# 1 (Some) Fairness Metrics

### Independence

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

### Separation

- Equalized Odds: $P(\hat{Y} = 1|Y = y, A = a) = P(\hat{Y} = 1|Y = y, A = b) \forall y \in \{0, 1\}$
  mlr3: `fairness.equalized.odds`
- Equal Opportunity/ False negative error rate balance: $P(\hat{Y} = 1|Y = 1, A = a) = P(\hat{Y} = 1|Y = 1, A = b)$
  mlr3: `fairness.tpr`
- Predictive Equality/ False positive error rate balance: $P(\hat{Y} = 1|Y = 0, A = a) = P(\hat{Y} = 1|Y = 0, A = b)$ or
  $P(\hat{Y} = 0|Y = 0, A = a) = P(\hat{Y} = 0|Y = 0, A = b)$
  mlr3: `fairness.fpr`, `fairness.tnr`
- Treatment Equality: $\frac{\text{FN}}{\text{FP}}\Big|_{A=a} = \frac{\text{FN}}{\text{FP}}\Big|_{A=b}$

### Sufficiency

- Predictive parity/ outcome test: $P(Y = 1|\hat{Y} = 1, A = a) = P(Y = 1|\hat{Y} = 1, A = b)$
  mlr3: `fairness.ppv`
- Equal true negative rate: $P(Y = 0|\hat{Y} = 0, A = a) = P(Y = 0|\hat{Y} = 0, A = b)$
  mlr3: `fairness.npv`
- Equal false omissionr rate*: $P(Y = 1|\hat{Y} = 0, A = a) = P(Y = 1|\hat{Y} = 0, A = b)$
  mlr3: `fairness.fomr`
- Equal false discovery rate*: $P(Y = 0|\hat{Y} = 1, A = a) = P(Y = 0|\hat{Y} = 1, A = b)$
- Conditional use accuracy equality: $P(Y = 1|\hat{Y} = 1, A = a) = P(Y = 1|\hat{Y} = 1, A = b) \wedge P(Y = 0|\hat{Y} = 0, A = a) = P(Y = 0|\hat{Y} = 0, A = b)$

### Score-based

- Calibration: $P(Y = 1|S = s, A = a) = P(Y = 1|S = s, A = b)$
- Well-calibration: $P(Y = 1|S = s, A = a) = P(Y = 1|S = s, A = b) = s$
- Balance for positive class: $E(S \mid Y = 1, A = a) = E(S \mid Y = 1, A = b)$
- Balance for negative class: $E(S \mid Y = 0, A = a) = E(S \mid Y = 0, A = b)$

### Other

- Overall Accuracy Equality: $P(\hat{Y} = Y|A = a) = P(\hat{Y} = Y|A = b)$
  mlr3: `fairness.acc`

* not officially defined in any of the three papers, but following the same principles as all confusion matrix based metrics

## 2 Fairness Methods

- Preprocessing
- Inprocessing
- Postprocessing

## 3 Sources of bias and the feedback loop