Task

Prove that except for trivial situations (independence between a decision with a group), no two of the three fairness equalities (demographic parity, equal opportunity, predictive rate parity) can occur simultaneously.

Definitions

($*$) Demographic parity: $\mathbb{P}\left(\hat{Y}=\hat{y} \mid A=a\right)=\mathbb{P}\left(\hat{Y}=\hat{y} \mid A=b\right) \quad \forall \hat{y} \in\{0,1\}$

($**$) Equal opportunity: $\mathbb{P}\left(\hat{Y}=\hat{y} \mid Y=y, A=a\right)=\mathbb{P}\left(\hat{Y}=\hat{y} \mid Y=y, A=b\right) \forall \hat{y}, y \in\{0,1\}$

($***$) Predictive rate parity: $\mathbb{P}\left(Y=y \mid \hat{Y}=\hat{y}, A=a\right)=\mathbb{P}\left(Y=y \mid \hat{Y}=\hat{y}, A=b\right) \forall \hat{y}, y \in\{0,1\}$

Part 1

Assume demographic parity and predictive rate parity occurs simultaneously. Multiplying ($* * *$) and ($*$) side by side and using conditional probability definition we get

$$\frac{\mathbb{P}\left(Y=y, \hat{Y}=\hat{y}, A=a\right)}{\mathbb{P}\left(\hat{Y}=\hat{y}, A=a\right)} \cdot \frac{\mathbb{P}\left(\hat{Y}=\hat{y}, A=a\right)}{\mathbb{P}(A=a)}=\frac{\mathbb{P}\left(Y=y, \hat{Y}=\hat{y}, A=b\right)}{\mathbb{P}\left(\hat{Y}=\hat{y}, A=b\right)} \cdot \frac{\mathbb{P}\left(\hat{Y}=\hat{y}, A=b\right)}{\mathbb{P}(A=b)}$$

Simplifying

$$\frac{\mathbb{P}\left(Y=y, \hat{Y}=\hat{y}, A=a\right)}{\mathbb{P}(A=a)}=\frac{\mathbb{P}\left(Y=y, \hat{Y}=\hat{y}, A=b\right)}{\mathbb{P}(A=b)}$$

And using definition again

$$\mathbb{P}\left(Y=y, \hat{Y}=\hat{y} \mid A=a\right)=\mathbb{P}\left(Y=y, \hat{Y}=\hat{y} \mid A=b\right)$$

So it means $(Y, \hat{Y})$ is independent of A so Y is independent of A.

Part 2

Assume demographic parity and equal opportunity occurs simultaneously.
We have following equation:

$$\mathbb{P}\left(\hat{Y}=\hat{y} \mid A=a\right)=\frac{\mathbb{P}\left(\hat{Y}=\hat{y}, A=a\right)}{\mathbb{P}(A=a)}=\sum_{y} \frac{\mathbb{P}\left(\hat{Y}=\hat{y}, Y=y, A=a\right)}{\mathbb{P}(A=a)}=$$

$$=\sum_{y} \frac{\mathbb{P}\left(\hat{Y}=\hat{y}, Y=y, A=a\right)}{\mathbb{P}(A=a, Y=y)} \cdot \frac{\mathbb{P}(A=a, Y=y)}{\mathbb{P}(A=a)}=$$

$$=\sum_{y} \mathbb{P}\left(\hat{Y}=\hat{y} \mid A=a, Y=y\right) \cdot \mathbb{P}(Y=y \mid A=a)$$

For $A=b$ we get analogic equation so from $(*)$ we get

$$\sum_{y} \mathbb{P}\left(\hat{Y}=\hat{y} \mid A=a, Y=y\right) \cdot \mathbb{P}(Y=y \mid A=a)=\mathbb{P}\left(\hat{Y}=\hat{y} \mid A=a\right)=\mathbb{P}\left(\hat{Y}=\hat{y} \mid A=b\right)=$$

$$=\sum_{y} \mathbb{P}\left(\hat{Y}=\hat{y} \mid A=b, Y=y\right) \cdot \mathbb{P}(Y=y \mid A=b)$$

From $(**)$ we can mark

$$p_y=\mathbb{P}\left(\hat{Y}=\hat{y} \mid A=a, Y=y\right)=\mathbb{P}\left(\hat{Y}=\hat{y} \mid A=b, Y=y\right)$$

So our equation becomes

$$\sum_{y} p_y \cdot \mathbb{P}(Y=y \mid A=a)=\sum_{y} p_y \cdot \mathbb{P}(Y=y \mid A=b)$$

Moving it to one side we get:

$$\sum_{y} p_y \cdot(\mathbb{P}(Y=y \mid A=a)-\mathbb{P}(Y=y \mid A=b))=0$$

Knowing $y \in \{0,1\}$ we now have:

$$p_0 \cdot(\mathbb{P}(Y=0 \mid A=a)-\mathbb{P}(Y=0 \mid A=b))+p_1 \cdot(\mathbb{P}(Y=1 \mid A=a)-\mathbb{P}(Y=1 \mid A=b))=0$$

Now knowing $\mathbb{P}(Y=0|A=a)+\mathbb{P}(Y=1|A=a)=1$ we get:

$$(p_0-p_1) \cdot(\mathbb{P}(Y=0|A=a)-\mathbb{P}(Y=0|A=b))=0$$

And from there it means that $\hat{Y} \perp Y$ or $A \perp Y$ so thats also trivial situations.

Part 3

Assume equal opportunity and predictive rate parity occurs simultaneously. Further we assume that $\mathbb{P}(\hat{Y} = 1 \,|\, Y = 0, A = a) \neq 0$ (its the same for all $a \in A$) and we assume Y is not independent of A. We have following equations:

$$\mathbb{P}\left(Y = y \mid \hat{Y} = \hat{y}, A = a\right) = \frac{\mathbb{P}\left(Y = y, \hat{Y} = \hat{y}, A = a\right)}{\mathbb{P}\left(\hat{Y} = \hat{y}, A = a\right)} = \frac{\mathbb{P}\left(Y = y, \hat{Y} = \hat{y}, A = a\right)}{\mathbb{P}\left(\hat{Y} = \hat{y} \mid A = a\right) \cdot \mathbb{P}\left(A = a\right)}$$

$$\mathbb{P}\left(\hat{Y} = \hat{y} \mid Y = y, A = a\right) = \frac{\mathbb{P}\left(Y = y, \hat{Y} = \hat{y}, A = a\right)}{\mathbb{P}\left(Y = y, A = a\right)} = \frac{\mathbb{P}\left(Y = y, \hat{Y} = \hat{y}, A = a\right)}{\mathbb{P}\left(Y = y \mid A = a\right) \cdot \mathbb{P}\left(A = a\right)}$$

From that to equations we get

$$\mathbb{P}\left(Y = y \mid \hat{Y} = \hat{y}, A = a\right) = \frac{\mathbb{P}\left(\hat{Y} = \hat{y} \mid Y = y, A = a\right) \cdot \mathbb{P}\left(Y = y \mid A = a\right)}{\mathbb{P}\left(\hat{Y} = \hat{y} \mid A = a\right)}$$

We observe that

$$\mathbb{P}\left(\hat{Y} = \hat{y} \mid A = a\right) = \frac{\mathbb{P}\left(\hat{Y} = \hat{y}, A = a\right)}{\mathbb{P}\left(A = a\right)} = \sum_y \frac{\mathbb{P}\left(Y = y, \hat{Y} = \hat{y}, A = a\right)}{\mathbb{P}\left(A = a\right)} =$$

$$= \sum_y \mathbb{P}\left(\hat{Y} = \hat{y} \mid Y = y, A = a\right) \cdot \mathbb{P}\left(Y = y \mid A = a\right)$$

From assumptions there exists $a, b \in A$ that

$$p_a := \mathbb{P}(Y = 1 \,|A = a) \neq \mathbb{P}(Y = 1 \mid A = b) := p_b$$

From this observation we countinue setting $y = \hat{y} = 1$:

$$\mathbb{P}\left(Y = 1 \mid \hat{Y} = 1, A = a\right) = \frac{\mathbb{P}\left(\hat{Y} = 1 \mid Y = 1, A = a\right) \cdot p_a}{\mathbb{P}\left(\hat{Y} = 1 \mid Y = 1, A = a\right) \cdot p_a + \mathbb{P}\left(\hat{Y} = 1 \mid Y = 0, A = a\right) \cdot (1 - p_a)}$$

From $(***)$ we have that $\mathbb{P}\left(Y = 1 \mid \hat{Y} = 1, A = a\right) = \mathbb{P}\left(Y = 1 \mid \hat{Y} = 1, A = b\right)$ so we have 2 cases. One case is $\mathbb{P}\left(\hat{Y} = 1 \mid Y = 1, A = a\right) \neq 0$. Then $p_a \neq 0$ and $p_b \neq 0$ cause then we would have one side equals 0 and other not. But having this it's impossible cause then following our assumption about left factor of right component of denominator being nonzero we would have

$$\frac{1 - p_a}{p_a} = \frac{1 - p_b}{p_b}$$

which implies $p_a = p_b$ which is a contradiction. Second case is $\mathbb{P}\left(\hat{Y} = 1 \mid Y = 1, A = a\right) = 0$ so we have $\mathbb{P}\left(\hat{Y} = 0 \mid Y = 1, A = a\right) = 1$. Thus:

$$\mathbb{P}\left(Y = 0 \mid \hat{Y} = 0, A = a\right) = \frac{\mathbb{P}\left(\hat{Y} = 0 \mid Y = 0, A = a\right) \cdot (1 - p_a)}{\mathbb{P}\left(\hat{Y} = 0 \mid Y = 0, A = a\right) \cdot (1 - p_a) + \mathbb{P}\left(\hat{Y} = 0 \mid Y = 1, A = a\right) \cdot p_a}$$

This is also a contradiction what ends the proof.

Final solution

Final solution is equivalent to all 3 parts which are solved for now.