DDA4210 Spring 2024 - Assignment 3

Yohandi

April 30, 2024

1.

Suppose we have the following structural causal model. Assume all exogenous variables are independent and the expected value of each is 0.

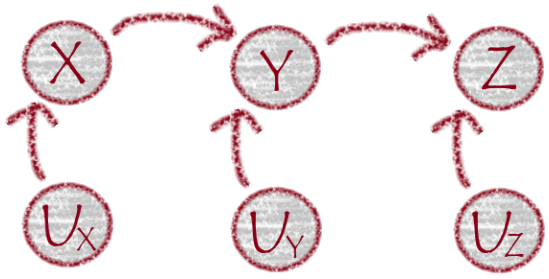
$$V = \{X, Y, Z\}, \quad U = \{U_X, U_Y, U_Z\}, \quad F = \{f_X, f_Y, f_Z\}$$

$$f_X : X = U_X$$
$$f_Y : Y = \frac{X}{3} + U_Y$$
$$f_Z : Z = \frac{Y}{16} + U_Z$$

1. Draw the graph that complies with the model.
2. Determine the best guess of the value (expected value) of Z, given that we observe $Y = 3$.
3. Determine the best guess of the value of Z, given that we observe $X = 3$.
4. Determine the best guess of the value of Z, given that we observe $X = 1$ and $Y = 3$.

1.
   - X is directly influenced by an exogenous variable U_X
   - Y is directly influenced by X and an exogenous variable U_Y
   - Z is directly influenced by Y and an exogenous variable U_Z

   Hence, the graph that complies with the model is drawn as follows:



2. As $Z = \frac{Y}{16} + \underbrace{U_Z}_{\text{independent}}$, the best guess of the value of Z is $\mathbb{E}(Z|Y=3) = \frac{3}{16} + \underbrace{\mathbb{E}(U_Z)}_{0} = \frac{3}{16}$.
3. As $Y = \frac{X}{3} + \underbrace{U_Y}_{\text{independent}}$, $\mathbb{E}(Y|X=3) = \frac{3}{3} + \underbrace{\mathbb{E}(U_Y)}_{0} = 1$. As our expected value of Y would be 1, the best guess of the value of Z is $\mathbb{E}(Z|Y=1) = \frac{1}{16} + \underbrace{\mathbb{E}(U_Z)}_{0} = \frac{1}{16}$.
4. In this scenario, X does not affect the estimation as Y is already known. As computed in 2., the best guess of the value of Z is $\mathbb{E}(Z|Y=3) = \frac{3}{16}$.

2.

In the central model of differential privacy, a trusted curator aggregates all data and randomizes responses to queries. However, in the local model of differential privacy, users do not trust the aggregator, so they randomize their data locally before sending it to the aggregator. Consider the

Randomized Response (RR) mechanism, proposed by Warner in 1965, aiming to collect sensitive statistics while providing each participant some deniability.

Each of the n users holds a private bit $b_i \in \{0, 1\}$, and we want to estimate the average $a := \frac{1}{n}\sum_{i=1}^{n} b_i$. The RR mechanism, executed independently by each user, consists of flipping two unbiased coins and following these steps:

- If the first coin is heads, send b_i to the aggregator. Otherwise, look at the second coin:
- If heads, send 0 to the aggregator.
- If tails, send 1 to the aggregator.

1. Demonstrate that the RR mechanism ensures ϵ-differential privacy for each user's individual bit, with $\epsilon = \ln(3)$.
2. Let $\hat{b}_i$ be the i-th user's randomized response. Prove that the untrusted aggregator that receives all these noisy bits can compute an unbiased estimate $\hat{a}$ of a (i.e., $\mathbb{E}[\hat{a}] = a$).
3. Show that the estimation error $\hat{a} - a$ has a standard deviation of $O\left(\frac{1}{\sqrt{n}}\right)$.
4. Compare this result with the central model, where all users send their bits to a trusted curator using the Laplace mechanism to output a noisy estimate $\hat{a}$ of a that is $\ln(3)$-differentially private. Show that the estimation error $\hat{a} - a$ has a standard deviation of $O\left(\frac{1}{n}\right)$.
5. Design a generalized RR mechanism that provides ϵ-differential privacy for each user's individual bit, for any fixed $\epsilon > 0$. Show that your mechanism satisfies ϵ-DP in the local model, and that the standard deviation of the untrusted aggregator's estimation error is $O\left(\frac{1}{\epsilon\sqrt{n}}\right)$.

1. Let $\mathcal{M}$ be the corresponding mechanism. Consider the private bit b_i cases:
   - Case $b_i = 0$:
     - $\Pr(\mathcal{M}(b_i) = 0) = \frac{3}{4}$
     - $\Pr(\mathcal{M}(b_i) = 1) = \frac{1}{4}$
   - Case $b_i = 1$:
     - $\Pr(\mathcal{M}(b_i) = 0) = \frac{1}{4}$
     - $\Pr(\mathcal{M}(b_i) = 1) = \frac{3}{4}$

   For neighboring datasets $\mathcal{D}_1 = \mathcal{D} \cup \{b_i = 0\}$ and $\mathcal{D}_2 = \mathcal{D} \cup \{b_i = 1\}$,

   $$\frac{\Pr(\mathcal{M}(\mathcal{D}_1) = 0)}{\Pr(\mathcal{M}(\mathcal{D}_2) = 0)} \leq \frac{\Pr(\mathcal{M}(b_i = 0) = 0)}{\Pr(\mathcal{M}(b_i = 1) = 0)} = \frac{\frac{3}{4}}{\frac{1}{4}} = 3 \leq e^{\epsilon}$$

   and

   $$\frac{\Pr(\mathcal{M}(\mathcal{D}_1) = 1)}{\Pr(\mathcal{M}(\mathcal{D}_2) = 1)} \geq \frac{\Pr(\mathcal{M}(b_i = 0) = 1)}{\Pr(\mathcal{M}(b_i = 1) = 1)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3} \geq e^{-\epsilon}$$

   Thus, the RR mechanism ensures ϵ-differential privacy with the according attainment $\epsilon^* = \ln(3)$.

2. The following proves that the untrusted aggregator that receives all these noisy bits can compute an unbiased estimate $\hat{a}$ of a.

   $$\mathbb{E}[\hat{b}_i] = \mathrm{P}(b_i = 0)\mathbb{E}[\hat{b}_i | b_i = 0] + \mathrm{P}(b_i = 1)\mathbb{E}[\hat{b}_i | b_i = 1]$$

$$
\begin{aligned}
&= (1-a)\cdot(0\cdot\frac{3}{4}+1\cdot\frac{1}{4})+a\cdot(0\cdot\frac{1}{4}+1\cdot\frac{3}{4}) \\
&= (1-a)\cdot\frac{1}{4}+a\cdot\frac{3}{4} \\
&= \frac{1}{4}+\frac{1}{2}a \\
\Rightarrow a &= 2\mathbb{E}[\hat{b}_i]-\frac{1}{2} \\
\Rightarrow \hat{a} &= \frac{2}{n}\sum_{i=1}^{n}\hat{b}_i-\frac{1}{2} \\
\Rightarrow \mathbb{E}[\hat{a}] &= \mathbb{E}[\frac{2}{n}\sum_{i=1}^{n}\hat{b}_i-\frac{1}{2}] \\
&= 2\mathbb{E}[\frac{1}{n}\sum_{i=1}^{n}\hat{b}_i]-\frac{1}{2} \\
&= 2\mathbb{E}[\hat{b}_i]-\frac{1}{2} \\
&= a
\end{aligned}
$$

3. The variance of our estimator is given by:

$$
\begin{aligned}
\mathrm{Var}[\hat{a}] &= \mathrm{Var}[\frac{2}{n}\sum_{i=1}^{n}\hat{b}_i-\frac{1}{2}] \\
&= \mathrm{Var}[\frac{2}{n}\sum_{i=1}^{n}\hat{b}_i] \\
&= \frac{4}{n^2}\mathrm{Var}[\sum_{i=1}^{n}\hat{b}_i] \\
&= \frac{4}{n}\mathrm{Var}[\hat{b}_i]
\end{aligned}
$$

   In our case, the estimator is unbiased; consequently, $\mathbb{E}[\hat{a}-a]=\mathbb{E}[\hat{a}]-a=0$, which further implies that:

$$
\begin{aligned}
\sigma(\hat{a}-a) &= \sigma(\hat{a}) \\
&= \sqrt{\mathrm{Var}[\hat{a}]} \\
&= \frac{2}{\sqrt{n}}\underbrace{\sqrt{\mathrm{Var}[\hat{b}_i]}}_{\mathcal{O}(1)} \\
&= \mathcal{O}(\frac{1}{\sqrt{n}})
\end{aligned}
$$

   Hence, the estimation error $\hat{a}-a$ is shown to have a standard deviation of $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$.

4. If we were to apply Laplace mechanism to all the private bits, the central model output would be $\hat{a}_c = \underbrace{\frac{1}{n}\sum_{i=1}^{n}b_i}_{a}+\underbrace{\eta}_{\text{noise}}$, for which the global l_1 sensitivity is given as $\Delta a=\frac{1}{n}$. Then, for $\epsilon=\ln(3)$, our

noise will follow the Laplace distribution as:

$$\eta \sim \text{Laplace}(0, \frac{\Delta a}{\epsilon} = \frac{1}{n\ln(3)})$$

Then, $\sigma(\hat{a}_c - a) = \sqrt{\text{Var}(\hat{a}_c - a)} = \sqrt{\text{Var}(\eta)} = \sqrt{2(\frac{\Delta a}{\epsilon})^2} = \frac{\sqrt{2}}{n\ln(3)} = \mathcal{O}(\frac{1}{n})$.

5. We let the first coin to have its weight adjusted such that p is the probability of it being the head and $1-p$ being the tail. Let $\mathcal{M}$ be the corresponding mechanism. Consider the private bit b_i cases:

   - Case $b_i = 0$:
     - $\Pr(\mathcal{M}(b_i) = 0) = \frac{1+p}{2}$
     - $\Pr(\mathcal{M}(b_i) = 1) = \frac{1-p}{2}$
   - Case $b_i = 1$:
     - $\Pr(\mathcal{M}(b_i) = 0) = \frac{1-p}{2}$
     - $\Pr(\mathcal{M}(b_i) = 1) = \frac{1+p}{2}$

   For neighboring datasets $\mathcal{D}_1 = \mathcal{D} \cup \{b_i = 0\}$ and $\mathcal{D}_2 = \mathcal{D} \cup \{b_i = 1\}$,

$$\frac{\Pr(\mathcal{M}(\mathcal{D}_1) = 0)}{\Pr(\mathcal{M}(\mathcal{D}_2) = 0)} \leq \frac{\Pr(\mathcal{M}(b_i = 0) = 0)}{\Pr(\mathcal{M}(b_i = 1) = 0)} = \frac{1+p}{1-p} \leq e^\epsilon \Rightarrow p \leq \frac{e^\epsilon - 1}{e^\epsilon + 1}$$

   The following proves that the untrusted aggregator that receives all these noisy bits can compute an unbiased estimate $\hat{a}$ of a (not only in the case of the first coin being a fair coin).

$$\begin{aligned}
\mathbb{E}[\hat{b}_i] &= \text{P}(b_i = 0)\mathbb{E}[\hat{b}_i | b_i = 0] + \text{P}(b_i = 1)\mathbb{E}[\hat{b}_i | b_i = 1] \\
&= (1-a) \cdot (0 \cdot \frac{1+p}{2} + 1 \cdot \frac{1-p}{2}) + a \cdot (0 \cdot \frac{1-p}{2} + 1 \cdot \frac{1+p}{2}) \\
&= (1-a) \cdot \frac{1-p}{2} + a \cdot \frac{1+p}{2} \\
&= \frac{1-p}{2} - \frac{a-ap}{2} + \frac{a+ap}{2} \\
&= ap - \frac{p-1}{2} \\
\Rightarrow a &= \frac{1}{p}(\frac{p-1}{2} + \mathbb{E}(\hat{b}_i)) \\
\Rightarrow \hat{a} &= \frac{1}{p}(\frac{p-1}{2} + \frac{1}{n}\sum_{i=1}^{n} \hat{b}_i) \\
\Rightarrow \mathbb{E}[\hat{a}] &= \mathbb{E}[\frac{1}{p}(\frac{p-1}{2} + \frac{1}{n}\sum_{i=1}^{n} \hat{b}_i)] \\
&= \frac{1}{p}\frac{p-1}{2} + \frac{1}{p}\mathbb{E}[(\frac{1}{n}\sum_{i=1}^{n} \hat{b}_i)] \\
&= \frac{1}{p}\frac{p-1}{2} + \frac{1}{p}\mathbb{E}[\hat{b}_i] \\
&= a
\end{aligned}$$

The variance of our estimator is given by:

$$\begin{aligned}
\text{Var}[\hat{a}] &= \text{Var}[\frac{1}{p}(\frac{p-1}{2} + \frac{1}{n}\sum_{i=1}^{n}\hat{b}_i)] \\
&= \text{Var}[\frac{1}{pn}\sum_{i=1}^{n}\hat{b}_i] \\
&= \frac{1}{p^2n^2}\text{Var}[\sum_{i=1}^{n}\hat{b}_i] \\
&= \frac{1}{p^2n}\text{Var}[\hat{b}_i] \\
&= \frac{1}{p^2n}(\mathbb{E}[\hat{b}_i^2] - \mathbb{E}[\hat{b}_i]^2) \\
&= \frac{1}{p^2n}(\underbrace{\mathbb{E}[\hat{b}_i]}_{\text{since } \hat{b}_i\in\{0,1\} \text{ then } \hat{b}_i^2=\hat{b}_i} - \mathbb{E}[\hat{b}_i]^2) \\
&= \frac{1}{p^2n}(\frac{1+p}{2}\hat{b}_i + \frac{1-p}{2}(1-\hat{b}_i) - (\frac{1+p}{2}\hat{b}_i + \frac{1-p}{2}(1-\hat{b}_i))^2) \\
&= \frac{1}{p^2n}(\frac{1+p}{2}\hat{b}_i + \frac{1-p}{2}(1-\hat{b}_i))(1 - \frac{1+p}{2}\hat{b}_i - \frac{1-p}{2}(1-\hat{b}_i)) \\
&= \frac{1}{p^2n}(p\hat{b}_i - \frac{\hat{b}_i}{2} - \frac{p}{2})(1 - (p\hat{b}_i - \frac{\hat{b}_i}{2} - \frac{p}{2})) \\
&= \underbrace{-\frac{\hat{b}_i^2}{n}}_{-\frac{\hat{b}_i}{2}} + \frac{\hat{b}_i}{n} - \frac{1}{4n} + \frac{1}{4np^2} \text{ (equation is obtained using Python's } \texttt{simplify} \text{ library)} \\
&= \frac{1}{4n}(\frac{1-p^2}{p^2}) \\
\Rightarrow \sigma(\hat{a}) &= \sqrt{\text{Var}[\hat{a}]} \\
&= \frac{1}{2\sqrt{n}}\sqrt{\frac{1-p^2}{p^2}}
\end{aligned}$$

We only need to show that $\sqrt{\frac{1-p^2}{p^2}} = \mathcal{O}(\frac{1}{\epsilon})$ so that $\sigma(\hat{a}) = \mathcal{O}(\frac{1}{\epsilon\sqrt{n}})$; equivalently,

$$\begin{aligned}
&\frac{1-p^2}{p^2} \leq c\frac{1}{\epsilon^2} \text{ (for some constant } c) \\
\Rightarrow &\frac{p^2}{1-p^2} \leq c'\epsilon^2 \text{ (let } c' = \frac{1}{c}) \\
&\Rightarrow p^2 \leq \frac{c'\epsilon^2}{c'\epsilon^2+1} \underbrace{\leq \left(\frac{e^\epsilon - 1}{e^\epsilon + 1}\right)^2}_{\text{ensure } \epsilon\text{-DF}} = 1 - \frac{4e^\epsilon}{(e^\epsilon+1)^2}
\end{aligned}$$

The above implies that

$$\frac{c'\epsilon^2}{c'\epsilon^2+1} = 1 - \frac{1}{c'\epsilon^2+1} \leq 1 - \frac{4e^\epsilon}{(e^\epsilon+1)^2}$$

$$\Rightarrow \frac{1}{c'\epsilon^2+1} \geq \frac{4e^\epsilon}{(e^\epsilon+1)^2}$$
$$\Rightarrow c'\epsilon^2+1 \leq \frac{(e^\epsilon+1)^2}{4e^\epsilon}$$

Claim that $c=4$ satisfies the above inequality. Then, the corresponding inequality

$$\frac{1}{4}\epsilon^2+1 = \frac{1}{4}\epsilon^2+\frac{1}{2}+\frac{1}{2} \leq \frac{e^\epsilon}{4}+\frac{1}{4e^\epsilon}+\frac{1}{2} = \frac{(e^\epsilon+1)^2}{4e^\epsilon}$$

is still satisfied $\forall\epsilon \geq 0$ as $\frac{1}{4}(e^\epsilon+\frac{1}{e^\epsilon}) \geq \frac{1}{4}\epsilon^2+\frac{1}{2}$ is true $\forall\epsilon \geq 0$. Proof will be shown in the latter part.

The above implies that $\exists c$ such that $\frac{1-p^2}{p^2} \leq c\frac{1}{\epsilon^2}, \forall\epsilon \geq \epsilon_0$ and $\epsilon_0 = 0$; thus,

$$\sigma(\hat{a}) = \frac{1}{2\sqrt{n}} \underbrace{\sqrt{\frac{1-p^2}{p^2}}}_{\mathcal{O}(\frac{1}{\epsilon})} = \mathcal{O}(\frac{1}{\epsilon\sqrt{n}})$$

is shown.

For $f(x) = \frac{1}{4}(e^x+e^{-x})-\frac{1}{4}x^2-\frac{1}{2}$, we have $\frac{\partial f}{\partial x}(x) = \frac{1}{4}e^x-\frac{1}{4}e^{-x}-\frac{1}{2}x$ and $\frac{\partial^2 f}{(\partial x)^2}(x) = \frac{1}{4}e^x+\frac{1}{4}e^{-x}-\frac{1}{2}$. Since $\frac{\partial^2 f}{(\partial x)^2}(x) \geq 0, \forall x \geq 0$, then $\frac{\partial f}{\partial x}(x) \geq 0, \forall x \geq 0$ as $\frac{\partial f}{\partial x}(0) = 0$, which further implies that $f(x) \geq 0, \forall x \geq 0$ as $f(0) = 0$. Then, $f(\epsilon) \geq 0 \Rightarrow \frac{1}{4}(e^x+e^{-x})-\frac{1}{4}x^2-\frac{1}{2} \geq 0 \Rightarrow \frac{1}{4}(e^x+e^{-x}) \geq \frac{1}{4}x^2+\frac{1}{2}$.

3.

Suppose you have two algorithms to predict whether a student can obtain an offer from Harvard University based on some features such as GPA, number of credits, internship, and research experience. The following table shows the ground-truth label and the predictions made by the two algorithms. The sensitive attribute considered in this fairness study is gender.

Gender	Label (truth)	Algorithm 1	Algorithm 2
Male	Yes	No	Yes
Male	No	Yes	Yes
Male	Yes	Yes	No
Male	Yes	No	No
Female	Yes	Yes	Yes
Female	No	No	Yes
Female	Yes	No	No
Female	Yes	Yes	No

1. Determine if the algorithms satisfy demographic parity. Show the derivation.
2. Determine if the algorithms satisfy equal opportunity. Show the derivation.
3. Determine if the algorithms satisfy equalized odds. Show the derivation.

Below are the notations used in the context of assessing the fairness of predictive algorithms for the above problem:

- $\hat{Y}$: represents the predicted label output by the corresponding algorithms.

- Y: represents the true label or ground truth.
- A: represents the sensitive attribute, which in this context is gender.
  - $A = 0$ denotes Male
  - $A = 1$ denotes Female

1.
   - Algorithm 1:
     - $P(\hat{Y} = 1|A = 0) = 0.5$
     - $P(\hat{Y} = 1|A = 1) = 0.5$

     Since $P(\hat{Y} = 1|A = 0) = P(\hat{Y} = 1|A = 1)$, algorithm 1 satisfies the demographic parity.
   - Algorithm 2:
     - $P(\hat{Y} = 1|A = 0) = 0.5$
     - $P(\hat{Y} = 1|A = 1) = 0.5$

     Since $P(\hat{Y} = 1|A = 0) = P(\hat{Y} = 1|A = 1)$, algorithm 2 satisfies the demographic parity.

2.
   - Algorithm 1:
     - $P(\hat{Y} = 1|A = 0, Y = 1) = 0.\bar{3}$
     - $P(\hat{Y} = 1|A = 1, Y = 1) = 0.\bar{6}$

     Since $P(\hat{Y} = 1|A = 0, Y = 1) \neq P(\hat{Y} = 1|A = 1, Y = 1)$, algorithm 1 does not satisfy the equalized opportunity.
   - Algorithm 2:
     - $P(\hat{Y} = 1|A = 0, Y = 1) = 0.\bar{3}$
     - $P(\hat{Y} = 1|A = 1, Y = 1) = 0.\bar{3}$

     Since $P(\hat{Y} = 1|A = 0, Y = 1) = P(\hat{Y} = 1|A = 1, Y = 1)$, algorithm 2 satisfies the equalized opportunity.

3.
   - Algorithm 1:
     - $P(\hat{Y} = 1|A = 0, Y = 0) = 1$
     - $P(\hat{Y} = 1|A = 1, Y = 0) = 0$
     - $P(\hat{Y} = 1|A = 0, Y = 1) = 0.\bar{3}$
     - $P(\hat{Y} = 1|A = 1, Y = 1) = 0.\bar{6}$

     Since $P(\hat{Y} = 1|A = 0, Y) \neq P(\hat{Y} = 1|A = 1, Y)$, algorithm 1 does not satisfy the equalized odds.
   - Algorithm 2:
     - $P(\hat{Y} = 1|A = 0, Y = 0) = 1$
     - $P(\hat{Y} = 1|A = 1, Y = 0) = 1$
     - $P(\hat{Y} = 1|A = 0, Y = 1) = 0.\bar{3}$
     - $P(\hat{Y} = 1|A = 1, Y = 1) = 0.\bar{3}$

     Since $P(\hat{Y} = 1|A = 0, Y) = P(\hat{Y} = 1|A = 1, Y)$, algorithm 2 satisfies the equalized odds.

4.

In machine learning, fairness is an important consideration when building models that are used to make decisions affecting people. Two common fairness metrics are demographic parity and equalized odds. Demographic parity requires that the classification rates (e.g., acceptance rates) for different demographic groups should be equal or close to equal. Equalized odds requires that the true positive rate (TPR) and false positive rate (FPR) are equal across groups. Consider a binary classification task for a loan approval system. The system predicts whether to approve or reject a loan application based on several features. You are given the following information about the loan applicants:

- There are 1000 applicants in total.
- 600 applicants belong to Group A and 400 applicants belong to Group B.
- 200 applicants from Group A are approved for loans.
- The overall loan approval rate for all applicants is 25%.
- The true positive rate (TPR) for Group A is 50%.
- The false positive rate (FPR) for Group A is 20%.

Your task is to analyze fairness in this loan approval system using demographic parity and equalized odds. Answer the following questions:

1. Calculate the loan approval rate for Group A.
2. Calculate the loan approval rate for Group B.
3. Determine whether demographic parity is achieved in this loan approval system. If not, what is the difference in approval rates between Group A and Group B?
4. Calculate the true positive rate (TPR) and false positive rate (FPR) for Group B, assuming equalized odds are achieved.
5. Suppose you want to achieve both demographic parity and equalized odds by adjusting the loan approval thresholds for Group A and Group B. Calculate the number of applicants from each group that need to be approved for loans to achieve both demographic parity and equalized odds while keeping the overall loan approval rate constant. If it is not possible to achieve both demographic parity and equalized odds simultaneously, explain why.

1. Loan approval rate for group A is calculated as $\frac{200}{600} = 0.\bar{3}$
2. Loan approval rate for group B is calculated as $\frac{25\% \times 1000 - 200}{400} = 0.125$
3. As approval rates for both groups are not equal (not even close), demographic parity is not achieved in this loan approval system. The difference between both approval rates of group A and B is $\frac{1}{3} - \frac{1}{8} = \frac{5}{24} = 0.208\bar{3}$.
4. In equalized odds criterion, both TPR and FPR for group A and B are equal. Then, TPR for group B is 50% and FPR for group B is 20%.
5. Both loan approval rates for group A and B can be adjusted to 25% each to satisfy the demographic parity aspect: 150 participants from group A are approved and 100 applicants from group B are

approved. However, we still lack the information for the proportion of the participants' eligibility in each group; hence, it is impossible to achieve both demographic parity and equalized odds simultaneously in this case.

5.

For a set of d players represented by $D = \{1, \ldots, d\}$ and a cooperative game $v : 2^D \to \mathbb{R}$, the Shapley value for each player $i \in D$ is defined as:

$$\phi_i(v) = \sum_{S \subseteq D \setminus \{i\}} \frac{|S|!(d - |S| - 1)!}{d!} (v(S \cup \{i\}) - v(S))$$

1. Prove the null player axiom of the Shapley value, which states that if a player contributes no value in a game $v : 2^D \to \mathbb{R}$, or $v(S \cup \{i\}) - v(S) = 0$ for all $S \subseteq D \setminus \{i\}$, then $\phi_i(v) = 0$.

2. Prove the linearity axiom of the Shapley value, which states that given two cooperative games $u : 2^D \to \mathbb{R}$ and $v : 2^D \to \mathbb{R}$ and a third game we defined as $w(S) = u(S) + v(S)$ for all $S \subseteq D$, the following holds for all players $i \in D$:

$$\phi_i(w) = \phi_i(u) + \phi_i(v).$$

3. Prove the efficiency axiom of the Shapley value, which states that the following holds for all games $v : 2^D \to \mathbb{R}$:

$$\sum_{i=1}^{d} \phi_i(v) = v(D) - v(\emptyset).$$

4. Prove that the Null Player, Symmetry, and Linearity axioms can be replaced by a single property

$$\phi_i(D) - \phi_i(D \setminus \{j\}) = \phi_j(D) - \phi_j(D \setminus \{i\})$$

for all $i, j \in D$ with $i \neq j$ where $\phi_i(D \setminus \{j\})$ denotes the Shapley value of the player i with player j removed.

Denote the coefficient of term w.r.t S as $k(S)$, where $k(S) = \frac{|S|!(d-|S|-1)!}{d!}$.

1. If player i contributes no value in a game $v : 2^D \to \mathbb{R}$, then $v(S \cup \{i\}) - v(S) = 0, \forall S \subseteq D \setminus \{i\}$. The null player axiom of the Shapley value is proved as follows:

$$\begin{aligned}
\phi_i(v) &= \sum_{S \subseteq D \setminus \{i\}} \frac{|S|!(d - |S| - 1)!}{d!} \underbrace{(v(S \cup \{i\}) - v(S))}_{0} \\
&= \sum_{S \subseteq D \setminus \{i\}} k(S) \cdot 0 \\
&= 0
\end{aligned}$$

2. Given $u : 2^D \to \mathbb{R}$ and $v : 2^D \to \mathbb{R}$ as cooperative games. If another cooperative game is defined as $w(S) = u(S) + v(S), \forall S \subseteq D$, then $\phi_i(w) = \phi_i(u) + \phi_i(v)$. The linearity axiom of the Shapley value is proved as follows:

$$\begin{aligned}
\phi_i(w) &= \sum_{S\subseteq D\setminus\{i\}} \frac{|S|!(d-|S|-1)!}{d!}(w(S\cup\{i\})-w(S)) \\
&= \sum_{S\subseteq D\setminus\{i\}} k(S)(u(S\cup\{i\})+v(S\cup\{i\})-u(S)-v(S)) \\
&= \sum_{S\subseteq D\setminus\{i\}} k(S)(u(S\cup\{i\})-u(S)) + \sum_{S\subseteq D\setminus\{i\}} k(S)(v(S\cup\{i\})-v(S)) \\
&= \phi_i(u)+\phi_i(v)
\end{aligned}$$

3. The efficiency axiom of the Shapley value is proved as follows:

$$\begin{aligned}
\sum_{i=1}^{d}\phi_i(v) &= \sum_{i=1}^{d}\left(\sum_{S\subseteq D\setminus\{i\}} \frac{|S|!(d-|S|-1)!}{d!}(v(S\cup\{i\})-v(S))\right) \\
&= \sum_{i=1}^{d}\left(\sum_{S\subseteq D\setminus\{i\}} k(S)v(S\cup\{i\}) - \sum_{S\subseteq D\setminus\{i\}} k(S)v(S)\right) \\
&= \sum_{S\subseteq D}\left(\mathbb{I}(S\neq\emptyset)\sum_{i\in S} k(S\setminus\{i\})v(S) - \mathbb{I}(S\neq D)\sum_{i\notin S} k(S)v(S)\right) \\
&= \sum_{i\in D} k(D\setminus\{i\})v(D) - \sum_{i\in D} k(\emptyset)v(\emptyset) + \sum_{S\subset D, S\neq\emptyset}\left(\sum_{i\in S} k(S\setminus\{i\})v(S) - \sum_{i\notin S} k(S)v(S)\right) \\
&= d\frac{(d-1)!0!}{d!}v(D) - d\frac{0!(d-1)!}{d!}v(\emptyset) + \sum_{S\subset D, S\neq\emptyset}\left(\sum_{i\in S} k(S\setminus\{i\})v(S) - \sum_{i\notin S} k(S)v(S)\right) \\
&= v(D)-v(\emptyset) + \sum_{S\subset D, S\neq\emptyset}\left(\sum_{i\in S} k(S\setminus\{i\})v(S) - \sum_{i\notin S} k(S)v(S)\right) \\
&= v(D)-v(\emptyset) + \sum_{S\subset D, S\neq\emptyset} v(S)\left(\sum_{i\in S} k(S\setminus\{i\}) - \sum_{i\notin S} k(S)\right) \\
&= v(D)-v(\emptyset) + \sum_{S\subset D, S\neq\emptyset} v(S)\left(\underbrace{|S|\frac{(|S|-1)!(d-|S|)!}{d!}}_{\binom{d}{|S|}^{-1}} - \underbrace{(d-|S|)\frac{|S|!(d-|S|-1)!}{d!}}_{\binom{d}{|S|}^{-1}}\right) \\
&= v(D)-v(\emptyset)
\end{aligned}$$

4. Consider $D(d)$ as a function that maps a positive integer d to the set $\{x\in\mathbb{N}\mid 1\leq x\leq d\}$. The proof proceeds by induction as follows:

   ◦ By the efficiency axiom of the Shapley value, when $d=1$,

$$\sum_{i=1}^{1}\phi_i(v) = v(\{1\})-v(\emptyset) \Rightarrow \phi_i(v) = \sum_{S\subseteq D\setminus\{i\}} \frac{|S|!(d-|S|-1)!}{d!}(v(S\cup\{i\})-v(S)), \forall i\in D(1)$$

◦ Assume both the efficiency axiom and the single property can derive the Shapley's formula for $d = k - 1$ and k is an integer larger than 1, then

$$\begin{aligned}
\phi_i(D(k-1)) - \phi_j(D(k-1)) &= \phi_i(D(k-1) \setminus \{j\}) - \phi_j(D(k-1) \setminus \{i\}) \\
&= \sum_{S \subseteq D(k-1) \setminus \{i\}} \frac{|S|!(k-|S|-2)!}{(k-1)!}(v(S \cup \{i\}) - v(S)) \\
&\quad - \sum_{S \subseteq D(k-1) \setminus \{j\}} \frac{|S|!(k-|S|-2)!}{(k-1)!}(v(S \cup \{j\}) - v(S)) \\
&= \sum_{S \subseteq D(k-1) \setminus \{i,j\}} \frac{|S|!(k-|S|-2)!}{(k-1)!}(v(S \cup \{i\}) - v(S \cup \{j\}))
\end{aligned}$$

◦ The following will demonstrate that the previous assertion remains valid for $d = k$:

$$\begin{aligned}
\phi_i(D(k)) &= \phi_j(D(k)) + \phi_i(D(k) \setminus \{j\}) - \phi_j(D(k) \setminus \{i\}), \forall j \in D(k) \\
&= \frac{1}{k} \sum_{j=1}^{k} (\phi_j(D(k)) + \phi_i(D(k) \setminus \{j\}) - \phi_j(D(k) \setminus \{i\})) \\
&= \frac{1}{k} \left(\underbrace{\sum_{j=1}^{k} \phi_j(D(k))}_{\text{efficiency axiom}} + \sum_{j=1}^{k} (\phi_i(D(k) \setminus \{j\}) - \phi_j(D(k) \setminus \{i\})) \right) \\
&= \frac{1}{k} \left(v(D(k)) - v(\emptyset) + \sum_{j=1}^{k} (\phi_i(D(k) \setminus \{j\}) - \phi_j(D(k) \setminus \{i\})) \right) \\
&= \frac{1}{k} \left(v(D(k)) - v(\emptyset) - \underbrace{\sum_{j=1, j \neq i}^{k} \phi_j(D(k) \setminus \{i\})}_{\text{efficiency axiom}} - \phi_j(D(k) \setminus \{i\}) + \sum_{j=1}^{k} \phi_i(D(k) \setminus \{j\}) \right) \\
&= \frac{1}{k} \left(v(D(k)) - v(D(k) \setminus \{i\}) - \phi_i(D(k) \setminus \{i\}) + \sum_{j=1}^{k} \phi_i(D(k) \setminus \{j\}) \right) \\
&= \frac{1}{k}(v(D(k)) - v(D(k) \setminus \{i\})) + \frac{1}{k} \sum_{j=1, j \neq i}^{k} \phi_i(D(k) \setminus \{j\}) \\
&= \frac{1}{k}(v(D(k)) - v(D(k) \setminus \{i\})) + \frac{1}{k} \sum_{j=1, j \neq i}^{k} \sum_{S \subseteq D(k) \setminus \{i,j\}} \frac{|S|!(k-|S|-2)!}{(k-1)!}(v(S \cup \{i\}) - v(S)) \\
&= \frac{1}{k}(v(D(k)) - v(D(k) \setminus \{i\})) + \sum_{j=1, j \neq i}^{k} \sum_{S \subseteq D(k) \setminus \{i,j\}} \frac{|S|!(k-|S|-2)!}{k!}(v(S \cup \{i\}) - v(S)) \\
&= \frac{1}{k}(v(D(k)) - v(D(k) \setminus \{i\})) + \sum_{S \subset D(k) \setminus \{i\}} \sum_{j=1, j \neq D(k) \cup i}^{k} \frac{|S|!(k-|S|-2)!}{k!}(v(S \cup \{i\}) - v(S)) \\
&= \frac{1}{k}(v(D(k)) - v(D(k) \setminus \{i\})) + \sum_{S \subset D(k) \setminus \{i\}} \frac{|S|!(k-|S|-1)!}{k!}(v(S \cup \{i\}) - v(S))
\end{aligned}$$

$$= \sum_{S \subseteq D(k) \setminus \{i\}} \frac{|S|!(k-|S|-1)!}{k!}(v(S \cup \{i\}) - v(S))$$

By induction, both the efficiency axiom and the single property can derive the Shapley's formula without additional axioms. Equivalently, the single property can replace the null player, symmetry, and linearity axioms.

6.

We consider some examples of cooperative games and calculate their Shapley values.

1. Calculate the Shapley value for all players $i \in \{1,2,3\}$ for the following cooperative game characterized by $v(S)$.

S	$\emptyset$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$
$v(S)$	0	2	3	4	5	6	7	8

2. Calculate the Shapley value for all players $i \in \{1,2,3\}$ in the game $v(S)$ given by

$$v(S) = 2x_1 + 3x_2 + 4x_3.$$

where x_i are binary variables that are equal to 1 if $i \in S$ and 0 otherwise.

3. Calculate the Shapley values for all players $i \in \{1,2,3,4,5\}$ in the game $v(S)$ given by

$$v(S) = 2x_2 + 3x_3 + 4x_4 + 5x_1x_3 + 7x_2x_5 - 12x_1x_2x_3$$

where x_i are binary variables that are equal to 1 if $i \in S$ and 0 otherwise.

1.
- For Player 1:

$$\phi_1(v) = \frac{1}{3}(v(\{1\})-v(\emptyset))+\frac{1}{6}(v(\{1,2\})-v(\{2\}))+\frac{1}{6}(v(\{1,3\})-v(\{3\}))+\frac{1}{3}(v(\{1,2,3\})-v(\{2,3\}))$$

, where:
  - $S = \emptyset$: $v(\{1\}) - v(\emptyset) = 2 - 0 = 2$
  - $S = \{2\}$: $v(\{1,2\}) - v(\{2\}) = 5 - 3 = 2$
  - $S = \{3\}$: $v(\{1,3\}) - v(\{3\}) = 6 - 4 = 2$
  - $S = \{2,3\}$: $v(\{1,2,3\}) - v(\{2,3\}) = 8 - 7 = 1$

$$\phi_1(v) = \frac{1}{3} \times 2 + \frac{1}{6} \times 2 + \frac{1}{6} \times 2 + \frac{1}{3} \times 1 = \frac{5}{3}$$

- For Player 2:

$$\phi_2(v) = \frac{1}{3}(v(\{2\})-v(\emptyset))+\frac{1}{6}(v(\{1,2\})-v(\{1\}))+\frac{1}{6}(v(\{2,3\})-v(\{3\}))+\frac{1}{3}(v(\{1,2,3\})-v(\{1,3\}))$$

, where:
  - $S = \emptyset$: $v(\{2\}) - v(\emptyset) = 3 - 0 = 3$
  - $S = \{1\}$: $v(\{1,2\}) - v(\{1\}) = 5 - 2 = 3$

- $S = \{3\}$: $v(\{2,3\}) - v(\{3\}) = 7 - 4 = 3$
- $S = \{1,3\}$: $v(\{1,2,3\}) - v(\{1,3\}) = 8 - 6 = 2$

$$\phi_2(v) = \frac{1}{3} \times 3 + \frac{1}{6} \times 3 + \frac{1}{6} \times 3 + \frac{1}{3} \times 2 = \frac{8}{3}$$

- For Player 3:

$$\phi_3(v) = \frac{1}{3}(v(\{3\}) - v(\emptyset)) + \frac{1}{6}(v(\{1,3\}) - v(\{1\})) + \frac{1}{6}(v(\{2,3\}) - v(\{2\})) + \frac{1}{3}(v(\{1,2,3\}) - v(\{1,2\}))$$

, where:

- $S = \emptyset$: $v(\{3\}) - v(\emptyset) = 4 - 0 = 4$
- $S = \{1\}$: $v(\{1,3\}) - v(\{1\}) = 6 - 2 = 4$
- $S = \{2\}$: $v(\{2,3\}) - v(\{2\}) = 7 - 3 = 4$
- $S = \{1,2\}$: $v(\{1,2,3\}) - v(\{1,2\}) = 8 - 5 = 3$

$$\phi_3(v) = \frac{1}{3} \times 4 + \frac{1}{6} \times 4 + \frac{1}{6} \times 4 + \frac{1}{3} \times 3 = \frac{11}{3}$$

2. - For Player 1:

$$\phi_1(v) = \frac{1}{3}(v(\{1\}) - v(\emptyset)) + \frac{1}{6}(v(\{1,2\}) - v(\{2\})) + \frac{1}{6}(v(\{1,3\}) - v(\{3\})) + \frac{1}{3}(v(\{1,2,3\}) - v(\{2,3\}))$$

, where:

- $S = \emptyset$: $v(\{1\}) - v(\emptyset) = 2 - 0 = 2$
- $S = \{2\}$: $v(\{1,2\}) - v(\{2\}) = 5 - 3 = 2$
- $S = \{3\}$: $v(\{1,3\}) - v(\{3\}) = 6 - 4 = 2$
- $S = \{2,3\}$: $v(\{1,2,3\}) - v(\{2,3\}) = 9 - 7 = 2$

$$\phi_1(v) = \frac{1}{3} \times 2 + \frac{1}{6} \times 2 + \frac{1}{6} \times 2 + \frac{1}{3} \times 2 = 2$$

- For Player 2:

$$\phi_2(v) = \frac{1}{3}(v(\{2\}) - v(\emptyset)) + \frac{1}{6}(v(\{1,2\}) - v(\{1\})) + \frac{1}{6}(v(\{2,3\}) - v(\{3\})) + \frac{1}{3}(v(\{1,2,3\}) - v(\{1,3\}))$$

, where:

- $S = \emptyset$: $v(\{2\}) - v(\emptyset) = 3 - 0 = 3$
- $S = \{1\}$: $v(\{1,2\}) - v(\{1\}) = 5 - 2 = 3$
- $S = \{3\}$: $v(\{2,3\}) - v(\{3\}) = 7 - 4 = 3$
- $S = \{1,3\}$: $v(\{1,2,3\}) - v(\{1,3\}) = 9 - 6 = 3$

$$\phi_2(v) = \frac{1}{3} \times 3 + \frac{1}{6} \times 3 + \frac{1}{6} \times 3 + \frac{1}{3} \times 3 = 3$$

- For Player 3:

$$\phi_3(v) = \frac{1}{3}(v(\{3\}) - v(\emptyset)) + \frac{1}{6}(v(\{1,3\}) - v(\{1\})) + \frac{1}{6}(v(\{2,3\}) - v(\{2\})) + \frac{1}{3}(v(\{1,2,3\}) - v(\{1,2\}))$$

, where:

- $S = \emptyset$: $v(\{3\}) - v(\emptyset) = 4 - 0 = 4$
- $S = \{1\}$: $v(\{1,3\}) - v(\{1\}) = 6 - 2 = 4$

- $S = \{2\}$: $v(\{2,3\}) - v(\{2\}) = 7 - 3 = 4$
- $S = \{1,2\}$: $v(\{1,2,3\}) - v(\{1,2\}) = 9 - 5 = 4$

$$\phi_3(v) = \frac{1}{3} \times 4 + \frac{1}{6} \times 4 + \frac{1}{6} \times 4 + \frac{1}{3} \times 4 = 4$$

3.
- For Player 1: $\phi_1(v) = -1.5$
- For Player 2: $\phi_2(v) = 1.5$
- For Player 3: $\phi_3(v) = 1.5$
- For Player 4: $\phi_4(v) = 4.0$
- For Player 5: $\phi_5(v) = 3.5$

The following code is used to calculate the Shapley value for each player:

```python
import math
from itertools import combinations

D = {1, 2, 3, 4, 5}
def v(S):
    ret = 0
    if 2 in S: ret += 2
    if 3 in S: ret += 3
    if 4 in S: ret += 4
    if 1 in S and 2 in S: ret += 5
    if 2 in S and 5 in S: ret += 7
    if 1 in S and 2 in S and 3 in S: ret -= 12
    return ret

def k(S, d):
    return math.factorial(len(S)) * math.factorial(d - len(S) - 1) / math.
   factorial(d)

def Shapley(D, i):
    D.discard(i)
    ret = 0
    for r in range(len(D) + 1):
        for S in combinations(D, r):
            S_with_i = list(S)
            S_with_i.append(i)
            ret += k(S, len(D) + 1) * (v(S_with_i) - v(S))
    D.add(i)
    return ret

for i in range(1, len(D) + 1):
    print("For player", i, Shapley(D, i))
```

Output:

```
For player 1 -1.5000000000000002
For player 2 1.5
For player 3 1.4999999999999998
For player 4 4.000000000000001
For player 5 3.5
```