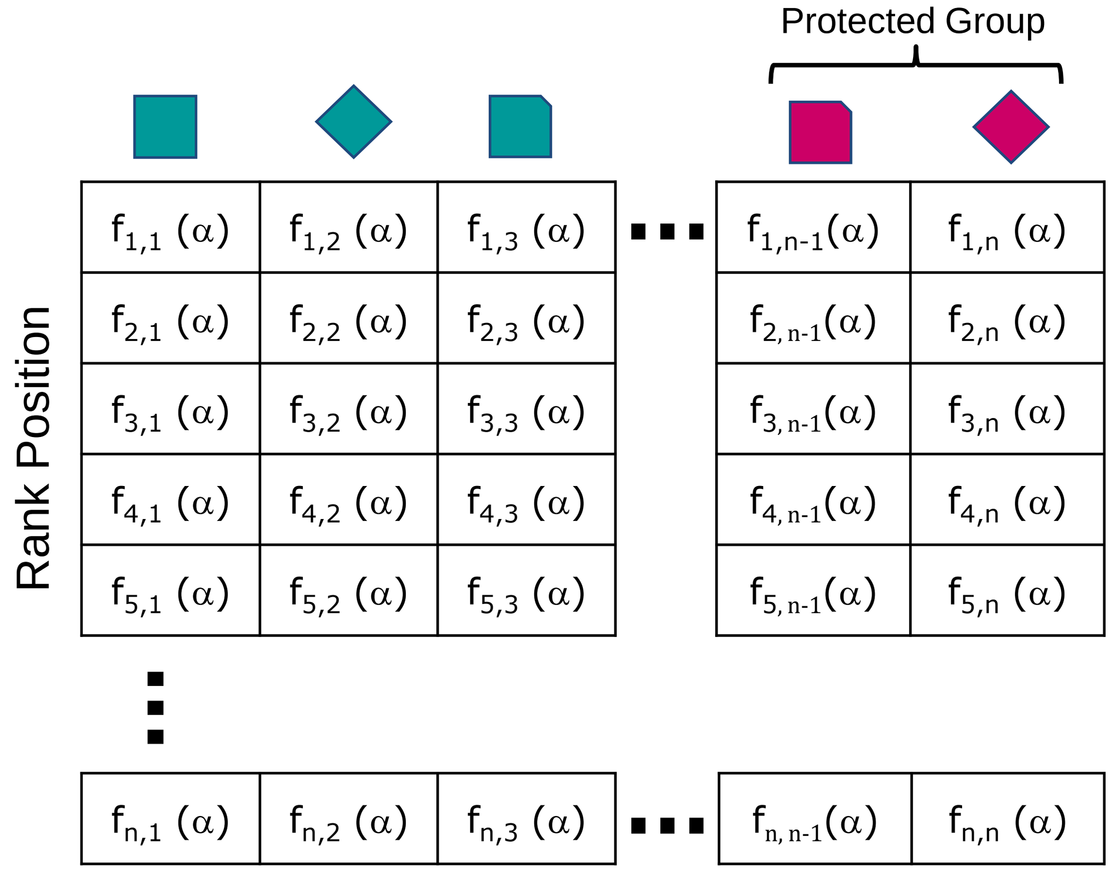
STOCHASTIC RANKING

- Item has probability of being in any rank position
- Models many real-world ranking scenarios
  - Multiple rankers
  - Unbiased click feedback in IR
- Statistical parity metrics evaluated on *expectations* of stochastic rankings



RANKING MATRIX FRAMEWORK

- Propose unified ranking matrix framework for metric evaluation and comparison
  - Columns = candidates
  - Rows = rank positions
- *i,j*-th cell represents probability candidate *j* is in position *i*
  - Probability represented by function $f_{i,j}$: $[0,1] \rightarrow [0,1]$
- Function of *advantage* α
  - α = 0 implies total disadvantage, α = 1 total advantage
  - Advantage with respect to "*protected group*"
    - One group designated as the protected group
- Assumptions: Probability that protected group monotonically increasing/decreasing if α greater/less than group proportion



α: Advantage of protected group

$f_{i,j}(\alpha)$: Probability object *j* is in position *i*

ASSUMPTION 1.

$$\sum_{j \in G_p} f_{i,j}(\alpha) \geq \sum_{j \in G_p} f_{i+1,j}(\alpha) \quad if\ \alpha \geq \frac{|G_p|}{N}$$

ASSUMPTION 2.

$$\sum_{j \in G_p} f_{i,j}(\alpha) \leq \sum_{j \in G_p} f_{i+1,j}(\alpha) \quad if\ \alpha \leq \frac{|G_p|}{N}$$

THEORETICAL ANALYSIS

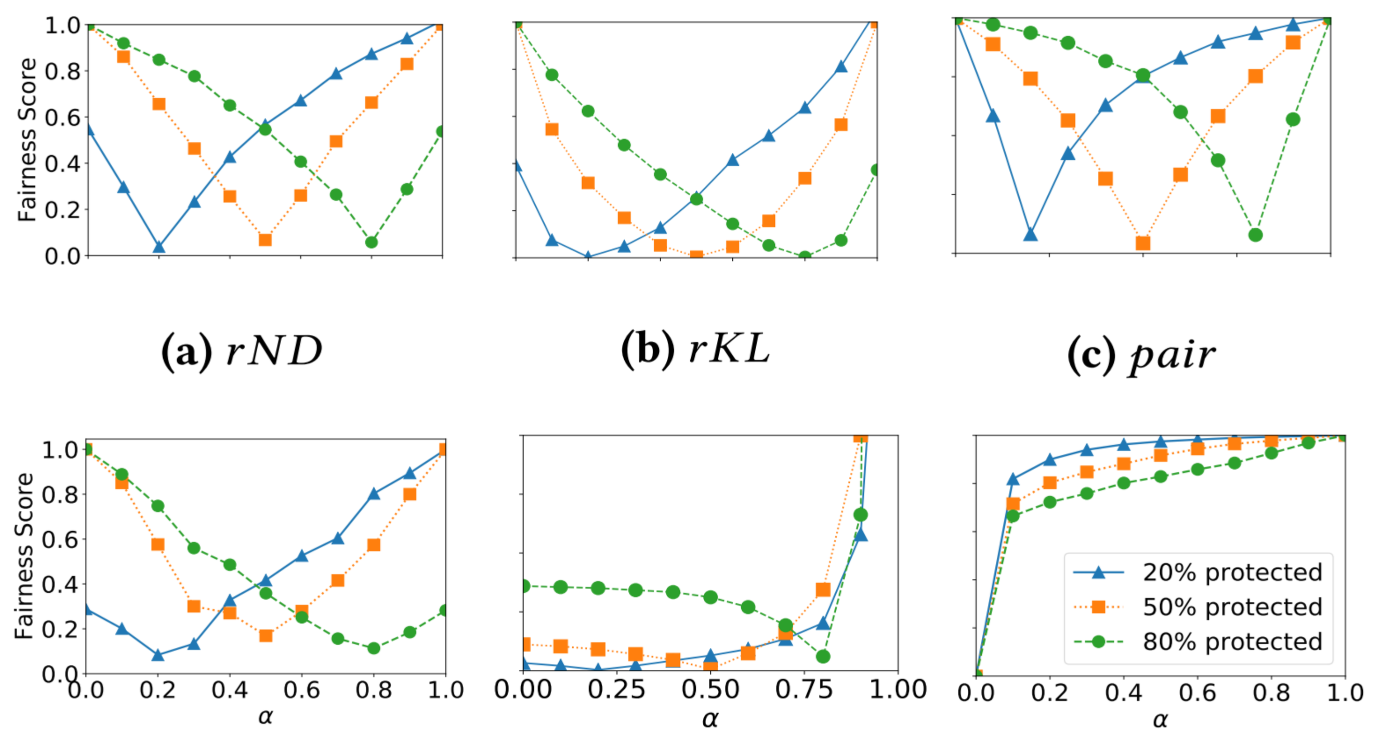
- We compare the standard statistical parity metrics:
- ➢ *rND, rRD, skew, rKL, expRR, pair*
- Surprising result: All metrics besides skew behave the same when Assumption 1+2 are met
- ➢ *Proof intuition: Metrics have same minima and signs of derivates are equal everywhere*
- ***Optimizing for one metric optimizes for all***
- We propose test to determine if theoretical assumptions hold for real data
- ➢ Statistical equivalence to monotonic function

THEOREM 1. *Given a ranking ρ with a protected group of candidates G_p and associated advantage α, if Assumptions 1 and 2 hold, then the rND, rRD, rKL, expRR, and pair metrics share the same minima.*

THEOREM 2. *Given a ranking ρ with a protected group of candidates G_p and associated advantage α, if Assumptions 1 and 2 hold, then signs of the derivative with respect to α of the rND, rKL, rRD, and expRR metrics are the same.*

EXPERIMENTAL RESULTS

- Simulated rankings with different group sizes
- Results match analysis: all but skew have same minima/slopes
- **Key difference**: *pair* treats absolute (dis)advantage equally regardless of group size, others won't flag complete disadvantage minority
- ➢ Crucial, as fairness evaluation might be most needed for small minority groups



(a) *rND* (b) *rKL* (c) *pair*

(d) *expRR* (e) *rRD* (f) *skew*

- Code for experiments, including cases where rankings do not conform to our assumptions, available at: https://github.com/waltergerych/AIES_2021_Measuring_Group_Advantage

ACKNOWLEDGEMENTS

- **Financial Sources**
  - NSF FAIR Grant 2007932, DARPA WASH Grant HR00111780032-WASH-FP-031
- **Collaborators**
  - Kathleen Kennedy, DSRG @ WPI, DAISY Lab @ WPI

REFERENCES

- Caitlin Kuhlman, MaryAnn VanValkenburg, and Elke Rundensteiner. 2019. FARE: Diagnostics for fair ranking using pairwise error metrics. WWW.
- Harikrishna Narasimhan, Andy Cotter, Maya Gupta, and Serena Lutong Wang. 2020. Pairwise Fairness for Ranking and Regression. AAAI.
- Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. 2017. Fa* ir: A fair top-k ranking algorithm. CIKM.
- Ke Yang and Julia Stoyanovich. 2017. Measuring fairness in ranked outputs. SSDBM.
- Sahin Cem Geyik, Stuart Ambler, and Krishnaram Kenthapadi. 2019. Fairnessaware ranking in search & recommendation systems with application to LinkedIn talent search. KDD.
- Ashudeep Singh and Thorsten Joachims. 2018. Fairness of Exposure in Rankings. KDD.
- Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. 2018. Equity of attention: Amortizing individual fairness in rankings. SIGIR.