DECOMPOSING CAUSALITY AND FAIRNESS

Anonymous authors
Paper under double-blind review

ABSTRACT

It is often informative to decompose key quantities of interest into smaller components, in order to develop a better understanding of the key quantity. In this paper, we focus causality and fairness, where bias attribution can be particularly useful. We show how quantities can be broken down based on independence, or conditional independence criteria, and show how such a decomposition can be used as a diagnosis tool.

1 INTRODUCTION

Both *causality* (Pearl, 2009; Peters et al., 2017; Neal, 2015; Schölkopf, 2022) and *fairness* (Chouldechova, 2017; Dwork et al., 2012; Donini et al., 2018) are large research topics in themselves, and are closely connected[1]. In both cases, there are key quantities that are of interest to compute. We show how we can decompose key quantities from each area into similar structures. This gives us a simpler way to calculate the quantity, and allows us to understand the how close a model is to respecting certain requirements, as well as diagnosing why the model isn't respecting these requirements.

2 CAUSALITY DECOMPOSITION

Closed-form adjustments We want to estimate the *causal impact* of a variable X on an outcome Y, in the presence of confounders S. The quantity $\mathbb{P}(Y|do(X = x)) - \mathbb{P}(Y|X = x)$ tells us how close the interventional distribution and the conditional distribution are. We obtain two decompositions for this quantity based on the *backdoor* and *frontdoor* criteria (see section A.1 of the Appendix), allowing us to consider a causal relationship using purely statistical tools.

Proposition 1. *a) Under the same conditions as the* ***backdoor criteria****, we have*

$$\mathbb{P}(Y|do(X=x)) - \mathbb{P}(Y|X=x) = \textstyle\sum_{x' \in \mathcal{X}} \mathbb{P}(Y|X=x, S=s)\mathbb{P}(S=s)\left[\underbrace{1 - \frac{\mathbb{P}(X=x, S=s)}{\mathbb{P}(X=x)\mathbb{P}(S=s)}}_{\chi_{x,s}}\right] \tag{1}$$

b) Under the same conditions as the ***frontdoor criterion****, we have*

$$\mathbb{P}(Y|do(X=x)) - \mathbb{P}(Y|X=x) = \sum_s \mathbb{P}(S=s|X=x)\left(\sum_{x'}[\mathbb{P}(Y|X=x', S=s) - \mathbb{P}(Y|X=x, S=s)]\mathbb{P}(X=x')\right). \tag{2}$$

Remark 1. Proposition 1 identifies closed form expressions for the difference between the interventional distribution of X on Y and the conditional distribution based on the underlying DAG. This allows us to decompose this quantity into sub components. There is a connection between the backdoor criterion and independence of X and S, through $\chi_{x,s}$: the closer X and S to independence, the closer $\chi_{x,s}$ is to 0. Thus the closer $\mathbb{P}(Y|do(X = x))$ and $\mathbb{P}(Y|X = x)$ are. Similarly, the frontdoor criterion is directly connected to the conditional independence of Y and X, given S.

Control Variates These decompositions naturally lead to the idea of control variates (Glasserman & Xu, 2014), where we can make use of an accurate existing estimate, or known value, of $\mathbb{P}(Y|X = x)$, to help model $\mathbb{P}(Y|do(X = x))$, by instead modelling the error from the decomposition.

[1]Further details, proofs, and experiments on causality and fairness are provided in the Appendix.

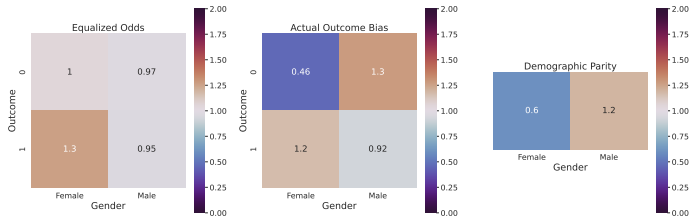


Figure 1: A heatmap showing a decomposition of the EOd-DP formula (Proposition 2) based on the Adult Income dataset. We compare fr_{DP} with fr_{EOd} and fr_{act} for $\hat{y} = 0$. One observes that are close to satisfying the equalised odds criterion (as these values are close to 0), but do not satisfy demographic parity.

3 Fairness Decomposition

We consider $S \in \mathcal{S}$ to be a binary protected attribute, $X \in \mathcal{X}$ to be a set of features (excluding S), $Y \in \mathcal{Y}$ to be a binary outcome variable, and $\hat{Y} \in \mathcal{Y}$ to be a predictor for Y. Narayanan (2018); Verma & Rubin (2018); Berk et al. (2018); Kim et al. (2020) have reviewed many of the different definitions of fairness that exist. These different definitions cannot all co-exist, as per the impossibility theorems (Kleinberg et al., 2017; Chouldechova, 2017).

Fairness metric	Equality requirement	Reference
Demographic parity	$\mathbb{P}(\hat{Y} = \hat{y} \mid S = s) = \mathbb{P}(\hat{Y} = \hat{y})$, for all $s \in \mathcal{S}, \hat{y} \in \mathcal{Y}$	(Calders & Verwer, 2010)
Equalized odds	$\mathbb{P}(\hat{Y} = \hat{y} \mid, Y = y, S = s) = \mathbb{P}(\hat{Y} = \hat{y} \mid Y = y)$, for all $s \in \mathcal{S}, \hat{y}, y \in \mathcal{Y}$	(Hardt et al., 2016)

Table 1: Fairness metrics used in this article

Fairness ratios We introduce "fairness ratios" as measures of deviation of a model from the "fair" situation in terms of ratios, according to a specified fairness metric, Table 1. For example, for Demographic parity, we consider $\text{fr}_{DP}(\widehat{y}, s) = \frac{\mathbb{P}(\widehat{y}|s)}{\mathbb{P}(\widehat{y})}$, whilst for Equalised odds, we consider $\text{fr}_{EOd}(\widehat{y}, s, y) = \frac{\mathbb{P}(\widehat{y}|s,y)}{\mathbb{P}(\widehat{y}|y)}$. We can also consider the actual outcome bias, $\text{fr}_{act}(s, y) = \frac{\mathbb{P}(s|y)}{\mathbb{P}(s)}$ for all $y \in \mathcal{Y}, s \in \mathcal{S}$. One of the interesting properties of fairness ratios is that they can be decomposed into finer fairness ratios and help diagnose where bias in a model may be coming from.

Proposition 2. ***EOd-DP Formula*** *Let $y, \widehat{y} \in \mathcal{Y}$ and $s \in \mathcal{S}$, then the following equality holds:*

$$fr_{DP}(\widehat{y}, s) = \sum_{y \in \mathcal{Y}} fr_{EOd}(\widehat{y}, s, y) \cdot fr_{act}(s, y) \cdot \mathbb{P}(y|\widehat{y}). \tag{3}$$

Thus, we can decompose demographic parity in terms of equalized odds and actual bias.

Experimentation We consider the Adult Income dataset (Dua & Graff, 2017), and have demonstrated the usefulness of this decomposition. We consider the protected attribute, S, of gender, and use a simple logistic regression model, without using S, to make predictions $\hat{y}$ on if an individual has an income of $> 50K$. We focus on the EOd-DP formula, Proposition 2, and are able to diagnose where DP unfairness is coming from (see Figure 1). To obtain a fair model with respect to DP, one must either tweak the model to be fair w.r.t EO, a form of inprocessing, or reweigh the underlying distribution, a form of pre-processing, to avoid outcome disparity.

4 Conclusion

In this paper, we create decompositions of key interest quantities in causality and fairness and provide closed form expressions for these decompositions. There are clear similarities that can be drawn between format of the decompositions and what they allow us to diagnose the source of the impact in each case, providing us with control variate approaches.

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2023 Tiny Papers Track.

REFERENCES

Richard Berk, Hoda Heidari, Shahin Jabbari, Michael Kearns, and Aaron Roth. Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, August 2018. doi: 10.1177/0049124118782533.

Toon Calders and Sicco Verwer. Three naive bayes approaches for discrimination-free classification. *Data Mining and Knowledge Discovery*, 21(2):277–292, 2010. doi: 10.1007/s10618-010-0190-x.

Alexandra Chouldechova. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, 5(2):153–163, June 2017. doi: 10.1089/big.2016.0047.

Michele Donini, Luca Oneto, Shai Ben-David, John S Shawe-Taylor, and Massimiliano Pontil. Empirical risk minimization under fairness constraints. *Advances in neural information processing systems*, 31, 2018.

Dheeru Dua and Casey Graff. UCI machine learning repository, 2017. URL `http://archive.ics.uci.edu/ml`.

Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*, pp. 214–226, 2012.

Paul Glasserman and Xingbo Xu. Robust risk measurement and model risk. *Quantitative Finance*, 14(1):29–58, 2014.

Moritz Hardt, Eric Price, and Nathan Srebro. Equality of opportunity in supervised learning. *Advances in Neural Information Processing Systems*, 29:3323–3331, December 2016. doi: 10.5555/3157382.3157469. URL `https://papers.nips.cc/paper/6374-equality-of-opportunity-in-supervised-learning`.

Joon Sik Kim, Jiahao Chen, and Ameet Talwalkar. Model-agnostic characterization of fairness trade-offs. In *Proceedings of the International Conference on Machine Learning*, volume 37, pp. 9339–9349, 2020.

Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. Inherent trade-offs in the fair determination of risk scores. In Christos H. Papadimitriou (ed.), *Proceedings of the 8th Innovations in Theoretical Computer Science Conference*, volume 67 of *Leibniz International Proceedings in Informatics (LIPIcs)*, Dagstuhl, Germany, 2017. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi: 10.4230/LIPIcs.ITCS.2017.43.

Arvind Narayanan. Translation tutorial: 21 fairness definitions and their politics. In *Proceedings of the Conference on Fairness, Accountability and Transparency*, FAT* 18, New York, USA, 2018.

Brady Neal. Introduction to causal inference. 2015.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. The MIT Press, 2017.

Bernhard Schölkopf. Causality for machine learning. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, pp. 765–804. 2022.

Sahil Verma and Julia Rubin. Fairness definitions explained. In *Proceedings of the International Conference on Software Engineering*, pp. 1–7, New York, NY, USA, 2018. ACM. doi: 10.1145/3194770.3194776.

A APPENDIX

In this Appendix, we consider the following:

- Additional details on causality (Section A.1
- Additional details on fairness (Section A.2)

A.1 ADDITIONAL DETAILS ON CAUSALITY

In this section, we provide additional details, proofs and experimental results in relation to causality.

A.1.1 INTRODUCTORY DETAILS

We typically want to estimate the *causal impact* of a variable X on an outcome Y. For example, this could be the causal impact of a treatment on a patient, or a marketing campaign on a KPI. Thorough introductions to the topic of causality can be found in Pearl (2009); Peters et al. (2017); Neal (2015); Schölkopf (2022).

We are able to condition of variable X, and deduce the conditional probability $\mathbb{P}(Y|X = x)$. However, this is not equivalent to understanding the causal effect of X on Y. Association is not the same as causation. This is due to the fact that there could be confounding variables, S, which could be common causes of both X and Y.

For example, in the case of a binary outcome Y, and binary treatment X, we are really interested in understanding the difference between the *potential outcome* of Y under X being 0 or 1. However, since only one of these outcomes can occur (either the treatment can be given or not), we cannot observe both outcomes. This is the Fundamental Problem of Causal Inference.

In order to remove the effect of the confounders, we must instead intervene on X, which we write as $do(X = x)$. Intervening allows us set the value of X, thus removing the affect of any confounding variables on X. We are then interested in $\mathbb{P}(Y|do(X = x))$, which is a *causal estimand*. For example, in the case of a binary variable $X \in 0, 1$ (such as being given a treatment or not), in order to establish the causal effect of X on Y, we must consider $\mathbb{E}[Y|do(X = 1)] - \mathbb{E}[Y|do(X = 0)]$.

In order to understand the causal impact, we must first introduce the idea of a Directed Acyclic Graph (DAG), which assumes the causal structure of the problem. For example, we could consider the causal structures outlined in Figures 2 or 3a. Key structures of causal graphs are *chains*, *forks* and *colliders*, which are outlined in Figure 2. In these cases, we refer to S as the middle node.

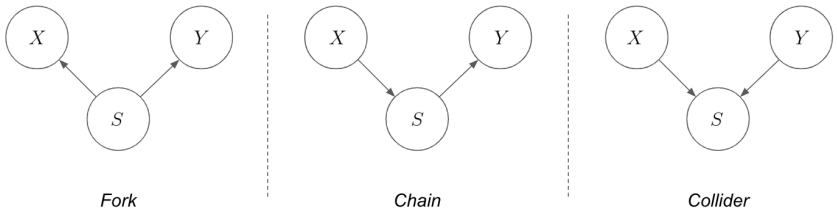


Figure 2: Forks, Chains and Colliders

Key results in causal inference theory allow us to relate these interventional distributions, which we are most interested in, to conditional distributions. This allows us to consider a causal relationship using purely statistical tools. We consider these results, in particular the backdoor and frontdoor criteria discussed in the following.

Definition 1. A *backdoor path* from X to Y is a non-causal path from X to Y.

Definition 2. A path between X and Y is blocked by a node S if either

- There is a fork, or a chain, on the path, with S as the middle node, where S is conditioned on.

- There is a collider on the path, with S as the middle node, where neither S, nor any descendants of S are conditioned on.

Theorem 1 (Backdoor Criterion)**.** *Suppose that a set S blocks all backdoor paths (non-causal paths) from X to Y, and does not contain any descendants of X, then*

$$\mathbb{P}(Y|do(X=x)) = \sum_{s\in\mathcal{S}} \mathbb{P}(Y|X=x, S=s)P(S=s) \tag{4}$$

Theorem 2 (Frontdoor Criterion)**.** *Suppose that a set S completely mediates the effect of X on Y, that there is no unblocked backdoor path from X to S, and that all backdoor paths from S to Y are blocked by X, then*

$$\mathbb{P}(Y|do(X=x)) = \sum_{s} \mathbb{P}(S=s|X=x) \sum_{x'} \mathbb{P}(Y|X=x', S=s)P(X=x'). \tag{5}$$

A.1.2 Proofs

Proposition 1. *a) Under the same conditions as the **backdoor criteria**, we have*

$$\mathbb{P}(Y|do(X=x)) - \mathbb{P}(Y|X=x) = \textstyle\sum_{x'\in\mathcal{X}} \mathbb{P}(Y|X=x, S=s)\mathbb{P}(S=s)\left[\underbrace{1 - \frac{\mathbb{P}(X=x, S=s)}{\mathbb{P}(X=x)\mathbb{P}(S=s)}}_{\chi_{x,s}}\right] \tag{1}$$

*b) Under the same conditions as the **frontdoor criterion**, we have*

$$\begin{aligned}&\mathbb{P}(Y|do(X=x)) - \mathbb{P}(Y|X=x) = \\ &\quad\sum_{s} \mathbb{P}(S=s|X=x)\left(\sum_{x'}[\mathbb{P}(Y|X=x', S=s) - \mathbb{P}(Y|X=x, S=s)]\mathbb{P}(X=x')\right).\end{aligned} \tag{2}$$

Proof of a). Firstly, we recall that, using the backdoor criteria, we have

$$\mathbb{P}(Y|do(X=x)) = \sum_{s\in\mathcal{S}} \mathbb{P}(Y|X=x, S=s)P(S=s). \tag{6}$$

Also, we note that

$$\mathbb{P}(Y|X=x) = \sum_{s\in\mathcal{S}} \mathbb{P}(Y|X=x, S=s)\mathbb{P}(S=s|X=x). \tag{7}$$

Combining these together, we get that

$$\begin{aligned}&\mathbb{P}(Y|do(X=x)) - \mathbb{P}(Y|X=x) \\ &\quad= \sum_{s}\Big(\mathbb{P}(Y|X=x, S=s)\mathbb{P}(S=s) - \mathbb{P}(Y|X=x, S=s)\mathbb{P}(S=s|X=x)\Big) \\ &\quad= \sum_{s} \mathbb{P}(Y|X=x, S=s)\Big(\mathbb{P}(S=s) - \mathbb{P}(S=s|X=x)\Big) \\ &\quad= \sum_{s} \mathbb{P}(Y|X=x, S=s)\mathbb{P}(S=s)\chi_{x,s}\end{aligned} \tag{8}$$

where the final line comes from the fact that

$$\mathbb{P}(S=s|X=x) = \frac{\mathbb{P}(S=s, X=x)}{\mathbb{P}(X=x)} \tag{9}$$

□

Proof of b). From the frontdoor criterion, we note that

$$\mathbb{P}(Y|do(X=x)) = \sum_{s} \mathbb{P}(S=s|X=x) \sum_{x'} \mathbb{P}(Y|X=x', S=s)P(X=x'). \tag{10}$$

Also, we note that

$$\mathbb{P}(Y|X=x) = \sum_s \mathbb{P}(Y|X=x, S=s)\mathbb{P}(S=s|X=x). \tag{11}$$

Combining these together we see that

$$\begin{aligned}
&\mathbb{P}(Y|do(X=x)) - \mathbb{P}(Y|X=x) \\
&= \sum_s \left(\mathbb{P}(S=s|X=x)\left(\sum_{x'} \mathbb{P}(Y|X=x', S=s)P(X=x')\right) - \mathbb{P}(Y|X=x, S=s)\mathbb{P}(S=s|X=x) \right) \\
&= \sum_s \mathbb{P}(S=s|X=x)\left(\left(\sum_{x'} \mathbb{P}(Y|X=x', S=s)P(X=x')\right) - \mathbb{P}(Y|X=x, S=s) \right)
\end{aligned} \tag{12}$$

Further note that

$$\mathbb{P}(Y|X=x, S=s) = \mathbb{P}(Y|X=x, S=s)\sum_{x'} \mathbb{P}(X=x') \tag{13}$$

$$= \sum_{x'} \mathbb{P}(Y|X=x, S=s)\mathbb{P}(X=x') \tag{14}$$

Combining the above, we see that

$$\begin{aligned}
&\mathbb{P}(Y|do(X=x)) - \mathbb{P}(Y|X=x) = \\
&\sum_s \mathbb{P}(S=s|X=x)\left(\sum_{x'} \mathbb{P}(Y|X=x', S=s)P(X=x') - \sum_{x'} \mathbb{P}(Y|X=x, S=s)\mathbb{P}(X=x') \right) \\
&= \sum_s \mathbb{P}(S=s|X=x)\left(\sum_{x'} [\mathbb{P}(Y|X=x', S=s) - \mathbb{P}(Y|X=x, S=s)]\mathbb{P}(X=x') \right)
\end{aligned} \tag{15}$$

□

A.1.3 Control Variates

If we wish to model $\mathbb{P}(Y|do(X=x))$, these decompositions naturally lead to the idea of control variates. The aim is to make use of existing solutions to similar problems, to solve a more difficult problem. Supposing that the conditional probability $\mathbb{P}(Y|X=x)$ was known, a control variates approach would be to model the differences from Proposition 1, and add it on to $\mathbb{P}(Y|X=x)$, rather than modelling $\mathbb{P}(Y|do(X=x))$ directly.

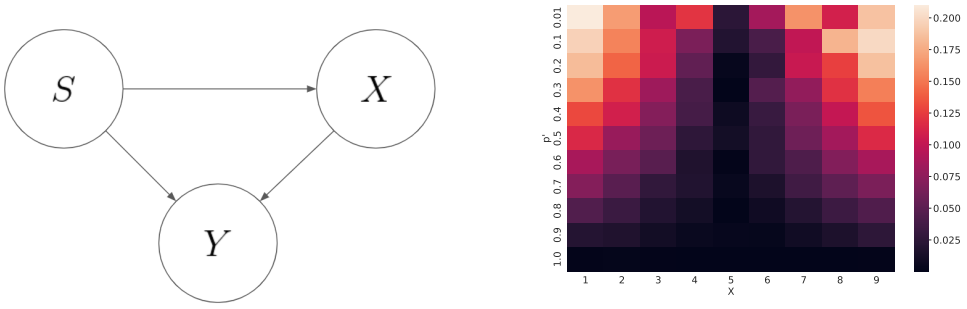
A.1.4 Experimentation

We run experiments in relation to the causality decompositions. We consider a synthetic dataset under the DAG represented in Figure 3a. Here, we are interested in understanding the causal relationship of X on Y. However, there is also a confounder, S. In order to generate the data, we randomly sample $s_i \in [1, \cdots, 10]$, and latent $u_i \in [1, \cdots, 10]$ for $i = 1, \cdots, 100000$. Then we consider $x_i = p'u_i + (1-p')s_i$, for some $p' \in (0,1)$, which controls the covariance between S and X. We model $Y_i \sim Bernoulli(p_i)$, where $p_i = \frac{x_i + s_i}{\max_i x_i + \max_i s_i}$. Note that S satsifies the backdoor criterion conditions, so we can use Proposition 1 a).

Figure 3b shows the value of the quantity $\mathbb{P}(Y|do(X=x)) - \mathbb{P}(Y|X=x)$, for each x and different values of p'. Varying p' from between 0 and 1 allows us to vary the covariance between X and S. For smaller p, we have higher covariance, which means that the quantity $\chi_{x,s}$ is larger, and thus the quantity $\mathbb{P}(Y|do(X=x)) - \mathbb{P}(Y|X=x)$ is also larger.

A.2 Additional Details on Fairness

In this section, we provide additional details, proofs and experimental results in relation to fairness.



(a) (b)

Figure 3: a) DAG used in the synthetic dataset. b) Absolute value of quantity from Proposition 1 a) for synthetic data, for different values of p'. A darker colour gives a value closer to 0, thus indicating less dependence between X and S. For each x as we decrease p' from 1 to 0, we introduce more dependency between X and S, and thus our quantity increases in absolute value, and the conditional distribution is further away from the interventional distribution.

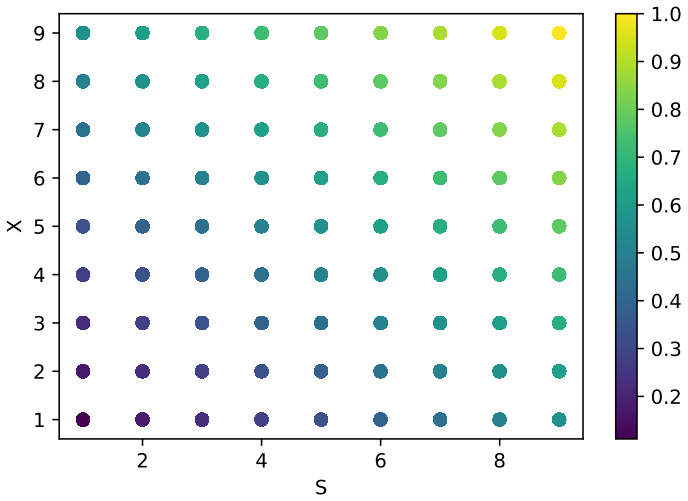


Figure 4: An example of the synthetic data used, with $p' = 0.5$. The colour indicates the value of p.

A.2.1 Proofs

Proposition 2. ***EOd-DP Formula*** *Let $y, \widehat{y} \in \mathcal{Y}$ and $s \in \mathcal{S}$, then the following equality holds:*

$$fr_{DP}(\widehat{y}, s) = \sum_{y \in \mathcal{Y}} fr_{EOd}(\widehat{y}, s, y) \cdot fr_{act}(s, y) \cdot \mathbb{P}(y|\widehat{y}). \tag{3}$$

Thus, we can decompose demographic parity in terms of equalized odds and actual bias.

Proof. Firstly, we note that this is equivalent to proving the following:

$$\frac{\mathbb{P}(\widehat{y}|s)}{\mathbb{P}(\widehat{y})} = \sum_{y \in \mathcal{Y}} \frac{\mathbb{P}(\widehat{y}|s, y)}{\mathbb{P}(\widehat{y}|y)} \cdot \frac{\mathbb{P}(s|y)}{\mathbb{P}(s)} \cdot \mathbb{P}(y|\widehat{y}). \tag{16}$$

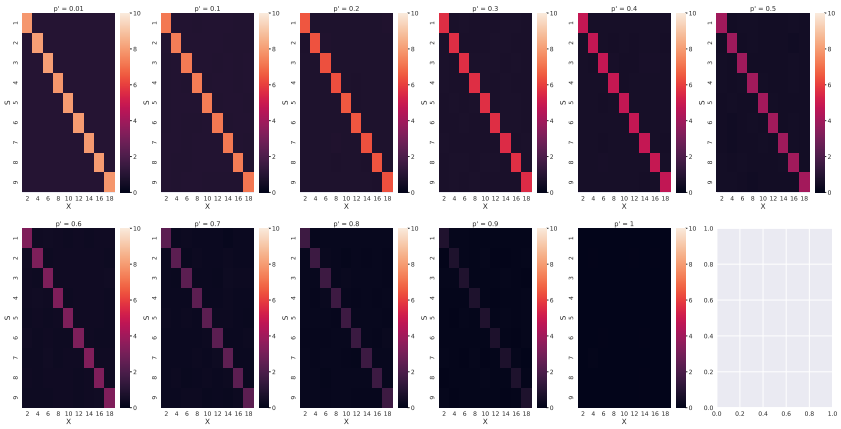


Figure 5: The values of $\chi_{x,s}$ for different values of p' using the synthetic data. We see that the closer p' is to 1, the closer to independence we are, and thus the closer $\chi_{x,s}$ is to 0.

We note that

$$\frac{\mathbb{P}(\widehat{y}|s)}{\mathbb{P}(\widehat{y})} = \sum_{y\in\mathcal{Y}} \frac{\mathbb{P}(\widehat{y},y|s)}{\mathbb{P}(\widehat{y})} \tag{17}$$

$$= \sum_{y\in\mathcal{Y}} \frac{\mathbb{P}(\widehat{y},|y,s)\mathbb{P}(y|s)}{\mathbb{P}(\widehat{y})} \tag{18}$$

$$= \sum_{y\in\mathcal{Y}} \frac{\mathbb{P}(\widehat{y},|y,s)\mathbb{P}(s|y)\mathbb{P}(y)}{\mathbb{P}(\widehat{y})\mathbb{P}(s)}, \tag{19}$$

where the second line comes from the definition of conditional probability, and the third line comes from Bayes Theorem. Also, note that

$$\mathbb{P}(\hat{y}|y) = \frac{\mathbb{P}(y,\hat{y})}{\mathbb{P}(y)} \tag{20}$$

and

$$\mathbb{P}(y|\hat{y}) = \frac{\mathbb{P}(y,\hat{y})}{\mathbb{P}(\hat{y})}. \tag{21}$$

Combining, we obtain,

$$\frac{\mathbb{P}(y)}{\mathbb{P}(\hat{y})} = \frac{\mathbb{P}(y|\hat{y})}{\mathbb{P}(\hat{y}|y)}. \tag{22}$$

Replacing this ratio in Equation 19, we obtain the result. □

Using this results, we can decompose demographic parity in terms of equalized odds and actual bias. In particular, if the assumption of equalized odds is verified, then $\text{fr}_{EOd}(\widehat{y},s,y) = 1$ for all $y \in \mathcal{Y}$, then $\text{fr}_{DP}(\widehat{y},s) = \sum_{y\in\mathcal{Y}} \text{fr}_{act}(s,y) \cdot \text{model probability}(y,\widehat{y})$. If, in addition, $\text{fr}_{act}(s,y) = 1$, then $\text{fr}_{DP}(\widehat{y},s) = 1$, but if $\text{fr}_{act}(s,y) \neq 1$, we recover the well-known result that demographic parity does not hold in general (see (Kleinberg et al., 2017)).

A.2.2 Experimentation

Figure 6 demonstrates the decomposition of the EOd-DP Formula for the Adult Income dataset, with the set-up outlined in the main text. The top row corresponds to $\hat{y} = 0$, whilst the bottom row corresponds to $\hat{y} = 1$.

Figure 7 demonstrates the decomposition of the EOd-DP Formula (Propopsition 2) when we fix $\text{fr}_{EOd} = 1$. It shows us that even if we have a completely fair model with respect to Equalised odds, this does not guarantee that our model will be fair under Demographic parity.

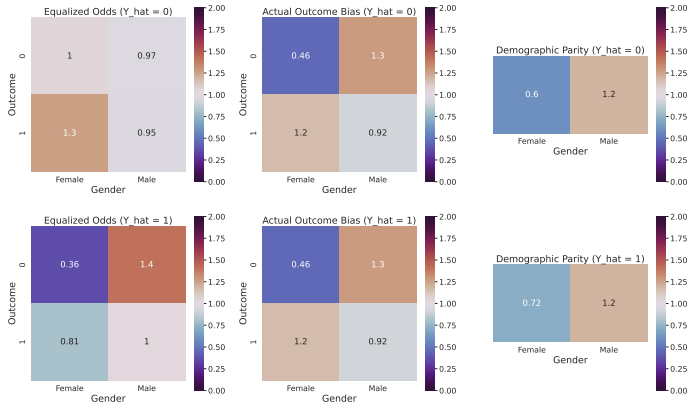


Figure 6: A heatmap showing a decomposition of the EOd-DP formula (Proposition 2) based on the Adult Income dataset. Comparing fr_{DP} with fr_{EOd} and fr_{act}. We see that for $\hat{y} = 0$ (the top row), we are close to satisfying the equalised odds criterion, but do not satisfy demographic parity, whilst in the case $\hat{y} = 1$ (the bottom row), we do not generally satisfy either condition.

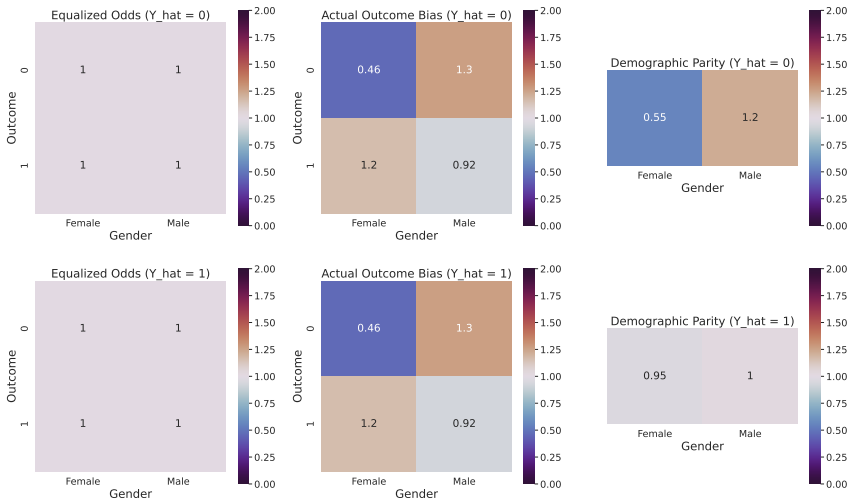


Figure 7: A heatmap showing a decomposition of the EOd-DP formula (Proposition 2) based on the Adult Income dataset, when we fix $\text{fr}_{EOd} = 1$, thus assuming a fair model with respect to Equalised Odds. We see that we aren't guaranteed to obtain fairness with respect to Demographic Parity.