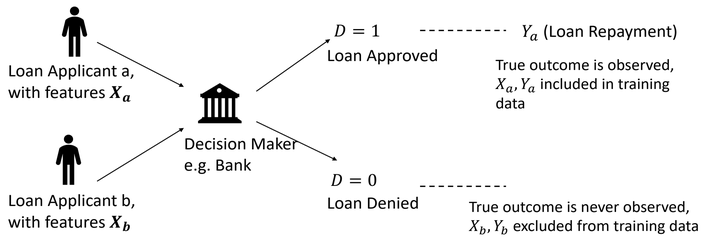
The Importance of Modeling Data Missingness in Algorithmic Fairness

Naman Goel, Alfonso Amayuelas, Amit Deshpande, Amit Sharma

Motivation

- In most fair machine learning settings, the training data has some form of missingness.



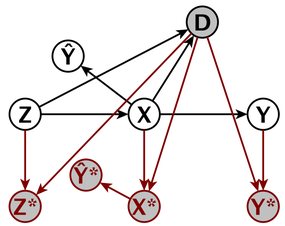
- The above figure shows a very common type of missingness in the training data: only those training instances, for which decisions in the past were positive, appear in the training data.
- Many benchmark training datasets used in the fair machine learning literature (e.g. the German Credit dataset) have this kind of missingness.
- Most of the state-of-the-art fair machine learning algorithms do not take data missingness into account. Due to this, a supposedly fair classifier can be arbitrarily unfair in the real world.

Contributions

- Using a causal graph based framework (based on *Mohan and Pearl, 2020*), we formally discuss how past decisions affect data missingness.
- With motivating examples for different kinds of past decision-making, we show which parts of the joint distribution can be recovered from the incomplete training data and which can not be recovered.
- Interestingly, in many scenarios of missingness, the distributions used in common fairness algorithms are not recoverable.
- We show how the above results can guide the design of fair algorithms in practice by proposing a detail-free, decentralized and fair algorithm for multi-stage setting.
- Our theoretical and empirical analysis shows that the algorithm provides same utility as an oracle algorithm which assumes full centralization and knowledge of non-recoverable distributions

Implications of Data Missingness for Fair ML

D - Past Decision, X - Non-Sensitive Features, Z - Sensitive Feature
Y – Outcome, $\hat{Y}$ - Classifier's prediction, V^* - Observed V



$$P(\hat{Y}^*|Y^*, Z^*) = P(\hat{Y}|Y, Z, D = 1)$$
$$\neq P(\hat{Y}|Y, Z) \quad \text{because} \quad \hat{Y} \not\perp D|Y, Z$$

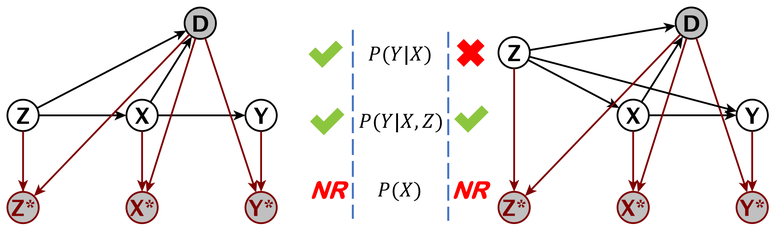
$\Rightarrow$ Equality of opportunity constraints, estimated naively from training data, are inconsistent.

Also true for demographic parity constraints.

$P(Y|X,Z)$ $P(Y|X)$ and $P(X)$ are other distributions assumed to be known in many fair ML algorithms. Using causal graphs, we reason about their identifiability.

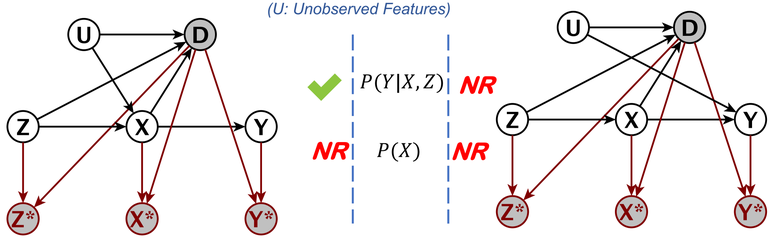
✔ Can be estimated consistently | ***NR*** Non-recoverable | ✖ Naïve estimate not consistent.

➢ **Case 1: When fully automated decisions cause missingness.**

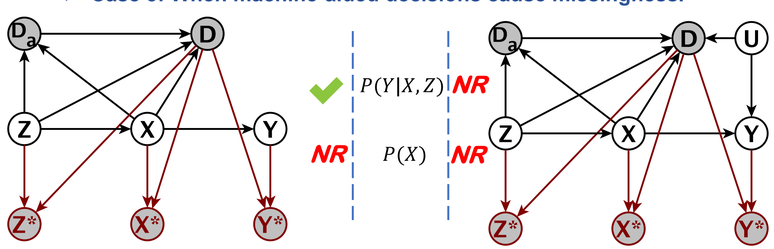


➢ **Case 2: When human decisions cause missingness.**

(U: Unobserved Features)

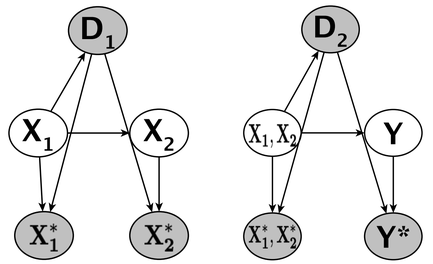


➢ **Case 3: When machine-aided decisions cause missingness.**



Application: Multi-Stage Decision Making

Multi-stage decision-making processes are common, e.g., in hiring, university admissions and lending. At each stage of the selection process, decision makers request or collect new features about the individuals and make decisions on whether to forward an individual to the next stage or not. Each stage of the selection process narrows down (subject to budget constraints) the pool of individuals and more features are observed in the subsequent stages for individuals who pass the previous stage.



Causal Graphs for Data Missingness in 2-Stage Decision Making Process

Joint distribution $P(X_1, X_2)$ is non-recoverable.

On the other hand, $P(Y|X_1, X_2)$ can be consistently estimated.

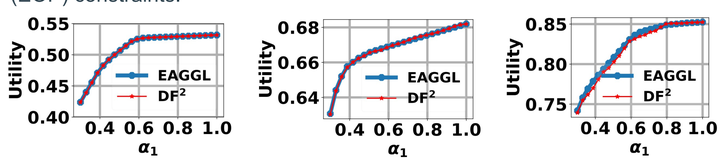
$P(Y|X_1)$ can be recovered by factorization.

The DF^2 (Detail-Free, Decentralized and Fair) Algorithm for Multi-State Decision-Making

The DF^2 algorithm solves the following optimization problem at every stage $i \in \{1, 2, \ldots k\}$:

$$\max_{D_i} \quad P(Y = 1|\hat{Y}_i = 1)$$
$$\text{s.t.} \quad P(\hat{Y}_i = 1) = \alpha_i$$
$$f_i(\hat{Y}_i) = 0$$

where $P(Y = 1 \mid \hat{Y}_i = 1)$ is the precision of the decisions taken at stage i, $P(\hat{Y}_i = 1) = \alpha_i$ is the budget constraint at stage i, and $f_i(\hat{Y}_i) = 0$ is the fairness constraint at stage i. In the paper, we show how to write the objective and the constraints using only the recoverable distributions $P(Y|X_1, X_2)$ and $P(Y|X_1)$. The performance of the DF^2 algorithm is guaranteed relative to an oracle algorithm (EAGGL), if the features used in different stages provide *coherent* information about the outcome Y. The following figure compares the utility of the two algorithms under fairness (EOP) constraints.



Adult Dataset | COMPAS Dataset | German Dataset