Fairness Metrics

Independence

- **Statistical Parity (Demographic Parity)** [XX]
  - $P(\hat{Y} = 1|A = a) = P(\hat{Y} = 1|A = b)$
- **Conditional Statistical Parity** [XX]
  - $P(\hat{Y} = 1|E = e, A = a) = P(\hat{Y} = 1|E = e, A = b)$
  - *E is a set of legitimate features that may affect the outcome.*

Separation

- **Equalized Odds** [XX]
  - $P(\hat{Y} = 1|Y = y, A = a) = P(\hat{Y} = 1|Y = y, A = b)$
    $\forall y \in \{0, 1\}$
  - `mlr3`: `fairness.equalized.odds`
    (Averages `fairness.fpr` and `fairness.tpr`)
- **Equal Opportunity** [XX]
  - $P(\hat{Y} = 1|Y = 1, A = a) = P(\hat{Y} = 1|Y = 1, A = b)$
  - *Requires equal TPR and FNR to be satisfied at the same time.*
  - `mlr3`: `fairness.tpr`
- **Predictive Equality** [XX]
  - $P(\hat{Y} = 1|Y = 0, A = a) = P(\hat{Y} = 1|Y = 0, A = b)$
    $P(\hat{Y} = 0|Y = 0, A = a) = P(\hat{Y} = 0|Y = 0, A = b)$
  - *Requires equal FPR and TNR to be satisfied at the same time.*
  - `mlr3`: `fairness.fpr`, `fairness.tnr`

Sufficiency

- **Overall Accuracy Equality** [XX]
  - $P(\hat{Y} = Y|A = a) = P(\hat{Y} = Y|A = b)$
  - `mlr3`: `fairness.acc`
- **Treatment Equality** [XX]
  - $\frac{\text{FN}}{\text{FP}}\big|_{A=a} = \frac{\text{FN}}{\text{FP}}\big|_{A=b}$