According to Debreu-Groman Theorem, according to our measure: The "leveling down" objection to equality.

Axiomatic characterization: According to inequality:

\[ y \hat{=} b \] for \( 0 \leq \beta \neq 1 \)

Often limits individual level inequality

Enjoys convex formulation

Beyond binary classification

More than one group

Beyond binary classification

Useful for reasoning about both individual and group level fairness

Proposition: Consider two benefit vectors \( b, b' > 0 \) with equal means \( \{ \mu = \mu' \} \).

For \( 0 < \alpha < 1 \), \( A_{\alpha}(b) \geq A_{\alpha}(b') \) if and only if \( W_a(b) \leq W_a(b') \).

For a fixed mean benefit \( \mu \), our measure and Atkinson index lead the same indifference curves and total ordering.

A Convex Formulation

- Our formulation:
  \[
  \min_{b \in \mathcal{B}} C(b, D) \text{ s.t. } W_a(b) \geq \tau
  \]
  \[
  \min_{b \in \mathcal{B}} \frac{1}{n} \sum_{i=1}^{n} (b_i - y_i)^2
  \]
  \[
  \text{s.t. } \frac{1}{n} \sum_{i=1}^{n} (b_i - y_i + 1)^n \geq \tau
  \]

- Impact of our in-processing on model parameters:

Impact on Previous Notions of Fairness

- Core idea: social welfare as fairness behind a veil of ignorance
- Axiomatic characterization:
  - Monotonicity: \( b' \geq b \Rightarrow W(b') \geq W(b) \).
  - Independence of unconcerned agents: \( \forall b, b', a, c, \quad \{b(a) \geq b'(a)\} \leftrightarrow \{b(c) \geq b'(c)\} \).
  - Independence of common scale: \( \forall \mu > 0, \quad W(b) \geq W(b') \iff W(\mu b) \geq W(\mu b') \).
  - Anonymity
  - Progressive transfer principle
  - According to Debreu-Groman Theorem, \( W_a(b_1, \ldots, b_n) = \sum_{i=1}^{n} w_a(b_i) \), where
    - for \( 0 < a \leq 1 \), \( w_a(b) = b^a \);
    - for \( a = 0 \), \( w_a(b) = \ln(b) \);
    - for \( a < 0 \), \( w_a(b) = -b^a \)

Connection to Inequality

Atkinson Index is a welfare-based measure of inequality

\[
A_{\beta}(b_1, \ldots, b_n) = 1 - \left( \frac{1}{n} \sum_{i=1}^{n} b_i^{-\beta} \right)^{1/(1-\beta)}, \quad \text{for } 0 \leq \beta \neq 1
\]

\( \mu \), the mean benefit

compared with the Equally Distributed Equivalent (EDE)

Connection to Inequality

Cardinal social welfare as a measure of fairness behind a veil of ignorance

Addresses the leveling down objection to inequality

Enjoys a convex formulation

Often limits individual level inequality

Previous notions only characterize conditions of fairness

Our work: a principled way of generalizing to more complicated settings

Beyond binary classification

More than one group

Useful for reasoning about both individual and group level fairness

Future Directions

- Extension to other learning tasks
- Extension to descriptive (as opposed to normative) behavioral theories
- Human perception of fairness in the context of automated decision making
- What is the right benefit function?

...