FARE - Fair Allocation RE-weighting

April Chen [1*] **and Prajna Soni** [2*]

[1] Harvard SEAS
[2] MIT IDSS
aprilchen@college.harvard.edu, prajna@mit.edu

Abstract

Resource allocation is a common problem in many setting such as lending, medicine, and policing. Such problems can be modeled as the allocation of a limited resource to different groups. The problem of unknown distribution, missing-not-at-random (MNAR) data, and biased historical data are commonly encountered in these settings and thus increase the challenge of fairly allocating resources in a manner that does not perpetuate biased cycles. In this paper, we analyze the fair allocation algorithm proposed by (Elzayn et al. 2019) and argue that assuming observed distributions as ground truth inputs to learn a fair allocation can further biased cycles in many such high-risk applications. We illustrate this using the New York City Fare Evasion dataset and instead propose, FARE, a modified algorithm that implements subgroup-level reweighting to mitigate the effects of biases present in historical datasets and thus inform fairer allocations. We deploy this on the NYC Fare Evasion dataset and illustrate significantly fairer allocations in comparison to the original algorithm. We hope to highlight the necessity for context-driven modifications to algorithms to improve fairness.

Introduction

In this paper, we explore the validity of proposed fair algorithms for limited resource allocation, specifically that by (Elzayn et al. 2019) and propose an alternative method to improve subgroup fairness. We contextualize resource allocation and the algorithm in the case of subway fare evasion in New York City, using the NYC Fare Evasion dataset cleaned and collected as part of the MIT CSAIL NYC Fare Evasion project. We highlight the need for modifying historical data in the context of high risk settings like crime statistics and lending where datasets represent decades of historical societal biases, and illustrate that our proposed reweighting algorithm improves subgroup fairness over the baseline optimal allocation algorithm while maintaining or improving utility.

Background: Context of Crime Statistics

Arguably the greatest challenge in crime modeling is that crime datasets are one-sided and composed primarily of true positives. Crime datasets only contain records of those actually arrested and no record of criminals who go uncaught. Furthermore, because of well-documented bias in policing, we have good reason to believe that missing data in crime datasets are not missing at random (Spencer, Charbonneau, and Glaser 2016) (Goncalves and Mello 2021) (Knox, Lowe, and Mummolo 2019). If a location does not have police officers deployed, crimes are more likely to go unobserved and thus not recorded in the dataset. Similarly, if an officer is more biased against a subgroup and is more likely to arrest offenders from that subgroup then the dataset would indicate a higher number of arrests for that subgroup and thus imply that that race has a higher rate of offense. This being said, modeling crime is still important. One study by Sanchez-Martinez estimated that Boston loses \$ 3,600 every weekday due to fare evasion. (Sánchez-Martínez 2017). Ideally, one should find more efficient ways to deal with crime fairly.

Many researchers attempt to deal with missing data in modeling. This problem is simple when data are missing completely at random (Little and Rubin 2019). However, other methods attempt to address the more relevant case for us, involving missing and biased data. Bertsimas et al. attempt to address this by imputing and optimizing missing data (Bertsimas, Pawlowski, and Zhuo 2017). Penny et al. use attention and latent space regularization (Penny-Dimri, Bergmeir, and Smith 2022). However, all widely used methods require assumptions about the joint distributions of variables in the dataset which cannot be made about crime statistics. Lum and Isaac explore this feedback loop in the over-policing of poor and minority neighborhoods and raise important questions about the validity and impact of using biased data in the context of crime statistics and the bias it perpetuates (Lum and Isaac 2016). Garg and Liu identify reporting rates without using ground truth data by leveraging duplicate reports on the same incident and found substantial spatial and socio-economic disparities (Liu and Garg 2022).

Related Work

Our fair algorithm for resource allocation is based on the algorithm proposed by Elzayn et al. which models the censored feedback problem (Elzayn et al. 2019). In that work, the authors assume that the observed arrest data is the ground truth and design a greedy algorithm that iterates to maximize utility while ensuring the probability of discovery of

[*] These authors contributed equally.

any candidate in a group is at most α different from every other group. Donahue and Kleinberg build upon the family of resource allocation problems described by Elzayn et al. and provide a theoretical bound to such an α-fair allocation for various distributions to illustrate the interaction of utility and fairness (Donahue and Kleinberg 2019).

Specific to the context of fare evasion, some work has been done into exploring the use of Stackelberg games as a means of mitigating fare evasion (Yin et al. 2012) (Sandholm et al. 2012) (Varakantham, Lau, and Yuan 2013) (Fave et al. 2014). For example, in 2012, the Los Angeles Metro Rail system worked with researchers in the field to create TRUSTS, which was used to schedule randomized patrols for fare inspection in transit systems to deter fare evasion and maximize revenue (Yin et al. 2012) (Sandholm et al. 2012). While these are possible effective formulations of the question, in this paper, we focus on evaluating and extending the concept of fairness in such a setting using Elzayn et al's formulation (Elzayn et al. 2019).

Ethical Statement

The contents of this paper can be evaluated at 3 scales: 1) the criminalization of fare evasion and the use of data-driven insights to mitigate fare evasion, 2) the inherent properties of crime statistics and the implications of data-driven policing, and 3) the use of allocation algorithms in contexts with biased and missing-not-at-random (MNAR) data. We attempt to outline a few considerations at each of these scales, however, also recognize that the depth and nuance required to fully explore each of these scales would require separate papers. Thus, we reference works by other authors to provide more context to the conversation and note that this paper aims to highlight fairness considerations in resource allocation algorithms that have been proposed in the literature. We hope to contribute to the literature by providing a more nuanced and context-driven approach to these proposed algorithms in the general context of resource allocation and motivate the need for such an approach by illustrating adverse outcomes in the example of crime statistics.

The criminalization of fare evasion is a debated topic (Carter and Johnson 2021) (Stolper and Jones 2017) (Jones and Stolper 2018). Ethical concerns include "last mile" officer bias, affordability, and societal factors surrounding fare evasion – essentially, is criminalizing fare evasion just criminalizing poverty? Furthermore, should those caught evading fair be charged with an arrest, or are there non-disruptive ways to prevent it (Smith and Clarke 2000)?

While we explore the algorithm in the context of the NYC fare evasion dataset, we hope to show the broader concerns with crime statistics and data-driven policing approaches which are based on historical events laden with bias (Babuta and Oswald 2019) (Bennett Moses and Chan 2018). The exploration of NYC fare evasion as a context for the proposed algorithm **does not mean** that the authors of this paper support the criminalization of fare evasion or the use of data drive statistics to mitigate fare evasion. Instead, we hope to use this example to draw attention to the need for nuanced and context-driven approaches to fairness. In the case of criminal justice and policing, it is also important to acknowledge that many societal, systemic, and economic factors feed into the trends that are observed at any point (Roberts 2007) (Richardson 2017) (Brewer and Heitzeg 2008). Therefore, in addition to debiasing criminal statistics, much more work must be spent tackling these issues from their societal roots.

Problem Formulation

Defining Baseline Fair Allocation Algorithm

Elzayn et al. (2019) propose and contextualize their algorithm using the Philadelphia Crime Incidents dataset. They frame the question of allocating police officers to districts as resources (i.e. police officers) being allocated to groups (i.e. districts) given different distributions of candidates eligible to receive the resource across the groups (i.e. law offenders/criminals). The problem is a censored feedback setting where the algorithm does not observe the total number of candidates (i.e. offenders) in each group, but only the number of candidates that received the resource (i.e. arrested offenders). Here we use a similar framing of the problem as described by Elzayn et al. with V units of a resource being allocated across G groups each with c_i candidates drawn from a fixed unknown and group-dependent distribution C_i. The goal of the algorithm is to obtain allocation $v = [v_1, v_2, \ldots, v_G]$ given the number of observed candidates in each group given a certain allocation v_i. Using this problem formulation, Elzayn et al. define an α-fair allocation as an allocation where the probability of discovery of a candidate is within a certain bound, α, across all groups. To achieve this in the context of the Philadelphia open crime dataset, they fit historical arrests to a Poisson distribution and use these as ground truth distributions while iterating to learn an α-fair optimal allocation (Elzayn et al. 2019).

Inferring Distributions from Historical Data

We argue that in the context of crime statistics, where historical data is inherently ridden with human and societal biases against certain subgroups, assuming observations and inferred distributions as ground truths creates a feedback cycle that perpetuates the bias against these subgroups.

We can ground this argument in the context of the NYC fare evasion dataset to visualize and understand the effect of this. Given the previous formulation, we define the following for this context.

- Resource = subway officers
- Candidates = fare evaders
- Groups = subway stations

Assuming station-wise Poisson distributions derived from the historical arrests at each station are the ground truth distributions and feeding them into the algorithm would mean assuming that the distribution of arrests within each station is also the ground truth. To further clarify, it would imply that the race-wise and gender-wise composition of arrests at each station is also the ground truth. Often, these distributions are heavily skewed and as illustrated by Rothbacher et al. (2019) in the CSAIL NYC Fare Evasion project, minority groups like Black and Hispanic communities in NYC,

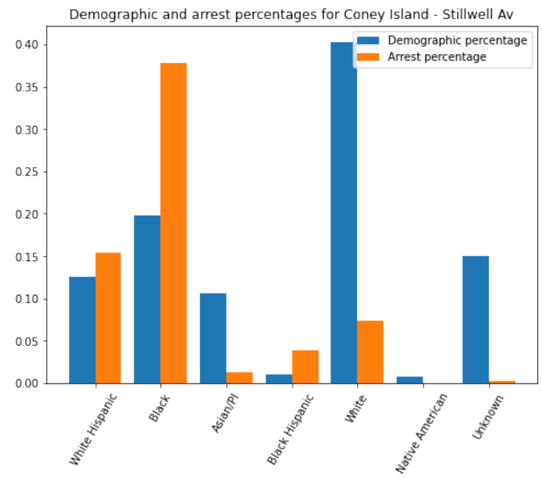


Figure 1: Black and Hispanic people are disproportionately arrested. This demonstrative example is for one station, but the trend holds true for the vast majority of stations.

face higher levels of policing and thus a higher number of absolute arrests.

In this paper, we formulate the distribution of arrests at each station as a composite distribution of independent per-race distributions (Eqn. 1).

$$dist_{station}(\mu) = \sum_{r \in races} dist_r(\mu) \tag{1}$$

Following Elzayn et al. (2019), we model the independent per-race distributions as Poisson distributions fit to the observed arrests at NYC subway stations. The NYC Fare Evasion dataset provides us with a distribution of race-wise arrests at a given subway stop. Therefore, given the number of race-wise entries and exits at a stop, we can calculate the expected race-wise arrest rate. The race-wise entries are calculated using turnstile entry data and demographic data near the station as provided by ACS. Given the lack of more detailed entry data, we assume that the demographic distribution of residents surrounding a subway stop is equal to the distribution of passengers entering a subway stop.

Reweighting Proposal

Provided this apparent bias in historical data, we argue that in order to break a biased feedback loop, an algorithm cannot plainly assume the data-driven inferences. Instead, we propose a novel reweighted discovery model that equalizes subgroup properties that are conditionally independent.

In the context of the NYC Fare Evasion dataset, we argue that it is unfair to assume that given a subway stop, one race is more likely to offend than another. While crime rates may be heavily tied to socio-economic factors, a dependency on race is the perpetuation of historical biases and stereotypes. Therefore, we introduce a median-reweighting of inferred Poissons to ensure that the assumed arrest rates of a passenger at a given subway stop are independent of their race.

$$\sum_{r \in races} dist_r(\mu_{reweighted}) = dist(\sum_{r \in races} \mu_{reweighted,r}) \tag{2}$$

To reweight the station-wise arrest rate, we choose the median arrest rate at a station and equate the total station arrest-rate to it. Note that median, in this case, is meant in the distributional sense – the median is the arrest rate corresponding to the 50th percentile of the population, when organized by race in order of arrest rate. Using the median arrest rate helps preserve any signal that the location of a subway stop might have while removing the effects of any extreme biases towards or against a given race which might influence the station's rate like the use of a mean might have. Depending on the fairness definition used in the specific resource allocation context of the algorithm, this reweighting can be flexibly adapted. For example, if an allocation is fair such that it counters historical biases or inequities, the reweighting factors can be adopted to skew resource allocation towards a certain group.

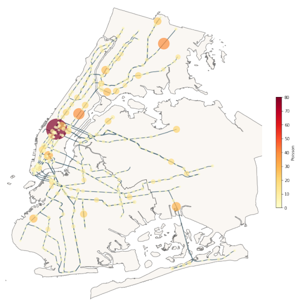
Experiments

To illustrate the necessity for a nuanced approach to learning distributions from historical data, we motivate this paper using the New York City (NYC) fare evasion dataset curated by Nick Rothbacher et al. (2019) and evaluate fair allocation algorithms in the context of crime statistics, specifically focusing on fare evasion arrests in NYC. We first fit the monthly crime rates at each station to a Poisson distribution, as described in Problem Formulation. We then implement the algorithms from Elzayn et al. where we seek to learn these true distributions λ^* based on simulated observed crime rates. In each iteration i, we attempt to learn λ^* of crime by using our existing knowledge to create a fair allocation v_i, using a simplifying discovery model $disc(v_i, c_i) = min(v_i, c_i)$ to calculate the number of crimes o_i observed given that allocation and true number of candidates c_i, and then using a Nelder-Mead optimizer to solve for the most likely distribution $\hat{\lambda}$ based on the history of crime observations. We repeat this for 10 iterations for each. Note that for the purposes of these experiments, we set $\alpha = 1$ to avoid setting fairness constraints across subway stations, since we are interested in the impact of reweighting race in an otherwise optimal case. We then analyze the output allocation. Due to computational constraints (the original algorithm runs in $O(n^2)$ time) we run this for sets of 10 stations each.

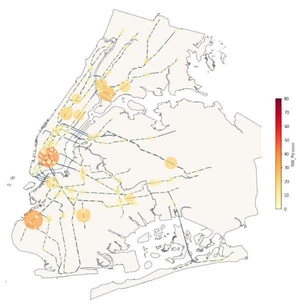
The NYC Fare Evasion dataset consists of arrests that took place within 250m of a subway stop in NYC from 2010 to 2018. It includes demographic information about arrested fare evaders including age, gender, and race, as well as metadata about the subway stops such as the number of entries and exits observed at the turnstiles. It is also joined with demographic data of each station's neighborhood obtained from the American Community Survey (Bureau 2022).

Results

First, we investigate the estimated number of arrests at each subway station, based on fitting Poisson distributions to historical data. As discussed earlier, these estimated rates contain implicit historical bias, which we attempt to mitigate with our reweighting algorithm. We then compare this to the estimated number of arrests at each subway station after reweighting in Figure 2. Note that the largest number of expected arrests, which correspond mostly to stations in majority-minority areas, decrease significantly.



(a) Arrest rates before reweighting.



(b) Arrest rates after reweighting.

Figure 2: Arrest rates as modeled by Poisson distribution parameters for subway stations before and after reweighting. Note the maximum values drop by more than half.

We also evaluate our method in terms of several fairness metrics. First, we use *Price of Fairness* (PoF). Intuitively, this is the ratio of crimes discovered by an optimal allocation divided by the number of crimes discovered by an allocation given our reweighted estimates, i.e.

$$PoF(rw) = \frac{\chi^*}{\chi^{rw}} \tag{3}$$

To measure the actual fairness of our method, we measure v_{BH}, the proportion of police allocated to majority Black-/Hispanic neighborhoods, as well as Δ_{BH}, the distance between the distribution of police and the distribution of the proportions of Black and Hispanic people in each neighborhood. Intuitively, in a fair police allocation, the first metric should be low and the second should be high. We note that while these are useful metrics in determining the trade-offs of fairness, it is also important to acknowledge that the historical data is itself biased, so the true cost of fairness cannot be determined unless the true underlying distributions of crimes were known.

As shown in Table 1, allocations generated using our reweighting method perform better in **all cases** on these fairness metrics, while keeping PoF relatively well-bounded.

Stations	**PoF**	$\mathbf{v}^*_{\mathbf{BH}} \downarrow$	$\mathbf{v}^{\mathbf{rw}}_{\mathbf{BH}} \downarrow$	$\mathbf{\Delta}^*_{\mathbf{BH}} \uparrow$	$\mathbf{\Delta}^{\mathbf{rw}}_{\mathbf{BH}} \uparrow$
1-10	1.36	0.76	0.10	0.14	0.98
11-20	1.033	0.2	0.07	0.14	0.75
21-30	0.87	0.03	0.03	0.5	0.75
31-40	1.29	0.4	0.16	0.47	1.79

Table 1: Price of Fairness and fairness metrics for sets of 10 stations at a time. Arrows indicate which direction is fairer. A PoF of 1 means utility does not decrease with reweighting.

Limitations and Implications

Human Bias: The algorithm can only try to redistribute and account for over-policing in a geographical sense based on observed historical disparities in arrest rates, however, the decisions of an officer on the ground in terms of who they decide to arrest/stop is a factor that is beyond the allocation algorithm.

Learning Time: The algorithm is currently set on a monthly cycle and the learning part to uncover lambdas requires deployment with an initial estimate and then requires calibration. In our simulation, police bias remains constant (i.e. as it has been historically).

Location Dependency: Our algorithm, unless it observes equal arrest rates across subway stops, will still vary its allocation across neighborhoods. Since neighborhoods are known to be racially segregated, this could still lead to bias.

Conclusion

We present a method for mitigating subgroup bias in data-driven resource allocation settings. We test this algorithm in the context of allocating officers to NYC subway stops to mitigate fare evasion and illustrate that while it remains an enormous challenge to accurately approximate underlying distributions given only biased historical data, our method creates fairer allocations without significant cost to utility. While such an intervention cannot replace systemic changes to reduce bias and factors contributing to crime, it provides an improvement upon existing algorithms and most importantly, illustrates the need to account for context-driven subgroup fairness in resource allocation.

Future work

Further evaluation should investigate ideal optimization parameters for this method, including the impact on fairness and utility of reweighting aggregation functions, such as average or minimum in contrast to the median. Additionally, the role of the group-fairness parameter α Elzayn et al. (2019), combined with the impact of our reweighting algorithm is an avenue yet to be explored. Furthermore, extending the proof of the bound in Donahue and Kleinberg (2019) to our multidimensional context would be interesting. Finally, significant future work could explore reweighting in positive allocation settings (e.g. allocation of healthcare, education, or funding).

References

Babuta, A.; and Oswald, M. 2019. Data analytics and algorithmic bias in policing. *The royal united services institute for defence and security studies*.

Bennett Moses, L.; and Chan, J. 2018. Algorithmic prediction in policing: assumptions, evaluation, and accountability. *Policing and society*, 28(7): 806–822.

Bertsimas, D.; Pawlowski, C.; and Zhuo, Y. D. 2017. From predictive methods to missing data imputation: an optimization approach. *J. Mach. Learn. Res.*, 18(1): 7133–7171.

Brewer, R. M.; and Heitzeg, N. A. 2008. The racialization of crime and punishment: Criminal justice, color-blind racism, and the political economy of the prison industrial complex. *American Behavioral Scientist*, 51(5): 625–644.

Bureau, C. 2022. 2011-2015 ACS 5-year Estimates.

Carter, T. J.; and Johnson, L. T. 2021. "Blacks Can't Jump": The Racialization of Transit Police Responses to Fare Evasion. *Race and Justice*, 21533687211007548.

Donahue, K.; and Kleinberg, J. M. 2019. Fairness and Utilization in Allocating Resources with Uncertain Demand. *CoRR*, abs/1906.09050.

Elzayn, H.; Jabbari, S.; Jung, C.; Kearns, M.; Neel, S.; Roth, A.; and Schutzman, Z. 2019. Fair Algorithms for Learning in Allocation Problems. New York, NY, USA: Association for Computing Machinery. ISBN 9781450361255.

Fave, F. M. D. F.; Jiang, A. X.; Yin, Z.; Zhang, C.; Tambe, M.; Kraus, S.; and Sullivan, J. 2014. Game-Theoretic Security Patrolling with Dynamic ExecutionUncertainty and a Case Study on a Real Transit System. *Journal of Artificial Intelligence Research 50 321-367*.

Goncalves, F.; and Mello, S. 2021. A few bad apples? Racial bias in policing. *American Economic Review*, 111(5): 1406–41.

Jones, J.; and Stolper, H. 2018. NO EASY PASS FOR TRANSIT RIDERS.

Knox, D.; Lowe, W.; and Mummolo, J. 2019. The bias is built in: How administrative records mask racially biased policing.

Little, R. J.; and Rubin, D. B. 2019. *Statistical analysis with missing data*, volume 793. John Wiley & Sons.

Liu, Z.; and Garg, N. 2022. Equity in Resident Crowdsourcing: Measuring Under-reporting without Ground Truth Data.

Lum, K.; and Isaac, W. 2016. To predict and serve? *Significance*, 13: 14–19.

Penny-Dimri, J. C.; Bergmeir, C.; and Smith, J. 2022. Dealing with missing data using attention and latent space regularization. *arXiv preprint arXiv:2211.07059*.

Richardson, L. S. 2017. Systemic triage: Implicit racial bias in the criminal courtroom. *Yale Law Journal*, 126(3).

Roberts, D. E. 2007. Constructing a criminal justice system free of racial bias: An abolitionist framework. *Colum. Hum. Rts. L. Rev.*, 39: 261.

Sandholm, T.; Yin, Z.; Jiang, A. X.; Tambe, M.; and Kiekintveld, C. 2012. Towards Optimal Patrol Strategies for Fare Inspection in Transit Systems. *AAAI Publications*.

Smith, M. J.; and Clarke, R. V. 2000. Crime and public transport. *Crime and Justice*, 27: 169–233.

Spencer, K. B.; Charbonneau, A. K.; and Glaser, J. 2016. Implicit bias and policing. *Social and Personality Psychology Compass*, 10(1): 50–63.

Stolper, H.; and Jones, J. 2017. The crime of being short $2.75. *Community Service Society*, 1–34.

Sánchez-Martínez, G. E. 2017. Estimating Fare Noninteraction and Evasion with Disaggregate Fare Transaction Data. *Transportation Research Record*, 2652(1): 98–105.

Varakantham, P.; Lau, H. C.; and Yuan, Z. 2013. Scalable Randomized Patrolling for Securing Rapid Transit Networks. *AAAI Conference on Artificial Intelligence; Vol. 27 No. 2: Twenty-Fifth Conference on Innovative Applications of Artificial Intelligence*.

Yin, Z.; Jiang, A. X.; Tambe, M.; and Kiekintveld, C. 2012. TRUSTS: Scheduling Randomized Patrols for Fare Inspection in Transit Systems Using Game Theory. *CMU*.