Modeling Equity in AI and Optimization

John Hooker

Carnegie Mellon University

INFORMS Pittsburgh

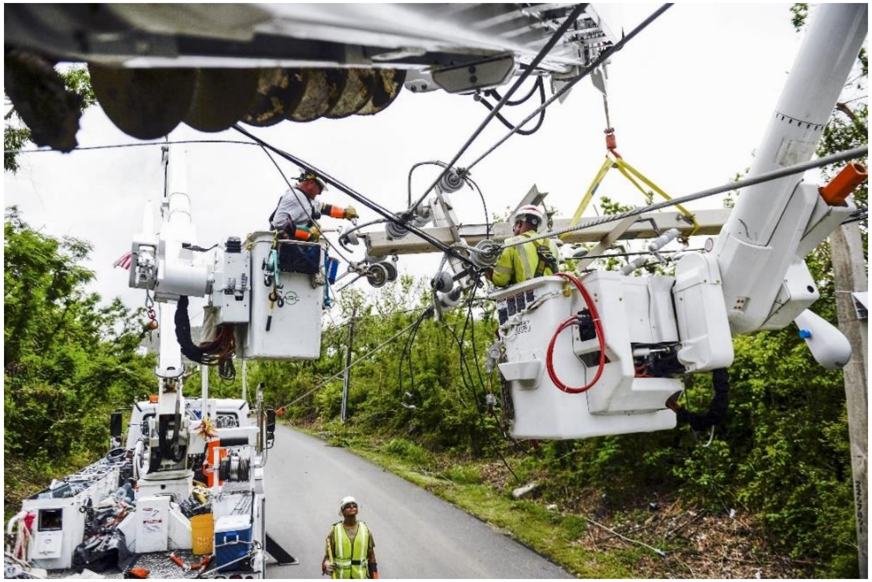
February 2022

Modeling Equity

- A growing interest in incorporating **equity** into models…
  - Health care resources.
  - Facility location (e.g., emergency services, infrastructure).
  - Telecommunications.
  - Traffic signal timing
  - Disaster recovery (e.g., power restoration)…

Modeling Equity

- Example: disaster relief
  - Power restoration can focus on **urban** areas first (**efficiency**).
  - This can leave rural areas without power for weeks/months.
  - This happened in Puerto Rico after Hurricane Maria (2017).
- A more **equitable** solution
  - …would give some priority to rural areas without overly sacrificing efficiency.



Modeling Equity

- It is far from obvious how to formulate equity concerns **mathematically**.
  - Less straightforward than maximizing total benefit or minimizing total cost.
  - Still less obvious how to **combine** equity with total benefit.

Modeling Equity

- There is **no one** concept of equity or fairness.
  - The appropriate concept **depends on the application**.
- We therefore survey a range of formulations.
  - Describe their **mathematical properties**.
  - Indicate their **strengths** and **weaknesses**.
  - State what appears to be the **most practical model**.
  - So that one can select the formulation that **best suits** a given application.
- Also a brief excursion into **social choice theory**.

References

- A more comprehensive tutorial (presented at CP 2021) is here: https://cp2021.lirmm.fr/submissions/2001

- References may be found in

V. Chen & J. N. Hooker, A guide to formulating equity and fairness in an optimization model, submitted, 2021.

Inequality measures

Criterion	*P-D?*	*C-M?*	*Linear?*	*Discrete?*
Relative range	yes	yes	yes	no
Relative mean deviation	yes	yes	yes	no
Coefficient of variation	yes	yes	no	no
Gini coefficient	yes	yes	yes	no
Hoover index	yes	yes	yes	no

Fairness for the disadvantaged

Criterion	*P-D?*	*C-M?*	*Linear?*	*Discrete?*
Maximin (Rawlsian)	yes	yes	yes	no
Leximax (lexicographic)	yes	yes	yes	no
McLoone index	no	yes	yes	yes

P-D = Pigou-Dalton
C-M = Chateauneuf-Moyes
Linear = all constraints linear
Discrete = some variables discrete

Combining efficiency & fairness
Convex combinations

Criterion	*P-D?*	*C-M?*	*Linear?*	*Discrete?*
Utility + Gini coefficient	yes	yes	no	no
Utility * Gini coefficient	yes	yes	yes	no
Utility + maximin	yes	yes	yes	no

Combining efficiency & fairness
Classical methods

Criterion	*P-D?*	*C-M?*	*Linear?*	*Discrete?*
Alpha fairness	yes	yes	yes	no
Proportional fairness (Nash bargaining)	yes	yes	yes	no
Kalai-Smorodinsky bargaining	yes	yes	no	no

P-D = Pigou-Dalton
C-M = Chateauneuf-Moyes
Linear = all constraints linear
Discrete = some variables discrete

Combining efficiency & fairness
Threshold methods

Criterion	*P-D?*	*C-M?*	*Linear?*	*Discrete?*
Utility + maximin – Utility threshold	no	yes	yes	yes
Utility + maximin – Equity threshold	yes	yes	yes	no
Utility + leximax – Predefined priorities	no	no	yes	yes
Utility + leximax – No predefined priorities	no	yes	yes	yes

Statistical fairness metrics

Criterion	*P-D?*	*C-M?*	*Linear?*	*Discrete?*
Demographic parity			yes	no
Equalized odds			yes	no
Accuracy parity			yes	no
Predictive rate parity			no	yes

P-D = Pigou-Dalton
C-M = Chateauneuf-Moyes
Linear = all constraints linear
Discrete = some variables discrete

Generic Model

- We formulate each fairness criterion as a **social welfare function (SWF)**.

Individual utilities

$$W(\boldsymbol{u}) = W(u_1, \ldots, u_n)$$

  - Measures desirability of the **magnitude and distribution of utilities** across individuals.
  - **Utility** can be wealth, health, negative cost, etc.
  - The **SWF** becomes the **objective function** of the optimization model.

Generic Model

The social welfare optimization problem

Vector of utilities enjoyed by individuals

Vector of resource allocations to individuals

$$\max_{\boldsymbol{u},\boldsymbol{x}} \left\{ W(\boldsymbol{u}) \;\middle|\; \boldsymbol{u} = U(\boldsymbol{x}),\; \boldsymbol{x} \in S_{\boldsymbol{x}} \right\}$$

Social welfare function

Vector of utility functions $U_1, \ldots, U_n$

Problem constraints

Generic Model

Example – *Medical triage*

QALYs without treatment

Additional QALYs due to treatment

= 1 if patient *i* is treated

Utility functions are $U_i(\boldsymbol{x}) = a_i + b_i x_i$.

$$\max_{\boldsymbol{u},\boldsymbol{x}} \left\{ W(\boldsymbol{u}) \;\middle|\; \begin{array}{l} u_i = a_i + b_i x_i,\; x_i \in \{0,1\},\; \text{all } i \\ \sum_i c_i x_i \le B \end{array} \right\}$$

Social welfare function

Budget constraint

Yes-or-no decision

Generic Model

The social welfare optimization problem

Incorporate $\boldsymbol{u} = U(\boldsymbol{x})$ into problem constraints.

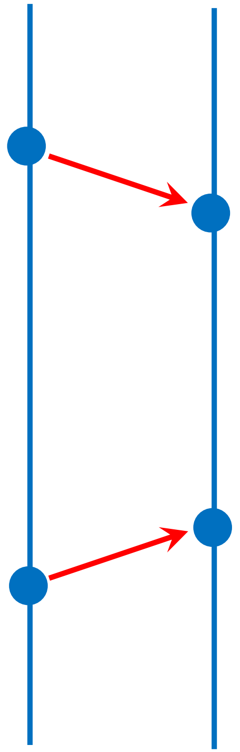
$$\max_{\boldsymbol{u},\boldsymbol{x}} \left\{ W(\boldsymbol{u}) \;\middle|\; (\boldsymbol{u}, \boldsymbol{x}) \in S \right\}$$

Social welfare function

Problem constraints

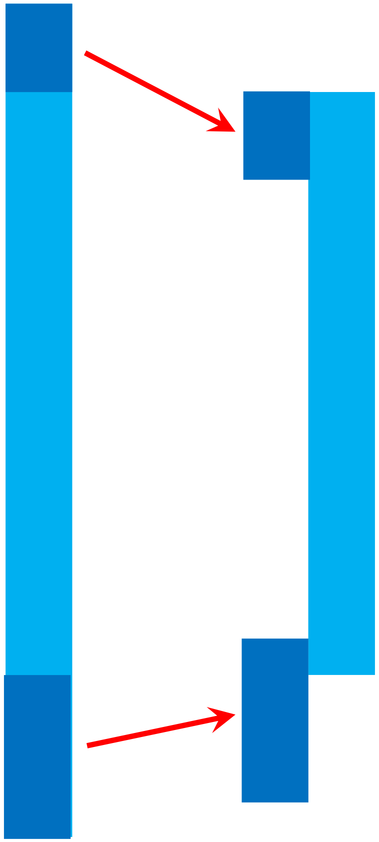
Pigou-Dalton Condition

- The Pigou-Dalton condition checks whether a SWF reflects **equality**.
    - A utility transfer from a **better-off** individual to a **worse-off** individual **never decreases** social welfare.
    - **Problem:** such a transfer can **increase inequality** with respect to some other individuals**.**



Chateauneuf-Moyes Condition

- Addresses weakness of Pigou-Dalton condition.
  - A utility transfer from **top of distribution** to **bottom of distribution** never decreases social welfare.
  - Loss/gain due to transfer is distributed equally in each class.



Chateauneuf & Moyes 2006

Inequality Measures

Criterion	*P-D?*	*C-M?*	*Linear?*	*Discrete?*
Relative range	yes	yes	yes	no
Relative mean deviation	yes	yes	yes	no
Coefficient of variation	yes	yes	no	no
Gini coefficient	yes	yes	yes	no
Hoover index	yes	yes	yes	no

Inequality Measures

Equality vs fairness

Two views on ethical importance of equality:

Parfit 1997

Scanlon 2003

- **Irreducible:** Inequality is inherently unfair.
- **Reducible:** Inequality is unfair only insofar as it reduces utility.

Frankfurt 2015

Possible problems with inequality measures:

- No preference for an identical distribution with **higher utility**.
- Even when average utility is fixed, no preference for reducing inequality at the **bottom** rather than the **top** of the distribution.

Inequality Measures

Equality vs fairness

We can perhaps agree on this much:

- Equality is **not the same concept** as fairness, even when it is closely related.
- An inequality metric can be appropriate when a specifically **egalitarian** distribution is the goal, **without regard** to efficiency and other forms of equity.

Inequality Measures

Relative range

$$W(\boldsymbol{u}) = -\frac{u_{\max} - u_{\min}}{\bar{u}}$$

Rationale:

- Perceived inequality is relative to the best off.
- So, move everyone closer to the best off.

Problem:

- Ignores distribution **between** extremes.

Inequality Measures

Relative range

- Problem is **linearized** using same change of variable as in linear-fractional programming.

Let $\boldsymbol{u} = \boldsymbol{u}'/t$ and $\boldsymbol{x} = \boldsymbol{x}'/t$. The optimization problem is

$$\min_{\substack{\boldsymbol{x}', \boldsymbol{u}', t \\ u'_{\min}, u'_{\max}}} \left\{ u'_{\max} - u'_{\min} \;\middle|\; \begin{array}{c} u'_{\min} \leq u'_i \leq u'_{\max}, \text{ all } i \\ \bar{u}' = 1, \; t \geq 0, \; (\boldsymbol{u}', \boldsymbol{x}') \in S' \end{array} \right\}$$

where $t, u'_{\min}, u'_{\max}$ are new variables.

Charnes & Cooper 1962

Inequality Measures

Relative range

Model:

$$\min_{\substack{\boldsymbol{x}',\boldsymbol{u}',t \\ u'_{\min},u'_{\max}}} \left\{ u'_{\max} - u'_{\min} \;\middle|\; \begin{array}{c} u'_{\min} \le u'_i \le u'_{\max},\ \text{all } i \\ \bar{u}' = 1,\ t \ge 0,\ (\boldsymbol{u}', \boldsymbol{x}') \in S' \end{array} \right\}$$

The difficulty of constraints $(\boldsymbol{u}', \boldsymbol{x}') \in S'$ depends on nature of S.

If S is linear $A\boldsymbol{u} + B\boldsymbol{x} \le \boldsymbol{b}$, it remains linear: $A\boldsymbol{u}' + B\boldsymbol{x}' \le t\boldsymbol{b}$.

If S is $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{x}) \le \boldsymbol{b}$ for homogeneous $\boldsymbol{g}$, it retains almost the same form: $\boldsymbol{g}(\boldsymbol{u}', \boldsymbol{x}') \le t\boldsymbol{b}$.

Inequality Measures

Linearity assumption

- From here out, we assume constraints $(\boldsymbol{u}, \boldsymbol{x}) \in S$ are **linear** when we describe the form of the optimization problem.

- This covers a wide variety of constraints.
- Convex feasible set can be approximated by piecewise linear constraints.

Inequality Measures

Relative mean deviation

$$W(\boldsymbol{u}) = -\frac{1}{\bar{u}} \sum_i |u_i - \bar{u}|$$

Rationale:

- Considers all utilities.

Model:

- Again, linearized by change of variable.

$$\min_{\boldsymbol{x}',\boldsymbol{u}',\boldsymbol{v},t} \left\{ \sum_i v_i \;\middle|\; \begin{array}{c} -v_i \leq u_i' - \bar{u}' \leq v_i, \text{ all } i \\ \bar{u}' = 1,\; t \geq 0,\; (\boldsymbol{u}', \boldsymbol{x}') \in S' \end{array} \right\}$$

where $\boldsymbol{v}$ is vector of new variables.

Inequality Measures

Coefficient of variation

$$W(\boldsymbol{u}) = -\frac{1}{\bar{u}}\left[\frac{1}{n}\sum_i (u_i - \bar{u})^2\right]^{\frac{1}{2}}$$

Rationale:

- Familiar. Outliers receive extra weight.

Problem:

- Nonlinear (but convex)

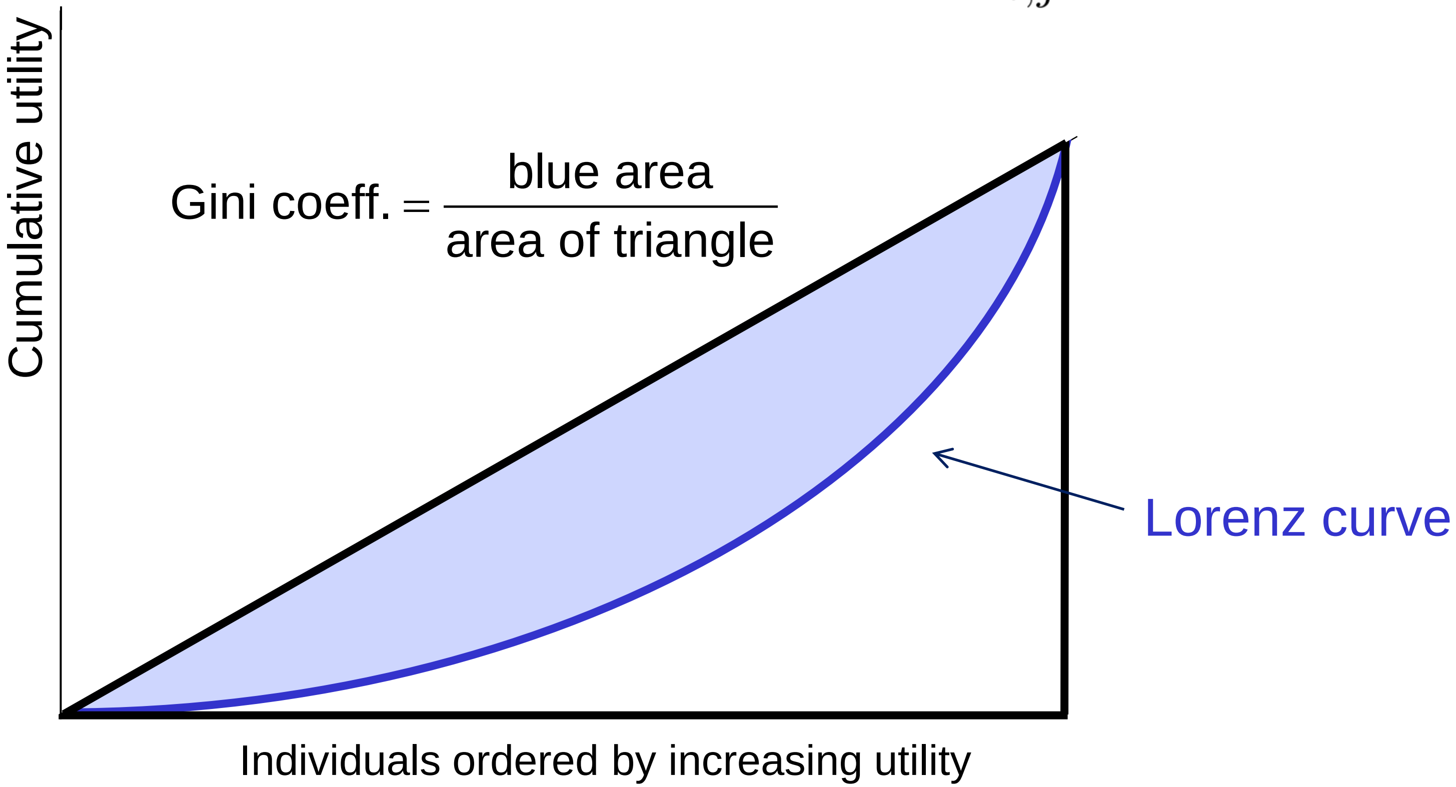
Model:

$$\min_{\boldsymbol{x}',\boldsymbol{u}',\boldsymbol{v},t}\left\{\frac{1}{n}\sum_i (u_i' - \bar{u}')^2 \;\middle|\; \begin{array}{c}\bar{u}' = 1,\; t \geq 0\\ (\boldsymbol{u}', \boldsymbol{x}') \in S'\end{array}\right\}$$

Inequality Measures

Gini coefficient

$$W(\boldsymbol{u}) = -G(\boldsymbol{u}), \quad \text{where } G(\boldsymbol{u}) = \frac{1}{2\bar{u}n^2} \sum_{i,j} |u_i - u_j|$$



Inequality Measures

Gini coefficient

$$W(\boldsymbol{u}) = -G(\boldsymbol{u}), \quad \text{where } G(\boldsymbol{u}) = \frac{1}{2\bar{u}n^2} \sum_{i,j} |u_i - u_j|$$

Rationale:

- Relationship to Lorenz curve.
- Widely used.

Model:

- Linear:

$$\min_{\boldsymbol{x}',\boldsymbol{u}',V,t} \left\{ \frac{1}{2n^2} \sum_{i,j} v_{ij} \;\middle|\; \begin{array}{c} -v_{ij} \le u_i' - u_j' \le v_{ij}, \text{ all } i,j \\ \bar{u}' = 1, \; t \ge 0, \; (\boldsymbol{u}', \boldsymbol{x}') \in S' \end{array} \right\}$$

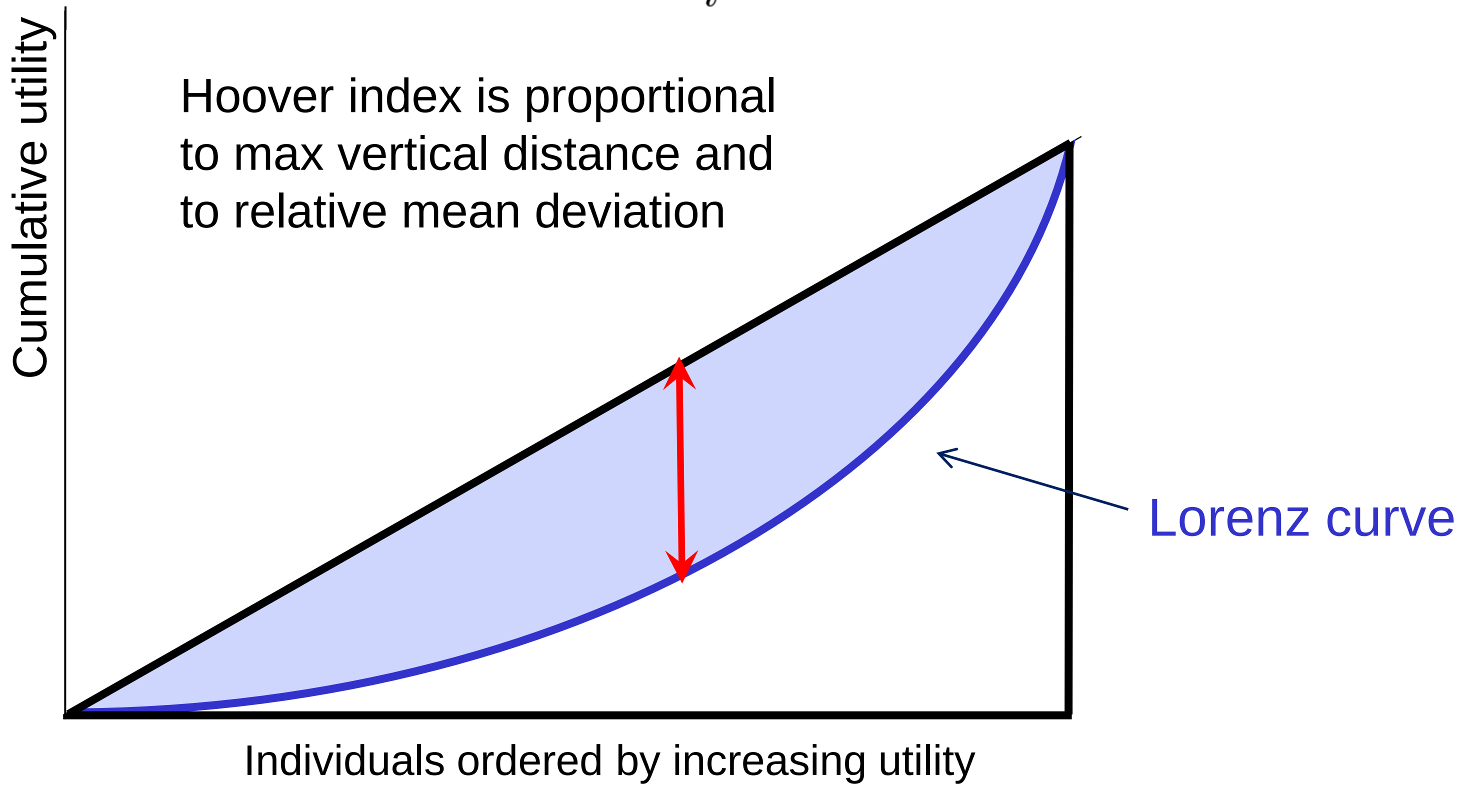
where V is a matrix of new variables.

Inequality Measures

Hoover index

$$W(\boldsymbol{u}) = -\frac{1}{2n\bar{u}} \sum_i |u_i - \bar{u}|$$

Hoover 1936



Inequality Measures

Hoover index

$$W(\boldsymbol{u}) = -\frac{1}{2n\bar{u}} \sum_i |u_i - \bar{u}|$$

Rationale:

- Hoover index is fraction of total utility that would have to be redistributed to achieve perfect equality.

Model:

- Same as relative mean deviation.

Fairness for the Disadvantaged

Criterion	*P-D?*	*C-M?*	*Linear?*	*Discrete?*
Maximin (Rawlsian)	yes	yes	yes	no
Leximax (lexicographic)	yes	yes	yes	no
McLoone index	no	yes	yes	yes

Fairness for the Disadvantaged

Maximin

$$W(\boldsymbol{u}) = \min_i \{u_i\}$$

Rationale:

- Based on **difference principle** of John Rawls.
- Inequality is justified only to the extent that it increases the utility of the worst-off.
- Originally intended only for the design of **social institutions** and distribution of **primary goods** (goods that any rational person would want).
- Can be adopted as a general principle of equity: maximize the minimum utility.

Rawls 1971, 1999

Fairness for the Disadvantaged

Maximin

$$W(\boldsymbol{u}) = \min_i \{u_i\}$$

Social contract argument:

- We decide on social policy in an "original position," behind a "veil of ignorance" as to our position on society.
- All parties must be willing to **endorse** the policy, no matter what position they end up assuming.
- No rational person can endorse a policy that puts him/her on the **bottom** of society – unless that person would be even **worse off** under another social arrangement.
- Therefore, an agreed-upon social policy must maximize the welfare of the worst-off.

Fairness for the Disadvantaged

Maximin

$$W(\boldsymbol{u}) = \min_i \{u_i\}$$

Model:

$$\max_{\boldsymbol{x},\boldsymbol{u},w} \left\{ w \;\middle|\; w \le u_i, \text{ all } i;\; (\boldsymbol{u},\boldsymbol{x}) \in S \right\}$$

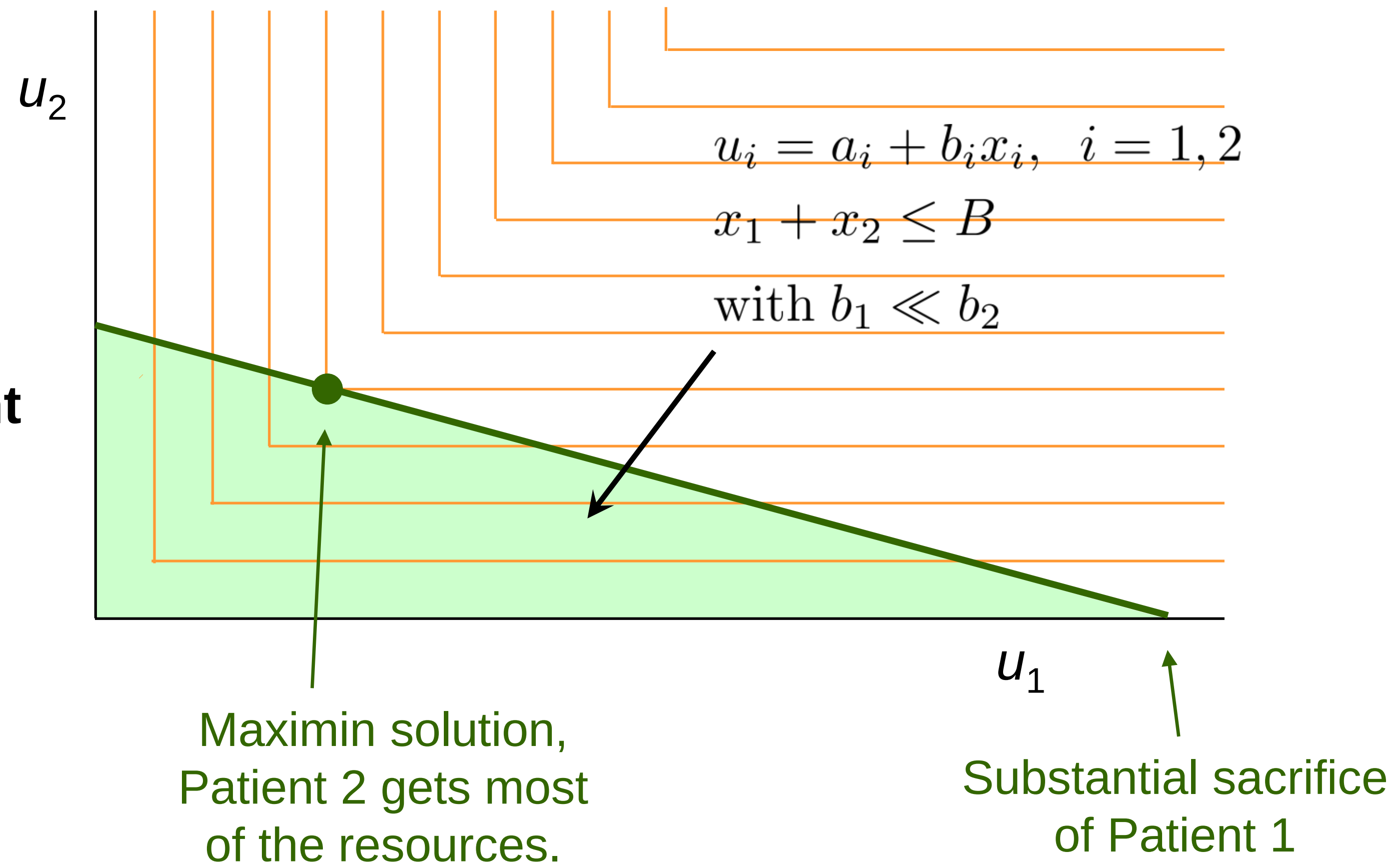
Problems:

- Can force equality even when this is extremely costly in terms of total utility.
- Does not care about 2nd worst off, etc., and so can waste resources.

Fairness for the Disadvantaged

Maximin

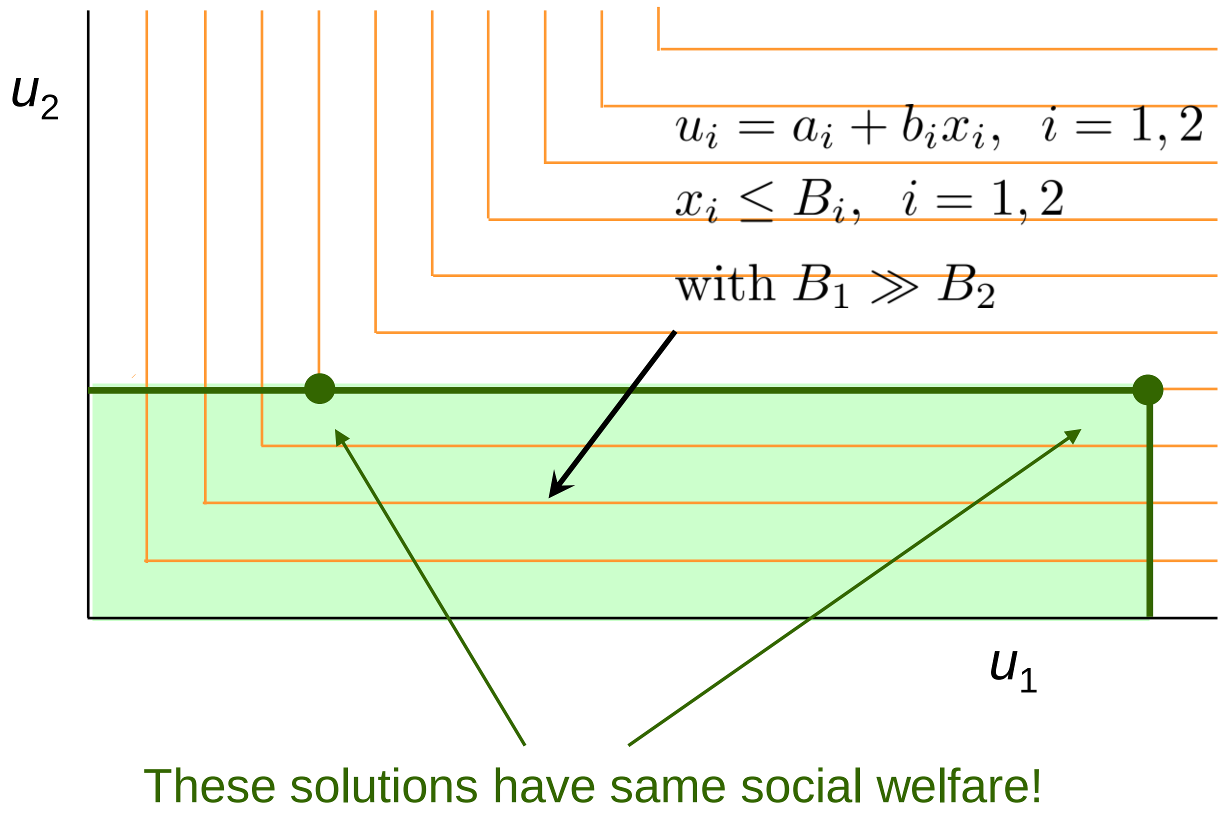
Medical example with budget constraint



Fairness for the Disadvantaged

Maximin

Medical example with resource bounds

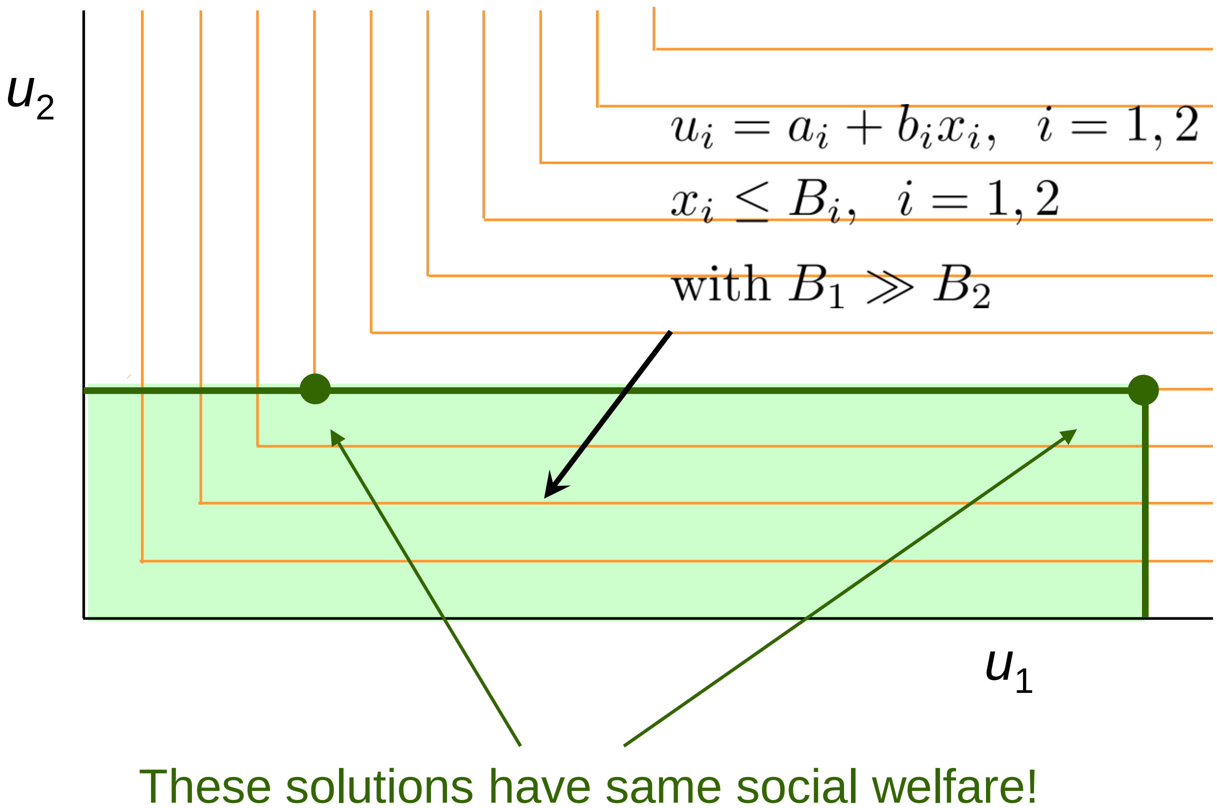


Fairness for the Disadvantaged

Maximin

Medical example with resource bounds

Remedy**:** use **leximax** solution



Fairness for the Disadvantaged

Leximax

Rationale:

- Takes in account 2nd worst-off, etc., and avoids wasting utility.
- Can be justified with Rawlsian argument.

Model:

Solve sequence of optimization problems

$$\max_{\boldsymbol{x},\boldsymbol{u},w} \left\{ w \;\middle|\; \begin{matrix} w \leq u_i, \; u_i \geq \hat{u}_{i_{k-1}}, \; i \in I_k \\ (\boldsymbol{u},\boldsymbol{x}) \in S \end{matrix} \right\}$$

for $k = 1, \ldots, n$, where i_k is defined so that $\hat{u}_{i_k} = \min_{i \in I_k}\{\hat{u}_i\}$, and where $I_k = \{1, \ldots, n\} \setminus \{i_1, \ldots, i_{k-1}\}$, $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{u}})$ is an optimal solution of problem k, and $\hat{u}_{i_0} = -\infty$.

If $\hat{u}_j = \min_{i \in I_k}\{\hat{u}_i\}$ for multiple j, **must enumerate all solutions** that result from breaking the tie.

Fairness for the Disadvantaged

McLoone index

$$W(\boldsymbol{u}) = \frac{1}{|I(\boldsymbol{u})|\tilde{u}} \sum_{i \in I(\boldsymbol{u})} u_i$$

where $\tilde{u}$ is the median of utilities in $\boldsymbol{u}$ and $I(\boldsymbol{u})$ is the set of indices of utilities at or below the median

Rationale:

- Compares total utility of those at or below the median to the utility that would result from bringing them up to the median.
- Index = 1 if no one is below median, $\rightarrow$ 0 for long lower tail.
- Focus on **all** the **disadvantaged**.
- Often used for public goods (e.g., educational benefits).
- Satisfies C-M condition, even though it violates P-D.

Fairness for the Disadvantaged

McLoone index

Model: Nonlinear, requires 0-1 variables.

$$\max_{\substack{\boldsymbol{x},\boldsymbol{u},m\\ \boldsymbol{y},\boldsymbol{z},\boldsymbol{\delta}}} \left\{ \frac{\sum_i y_i}{\sum_i z_i} \;\middle|\; \begin{array}{c} m - M\delta_i \le u_i \le m + M(1-\delta_i),\ \text{all } i \\ y_i \le u_i,\ y_i \le M\delta_i,\ \delta_i \in \{0,1\},\ \text{all } i \\ z_i \ge 0,\ z_i \ge m - M(1-\delta_i),\ \text{all } i \\ \sum_i \delta_i \le n/2,\ (\boldsymbol{u},\boldsymbol{x}) \in S \end{array} \right\}$$

Linearize with change of variable, obtain MILP.

$$\max_{\substack{\boldsymbol{x}',\boldsymbol{u}',m'\\ \boldsymbol{y}',\boldsymbol{z}',t,\boldsymbol{\delta}}} \left\{ \sum_i y_i' \;\middle|\; \begin{array}{c} u_i' \ge m' - M\delta_i, \text{all } i \\ u_i' \le m' + M(1-\delta_i),\ \text{all } i \\ y_i' \le u_i',\ y_i' \le M\delta_i,\ \delta_i \in \{0,1\},\ \text{all } i \\ z_i' \ge 0,\ z_i' \ge m' - M(1-\delta_i),\ \text{all } i \\ \sum_i z_i' = 1,\ t \ge 0 \\ \sum_i \delta_i \le n/2,\ (\boldsymbol{u}',\boldsymbol{x}') \in S' \end{array} \right\}$$

Social Choice Theory

- The economics literature derives social welfare functions from **axioms of rational choice**.
- The social welfare function depends on degree of **interpersonal comparability** of utilities.
- Arrow's impossibility theorem was the first result, but there are many others**.**

Social Choice Theory

Axioms

Anonymity (symmetry)

Social preferences are the same if indices of u_is are permuted.

Strict pareto

If $\boldsymbol{u} > \boldsymbol{u}'$, then $\boldsymbol{u}$ is preferred to $\boldsymbol{u}'$.

Independence

The preference of $\boldsymbol{u}$ over $\boldsymbol{u}'$ depends only on $\boldsymbol{u}$ and $\boldsymbol{u}'$ and not on what other utility vectors are possible.

Separability

Individuals i for which $u_i = u_i'$ do not affect the relative ranking of $\boldsymbol{u}$ and $\boldsymbol{u}'$.

Social Choice Theory

Interpersonal comparability

- The properties of social welfare functions that satisfy the axioms depend on the degree to which utilities can be **compared** across individuals.

Invariance transformations

- These are transformations of utility vectors that indicate the degree of interpersonal comparability.
- Applying an invariance transformation to utility vectors does not change the **ranking** of distributions.

An invariance transformation has the form $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)$, where ϕ_i is a transformation of individual utility i.

Social Choice Theory

Unit comparability.

- It is possible to compare utility **differences** across individuals.

$$u_i' - u_i > u_j' - u_j \text{ if and only if } \phi_i(u_i') - \phi_i(u_i) > \phi_j(u_j') - \phi_j(u_j)$$

Theorem. Given anonymity, strict pareto, and independence axioms, the social welfare criterion must be **utilitarian**.

$$W(\boldsymbol{u}) = \sum_i u_i$$

Level comparability.

- It is possible to compare utility **levels** across individuals.

$$u_i > u_j \text{ if and only if } \phi_i(u_i) > \phi_j(u_j)$$

Theorem. Given anonymity, strict pareto, independence, and separability axioms, the social welfare criterion must be **maximin or minimax**.

$$W(\boldsymbol{u}) = \min_i\{u_i\} \text{ or } W(\boldsymbol{u}) = -\max_i\{u_i\}$$

Social Choice Theory

Problem with the utilitarian proof.

- The proof assumes that utilities have **no more** than unit comparability.
- This immediately rules out a maximin criterion, since identifying the minimum utility presupposes that utility **levels** can be compared.

Problem with the maximin proof.

- The proof assumes that utilities have **no more** than level comparability.
- This immediately rules out criteria that consider the spread of utilities.
- So, it rules out all the criteria we consider after maximin.

Utility & Fairness – Convex Combinations

Criterion	*P-D?*	*C-M?*	*Linear?*	*Discrete?*
Utility + Gini coefficient	yes	yes	no	no
Utility * Gini coefficient	yes	yes	yes	no
Utility + maximin	yes	yes	yes	no

Utility & Fairness – Convex Combinations

Utility + Gini coefficient

$$W(\boldsymbol{u}) = (1-\lambda)\sum_i u_i + \lambda\big(1 - G(\boldsymbol{u})\big)$$

Rationale.

- Takes into account both efficiency and equity.
- Allows one to adjust their relative importance.

Problem.

- Combines utility with a dimensionless quantity.
- How to interpret λ, or choose a λ for a given application?
- Choice of λ is an issue with convex combinations in general.

Utility & Fairness – Convex Combinations

Utility * Gini coefficient

$$W(\boldsymbol{u}) = \big(1 - G(\boldsymbol{u})\big) \sum_i u_i$$

Rationale.

Eisenhandler & Tzur 2019

- Gets rid of λ.
- Equivalent to SWF that is easily linearized:

$$W(\boldsymbol{u}) = \sum_i u_i - \frac{1}{n} \sum_{i<j} |u_j - u_i|$$

Problem.

- It is still a convex combination of utility and an equality metric (mean absolute difference).
- Implicit multiplier λ = ½. Why this multiplier?

Utility & Fairness – Convex Combinations

Utility + Gini-weighted utility

$$W(\boldsymbol{u}) = \sum_i u_i + \mu\big(1 - G(\boldsymbol{u})\big) \sum_i u_i$$

Rationale.

- Combines quantities measured in same units.

Mostajabdaveh, Gutjahr, Salman 2019

Problem.

- Equivalent to utility*(1-Gini) with multiplier $\lambda = \mu\,(1 + 2\mu)^{-1}$.
- How to interpret μ?

Utility & Fairness – Convex Combinations

Utility + Maximin

$$W(\boldsymbol{u}) = (1-\lambda)\sum_i u_i + \lambda \min_i\{u_i\}$$

Rationale.

- Explicitly considers individuals other than worst off.

Problem.

- If u_k is smallest utility, this is simply the linear combination

$$W(\boldsymbol{u}) = u_k + (1-\lambda)\sum_{i \neq k} u_i$$

- How to interpret λ?

Utility & Fairness – Classical Methods

Criterion	*P-D?*	*C-M?*	*Linear?*	*Discrete?*
Alpha fairness	yes	yes	yes	no
Proportional fairness (Nash bargaining)	yes	yes	yes	no
Kalai-Smorodinsky bargaining	yes	yes	no	no

Alpha Fairness

$$W_\alpha(\boldsymbol{u}) = \begin{cases} \dfrac{1}{1-\alpha} \displaystyle\sum_i u_i^{1-\alpha} & \text{for } \alpha \geq 0, \ \alpha \neq 1 \\ \displaystyle\sum_i \log(u_i) & \text{for } \alpha = 1 \end{cases}$$

Mo & Walrand 2000; Verloop, Ayesta & Borst 2010

Rationale.

- Continuous and well-defined adjustment of equity/efficiency tradeoff.

  Utility u_j must be reduced by $(u_j/u_i)^\alpha$ units to compensate for a unit increase in u_i $(< u_j)$ while maintaining constant social welfare.

- Integral of power law $\Sigma_i u_i^{-\alpha}$
- Utilitarian when $\alpha = 0$, maximin when $\alpha \to \infty$
- Satisfies P-D (and therefore C-M).

Alpha Fairness

$$W_\alpha(\boldsymbol{u}) = \begin{cases} \dfrac{1}{1-\alpha} \displaystyle\sum_i u_i^{1-\alpha} & \text{for } \alpha \geq 0,\ \alpha \neq 1 \\ \displaystyle\sum_i \log(u_i) & \text{for } \alpha = 1 \end{cases}$$

Model

- Nonlinear but concave.

$$\max_{\boldsymbol{x},\boldsymbol{u}} \left\{ W_\alpha(\boldsymbol{u}) \;\middle|\; (\boldsymbol{u},\boldsymbol{x}) \in S \right\}$$

- Can be solved by efficient algorithms if constraints are linear (or perhaps if *S* is convex).

Alpha Fairness

$$W_\alpha(\boldsymbol{u}) = \begin{cases} \dfrac{1}{1-\alpha} \sum_i u_i^{1-\alpha} & \text{for } \alpha \geq 0,\ \alpha \neq 1 \\ \sum_i \log(u_i) & \text{for } \alpha = 1 \end{cases}$$

Possible problems

- Parameter α has no interpretation apart from the tradeoff rate.
- Unclear how to choose α in practice.
- An egalitarian distribution can have same social welfare as arbitrarily extreme inequality.

In a 2-person problem, the distribution $(u_1, u_2) = (1, 1)$ has the same social welfare as $(2^{1/(1-\alpha)}, \infty)$ when $\alpha > 1$.

Proportional Fairness

$$W(\boldsymbol{u}) = \sum_i \log(u_i)$$

Nash 1950

- Special case of alpha fairness ($\alpha = 1$).
- Also known as **Nash bargaining solution**, in which case bargaining starts with a default distribution ***d***.

$$W(\boldsymbol{u}) = \sum_i \log(u_i - d_i) \quad \text{or} \quad W(\boldsymbol{u}) = \prod_i (u_i - d_i)$$

Rationale

- Has nice geometric interpretation.
- Can be derived from axiomatic and bargaining arguments.
- Used in engineering applications (telecom, traffic signaling).

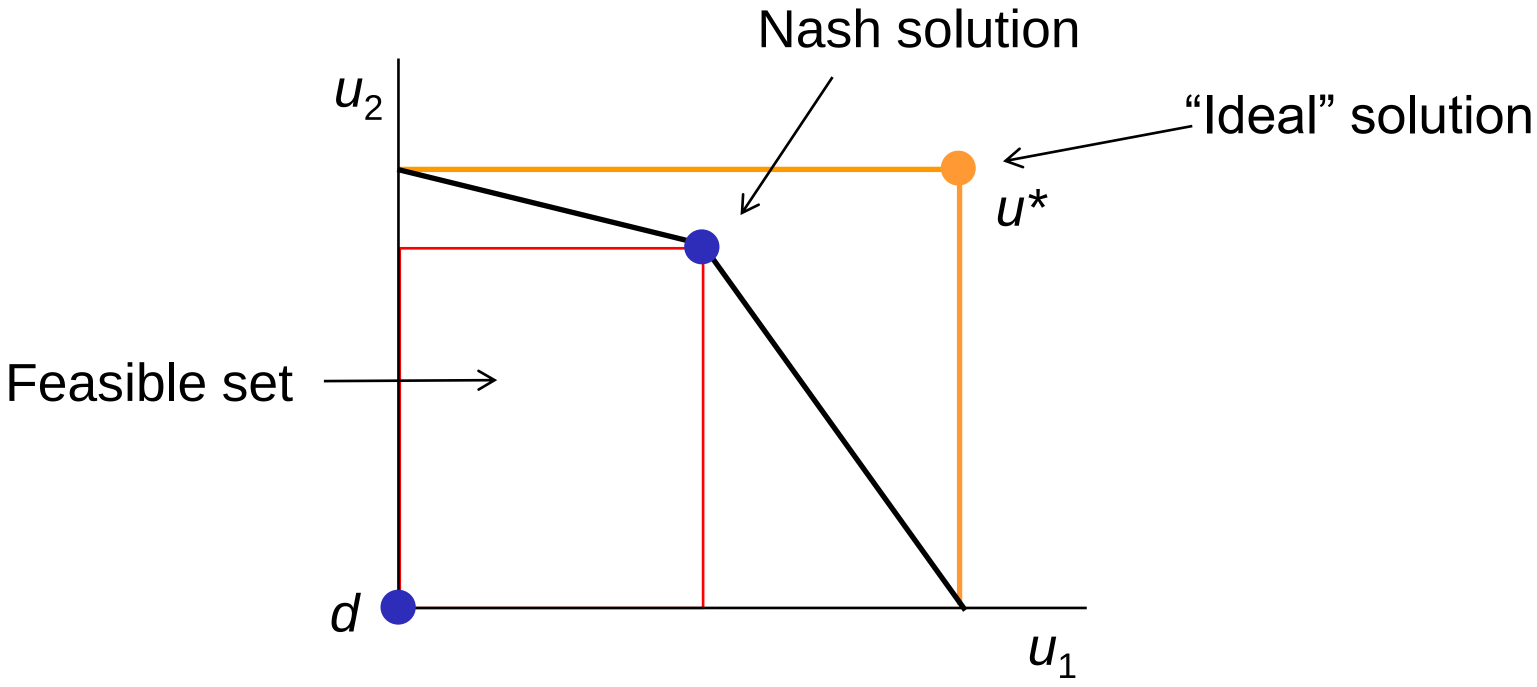
Proportional Fairness

Problems with bargaining justifications.

- Why should a bargaining procedure that is rational from an **individual** viewpoint result in a **just distribution?**
- Why should "**procedural justice**" = **justice**? For example, is the outcome of bargaining in a free market necessarily just?
- A deep question in political theory.

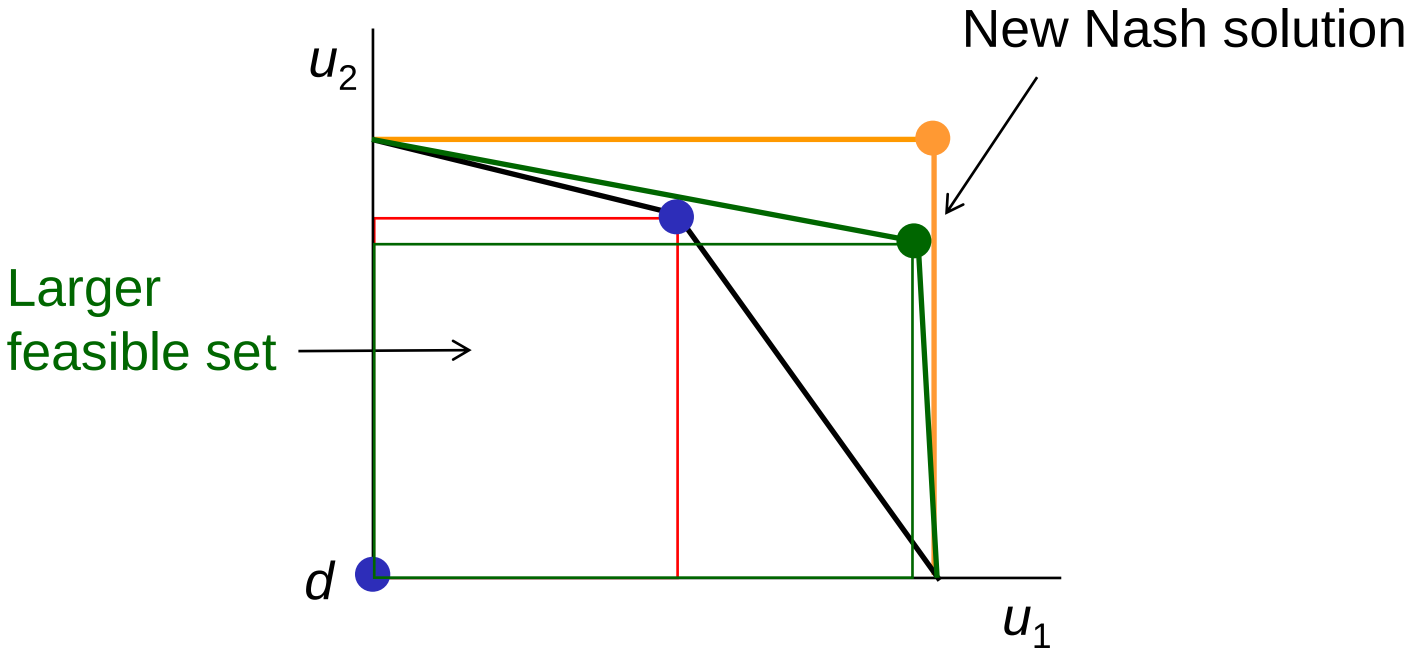
Kalai-Smorodinsky Bargaining

- Begins with a critique of the Nash bargaining solution.



Kalai-Smorodinsky Bargaining

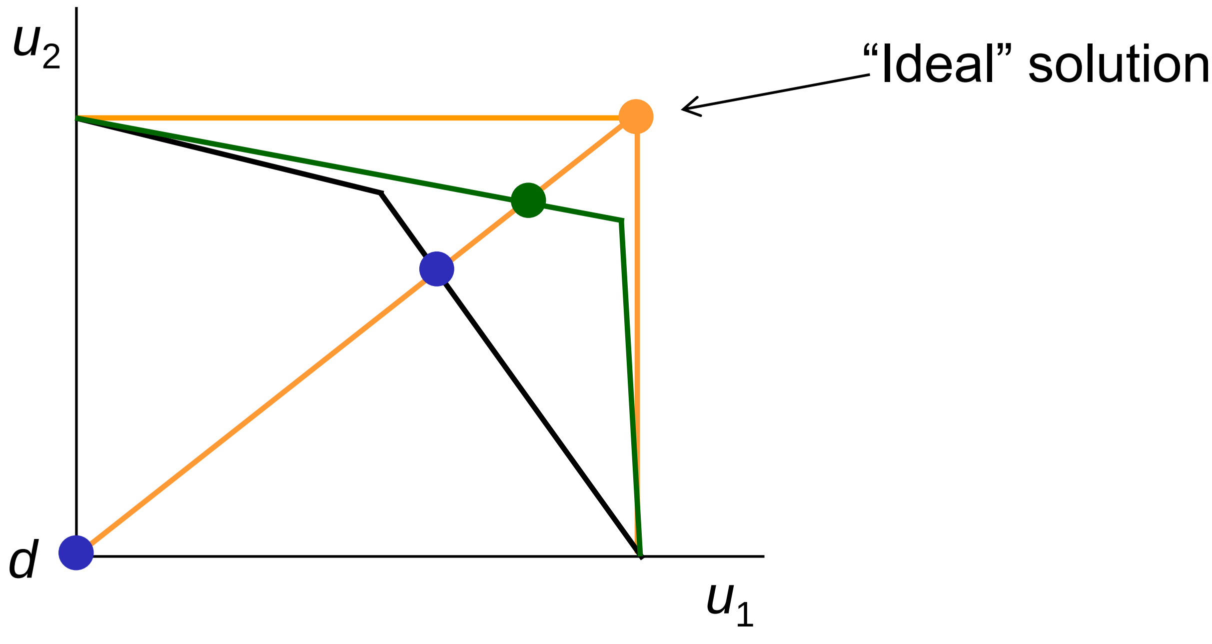
- Begins with a critique of the Nash bargaining solution.
- The new Nash solution is **worse** for player 2 even though the feasible set is **larger**.



Kalai-Smorodinsky Bargaining

- **Proposal**: Bargaining solution is pareto optimal point on line from *d* to ideal solution.

Kalai & Smorodinksy 1975



Kalai-Smorodinsky Bargaining

Social welfare function

$$W(\boldsymbol{u}) = \begin{cases} \sum_i u_i, & \text{if } \boldsymbol{u} = (1-\beta)\boldsymbol{d} + \beta \boldsymbol{u}^{\max} \text{ for some } \beta \text{ with } 0 \le \beta \le 1 \\ 0, & \text{otherwise} \end{cases}$$

where $u_i^{\max} = \max_{\boldsymbol{x},\boldsymbol{u}} \{u_i \mid (\boldsymbol{u},\boldsymbol{x}) \in S\}$.

Model

$$\max_{\beta,\boldsymbol{x},\boldsymbol{u}} \left\{ \beta \;\middle|\; \boldsymbol{u} = (1-\beta)\boldsymbol{d} + \beta \boldsymbol{u}^{\max},\ (\boldsymbol{u},\boldsymbol{x}) \in S,\ \beta \le 1 \right\}$$

Rationale

- Satisfies monotonicity.
- Seems reasonable for **price or wage negotiation**.
- Defended by some social contract theorists (e.g., “contractarians”)

Gauthier 1983, Thompson 1994

Kalai-Smorodinsky Bargaining

Possible problems

- Satisfies neither P-D nor C-M condition.
- In some contexts, it may not be ethical to allocate utility in proportion to one's potential.
- For example, when allocating resources to those with minor ailments vs chronic diseases.

Utility & Fairness – Threshold Methods

Criterion	*P-D?*	*C-M?*	*Linear?*	*Discrete?*
Utility + maximin – Utility threshold	no	yes	yes	yes
Utility + maximin – Equity threshold	yes	yes	yes	no
Utility + leximax – Predefined priorities	no	no	yes	yes
Utility + leximax – No predefined priorities	no	yes	yes	yes

Threshold Methods

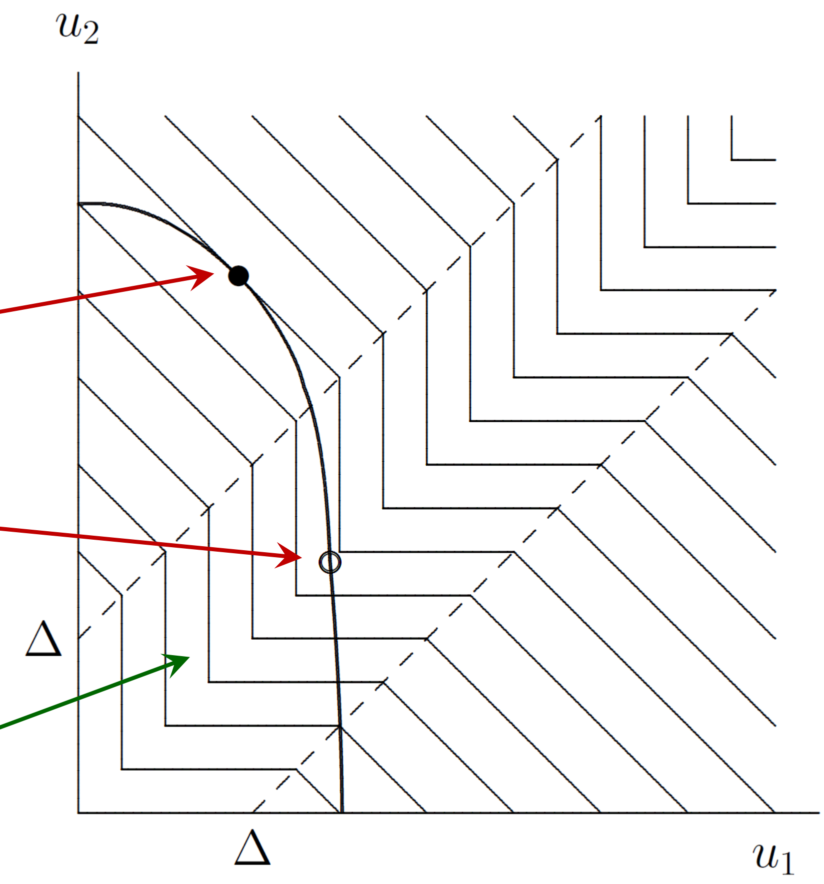
Combining utility and maximin

- **Utility threshold:** Use a maximin criterion until the utility cost becomes too great, then switch a utilitarian criterion.
- **Equity threshold:** Use a utilitarian criterion until the inequity becomes too great, then switch to a maximin criterion.

Williams & Cookson 2000

Threshold Methods

Utility threshold



Optimal solution

Maximin solution results in too much utility sacrifice for person 2

Feasible set

Williams & Cookson 2000

$$W(u_1, u_2) = \begin{cases} u_1 + u_2, & \text{if } |u_1 - u_2| \geq \Delta \\ 2\min\{u_1, u_2\} + \Delta, & \text{otherwise} \end{cases}$$

Threshold Methods

Utility threshold

Generalization to *n* persons

$$W(\boldsymbol{u}) = (n-1)\Delta + \sum_{i=1}^{n} \max\left\{u_i - \Delta, u_{\min}\right\}$$

$$\text{where } u_{\min} = \min_i\{u_i\}$$

JH & Williams 2012

Rationale

- Utilities within Δ of the lowest are in the **fair region**.
- Trade-off parameter Δ has a **practical interpretation**.
- Δ is chosen so that individuals in fair region are sufficiently deprived to **deserve priority**.
- Suitable when **equity** is the initial concern, but without paying **too high a cost** for fairness (healthcare, politically sensitive contexts).
- Δ = 0 corresponds to utilitarian criterion, Δ = ∞ to maximin.

Threshold Methods

Utility threshold

Model

$$\max_{\boldsymbol{x},\boldsymbol{u},\boldsymbol{\delta},\boldsymbol{v},w,z} \left\{ n\Delta + \sum_i v_i \;\middle|\; \begin{array}{c} u_i - \Delta \leq v_i \leq u_i - \Delta\delta_i, \text{ all } i \\ w \leq v_i \leq w + (M - \Delta)\delta_i, \text{ all } i \\ u_i - u_i \leq M, \text{ all } i, j \\ u_i \geq 0, \; \delta_i \in \{0,1\}, \text{all } i \\ (\boldsymbol{u}, \boldsymbol{x}) \in S \end{array} \right\}$$

- Tractable MILP model.
- Model is **sharp** without (***u***, ***x***) ∈ *S.*
- Easily generalized to differently-sized **groups** of individuals.

JH & Williams 2012

Problem

- Due to maximin component, many solutions with different equity properties have same social welfare value.

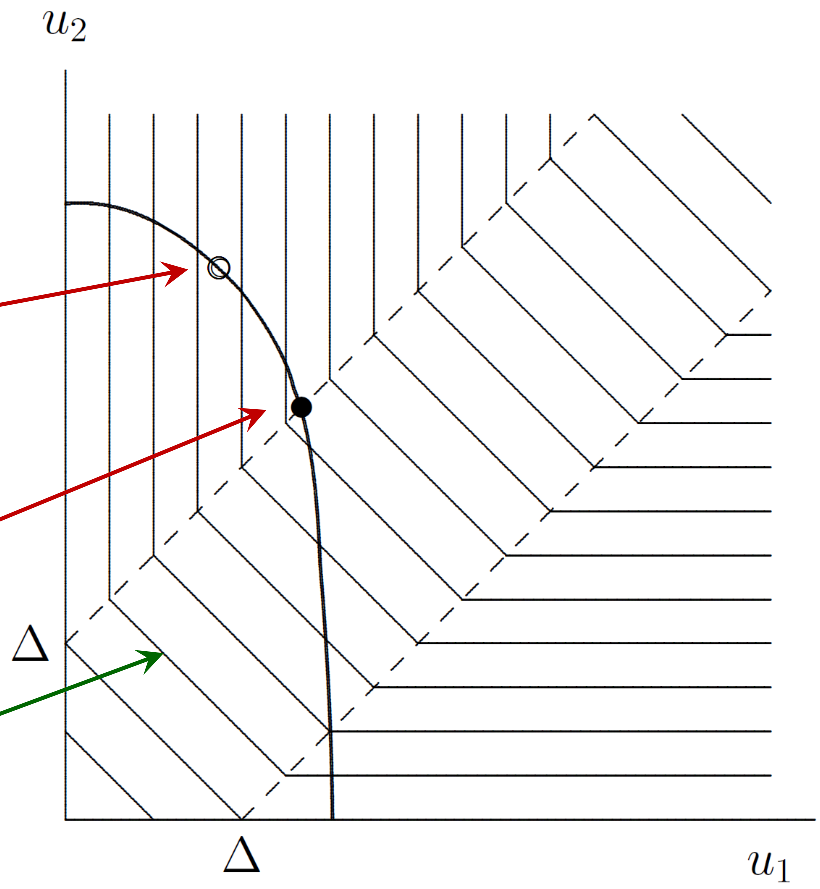
Threshold Methods

Equity threshold

Utilitarian solution leaves person 1 overly deprived

Optimal solution

Feasible set



Williams & Cookson 2000

$$W(u_1, u_2) = \begin{cases} 2\min\{u_1, u_2\} + \Delta, & \text{if } |u_1 - u_2| \geq \Delta \\ u_1 + u_2, & \text{otherwise} \end{cases}$$

Threshold Methods

Equity threshold

Generalization to *n* persons

$$W(\boldsymbol{u}) = n\Delta + \sum_{i=1}^{n} \min\{u_i - \Delta, u_{min}\}$$

Chen & JH 2021

Rationale

- Utilities more than Δ above the lowest are in the **fair region**.
- Trade-off parameter Δ has a **practical interpretation**.
- Δ is chosen so that well-off individuals (those in fair region) **do not deserve more utility** unless smaller utilities are also increased.
- Suitable when **efficiency** is the initial concern, but one does not want to create **excessive inequality** (traffic management, telecom, disaster recovery).

Threshold Methods

Equity threshold

Model

$$\max_{\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},w,z} \left\{ n\Delta + \sum_i v_i \;\middle|\; \begin{array}{c} v_i \leq w \leq u_i, \text{ all } i \\ v_i \leq u_i - \Delta, \text{ all } i \\ w \geq 0, v_i \geq 0, \text{ all } i \\ (\boldsymbol{u}, \boldsymbol{x}) \in S \end{array} \right\}$$

Chen & JH 2021

- Linear model.
- Easily generalized to differently-sized **groups** of individuals.

Problem

- As with threshold model, many solutions with different equity properties have same social welfare value.

Threshold Methods

Utility + leximax, predetermined preferences

$$W(\boldsymbol{u}) = \begin{cases} nu_1, & \text{if } |u_i - u_j| \le \Delta \text{ for all } i, j \\ \sum_i u_i + \text{sgn}(u_1 - u_i)\Delta, & \text{otherwise} \end{cases}$$

where preference order is $u_1, \ldots, u_n$.

McElfresh & Dickerson 2018

Rationale

- Takes into account utility levels of individuals in the fair region.
- Successfully applied to kidney exchange.

Threshold Methods

Utility + leximax, predetermined preferences

Model (MILP)

$$\max_{\substack{\boldsymbol{u},\boldsymbol{x}\\ w_1,w_2\\ \boldsymbol{y},\phi,\boldsymbol{\delta}}} \left\{ w_1 + w_2 \;\middle|\; \begin{array}{l} w_1 \le n u_1, \; w_1 \le M\phi \\ w_2 \le \sum_i (u_i + y_i), \; w_2 \le M(1-\phi) \\ u_i - u_j - \Delta \le M(1-\phi), \text{ all } i,j \\ y_i \le \Delta, \; y_i \le -\Delta + M\delta_i, \; u_i - u_1 \le M(1-\delta_i), \text{ all } i \\ (\boldsymbol{u},\boldsymbol{x}) \in S; \; \phi, \delta_i \in \{0,1\}, \text{ all } i \end{array} \right\}$$

where preference order is $u_1, \ldots, u_n$.

Also…

- The SWF combines utility and maximin.
- Leximax criterion applied only to optimal solutions of the SWF, and then only if some u_i's are in the fair region.

Threshold Methods

Utility + leximax, predetermined preferences

Possible problems

- SWF is discontinuous.
- SWF violates C-M and therefore P-D conditions.
- Preferences cannot be pre-ordered in many applications.
- Leximax is not incorporated in the SWF, but is applied only to SWF-maximizing solutions.

Threshold Methods

Utility + leximax, sequence of SWFs

SWFs $W_1, \ldots, W_n$ are maximized sequentially, where W_1 is the utility threshold SWF defined earlier, and W_k for $k \geq 2$ is

$$W_k(\boldsymbol{u}) = \sum_{i=1}^{k-1} (n-i+1) u_{\langle i \rangle} + (n-k+1) \min\left\{u_{\langle 1 \rangle} + \Delta, u_{\langle k \rangle}\right\} + \sum_{i=k}^{n} \max\left\{0,\ u_{\langle i \rangle} - u_{\langle 1 \rangle} - \Delta\right\}$$

where $u_{\langle 1 \rangle}, \ldots, u_{\langle n \rangle}$ are $u_1, \ldots, u_n$ in nondecreasing order.

Chen & JH 2021

Rationale

- Does not require pre-ordered preferences, satisfies C-M (not P-D).
- Tractable MILP models in practice, valid inequalities known.

Threshold Methods

Utility + leximax, sequence of SWFs

Model (MILP for W_k)

$$\max_{\substack{\boldsymbol{x},\boldsymbol{u},\boldsymbol{\delta},\boldsymbol{\epsilon}\\ \boldsymbol{v},w,\sigma,z}} \left\{ z \;\middle|\; \begin{aligned} & z \leq (n-k+1)\sigma + \sum_{i\in I_k} v_i \\ & 0 \leq v_i \leq M\delta_i, \;\; i \in I_k \\ & v_i \leq u_i - \bar{u}_{i_1} - \Delta + M(1-\delta_i), \;\; i \in I_k \\ & \sigma \leq \bar{u}_{i_1} + \Delta \\ & \sigma \leq w \\ & w \leq u_i, \; i \in I_k \\ & u_i \leq w + M(1-\epsilon_i), \; i \in I_k \\ & \sum_{i\in I_k} \epsilon_i = 1 \\ & w \geq \bar{u}_{i_{k-1}} \\ & u_i - \bar{u}_{i_1} \leq M, \; i \in I_k \\ & \delta_i, \epsilon_i \in \{0,1\}, \; i \in I_k \end{aligned} \right\}$$

where $\bar{u}_{i_k}$ is the value of the smallest utility in the optimal solution of the kth MILP model, and $I = \{1,\ldots,n\} \setminus \{i_1,\ldots,i_{k-1}\}$. The socially optimal solution is $(\bar{u}_1,\ldots,\bar{u}_n)$.

Threshold Methods – Healthcare Example

- Based on budget decisions in UK National Health Service
- Allocate limited treatment resources to disease/prognosis categories of patients.
- Based on cost, number of patients, and QALY estimates with and without treatment.*
- We will compare **2 utility-threshold SWFs**: utility + maximin and sequential utility + leximax.
- Solution time = fraction of second for each value of Δ.

Problem due to JH & Williams 2012

*QALY = quality adjusted life-year. Data reflect a particular situation and are not valid in general. Solutions presented here should not be taken as a general recommendation for healthcare resource allocation, but only as an illustration of social welfare functions.

QALY & cost data

Part 1

Intervention	Cost per person c_i (£)	QALYs gained q_i	Cost per QALY (£)	QALYs without intervention α_i	Subgroup size n_i
Pacemaker for atrioventricular heart block					
Subgroup A	3500	3	1167	13	35
Subgroup B	3500	5	700	10	45
Subgroup C	3500	10	350	5	35
Hip replacement					
Subgroup A	3000	2	1500	3	45
Subgroup B	3000	4	750	4	45
Subgroup C	3000	8	375	5	45
Valve replacement for aortic stenosis					
Subgroup A	4500	3	1500	2.5	20
Subgroup B	4500	5	900	3	20
Subgroup C	4500	10	450	3.5	20
CABG[1] for left main disease					
Mild angina	3000	1.25	2400	4.75	50
Moderate angina	3000	2.25	1333	3.75	55
Severe angina	3000	2.75	1091	3.25	60
CABG for triple vessel disease					
Mild angina	3000	0.5	6000	5.5	50
Moderate angina	3000	1.25	2400	4.75	55
Severe angina	3000	2.25	1333	3.75	60
CABG for double vessel disease					
Mild angina	3000	0.25	12,000	5.75	60
Moderate angina	3000	0.75	4000	5.25	65
Severe angina	3000	1.25	2400	4.75	70

QALY & cost data

Part 2

Intervention	Cost per person c_i (£)	QALYs gained q_i	Cost per QALY (£)	QALYs without intervention α_i	Subgroup size n_i
Heart transplant					
	22,500	4.5	5000	1.1	2
Kidney transplant					
Subgroup A	15,000	4	3750	1	8
Subgroup B	15,000	6	2500	1	8
Kidney dialysis					
Less than 1 year survival					
Subgroup A	5000	0.1	50,000	0.3	8
1-2 years survival					
Subgroup B	12,000	0.4	30,000	0.6	6
2-5 years survival					
Subgroup C	20,000	1.2	16,667	0.5	4
Subgroup D	28,000	1.7	16,471	0.7	4
Subgroup E	36,000	2.3	15,652	0.8	4
5-10 years survival					
Subgroup F	46,000	3.3	13,939	0.6	3
Subgroup G	56,000	3.9	14,359	0.8	2
Subgroup H	66,000	4.7	14,043	0.9	2
Subgroup I	77,000	5.4	14,259	1.1	2
At least 10 years survival					
Subgroup J	88,000	6.5	13,538	0.9	2
Subgroup K	100,000	7.4	13,514	1.0	1
Subgroup L	111,000	8.2	13,537	1.2	1

Threshold Methods – Healthcare Example

Budget constraint

$$\sum_j n_j c_j y_j \leq B$$

Size of treatment group *j*

Unit cost of treatment *j*

Fraction of group treated

Utility function

$$u_i = q_i y_i + \alpha_i$$

Treatment benefit (QALYs)

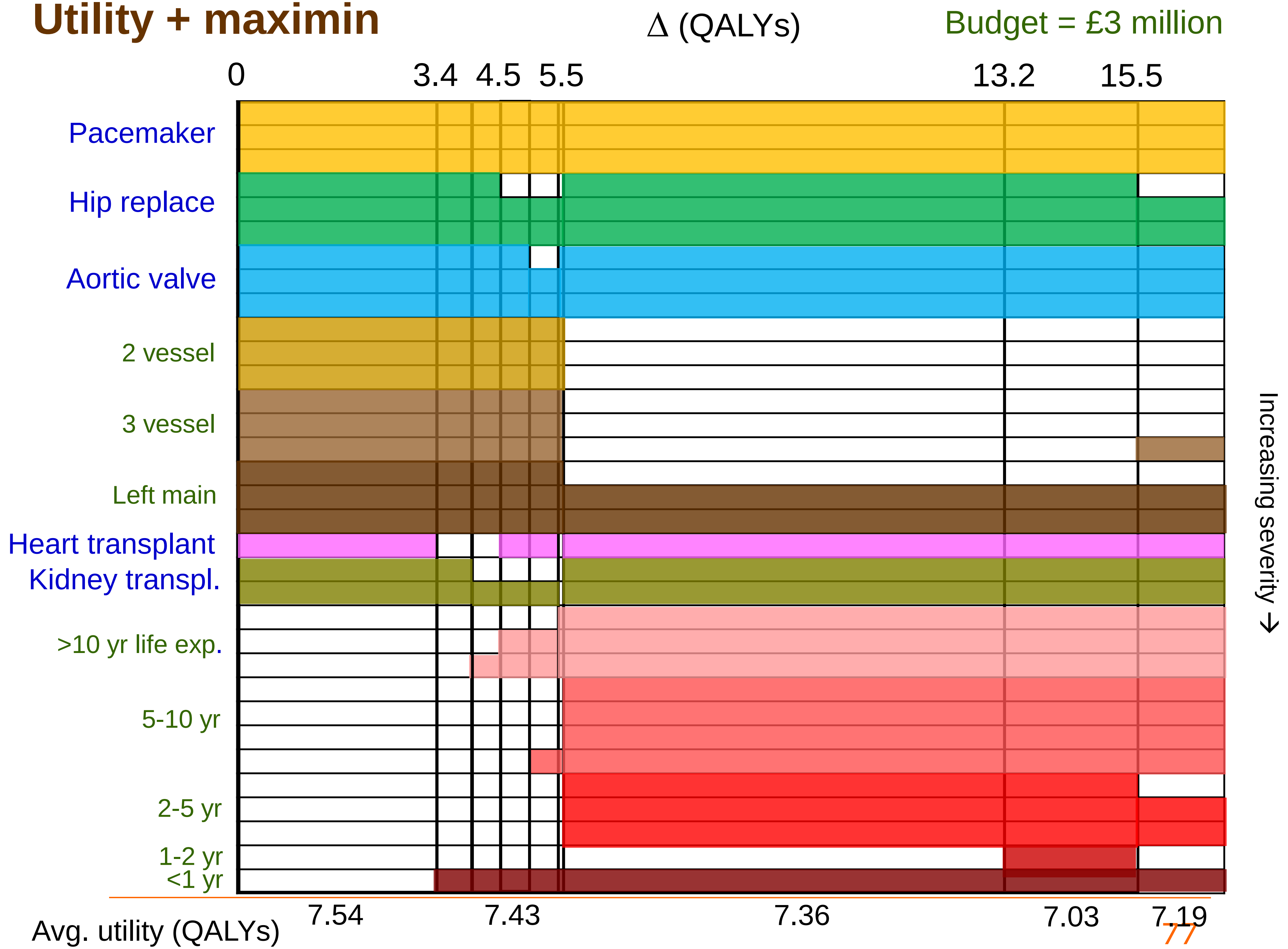
QALYs without treatment

which implies $y_i = (u_i - \alpha_i)/q_i$

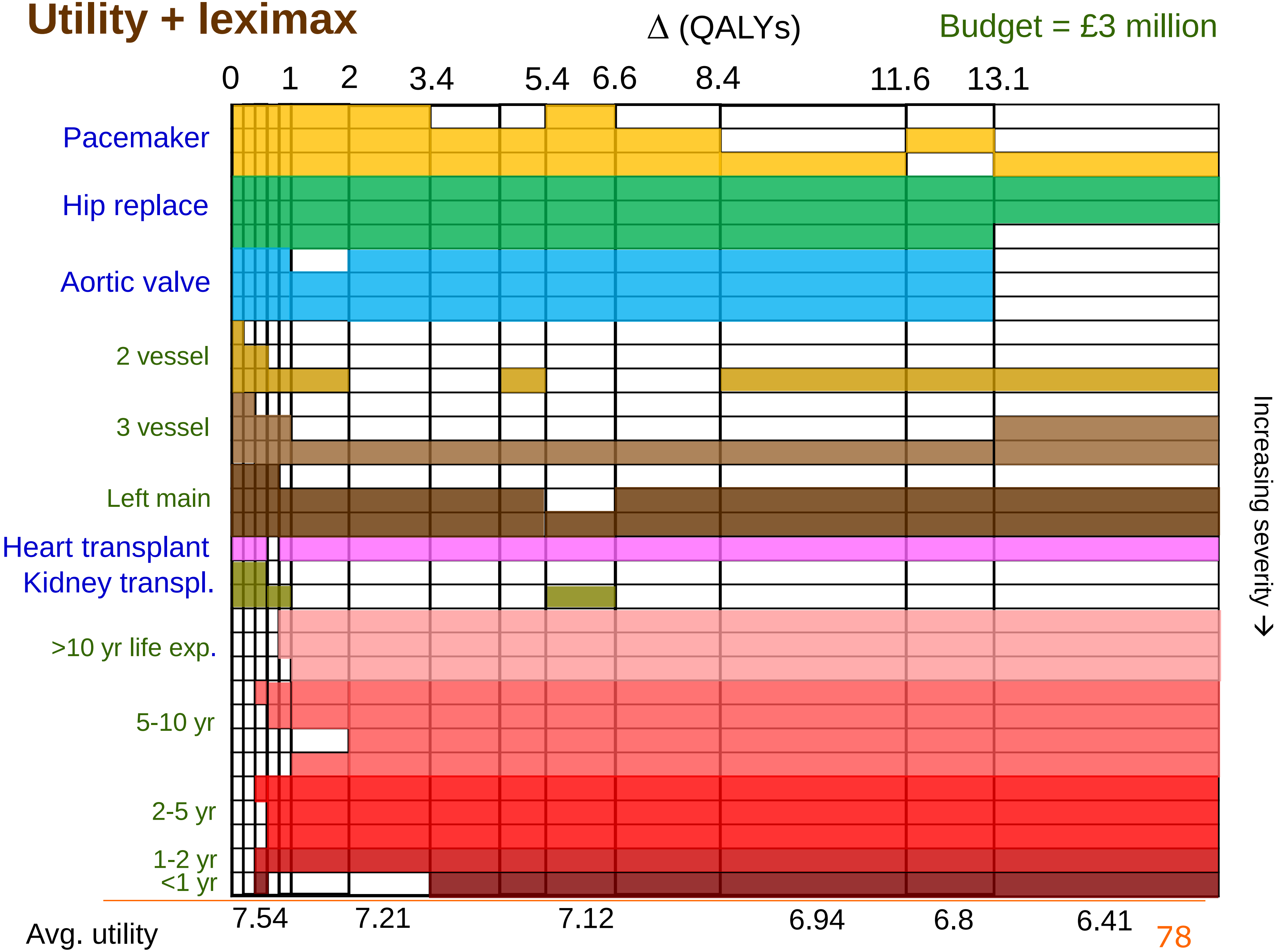
So the optimization problem becomes

$$\max_{\boldsymbol{u}} \left\{ W(\boldsymbol{u}) \;\middle|\; \sum_j \frac{n_j c_j}{q_j} u_j \leq B + \sum_j \frac{n_j c_j \alpha_j}{q_j}; \;\; \boldsymbol{\alpha} \leq \boldsymbol{u} \leq \boldsymbol{q} + \boldsymbol{\alpha} \right\}$$

Utility + maximin



Utility + leximax



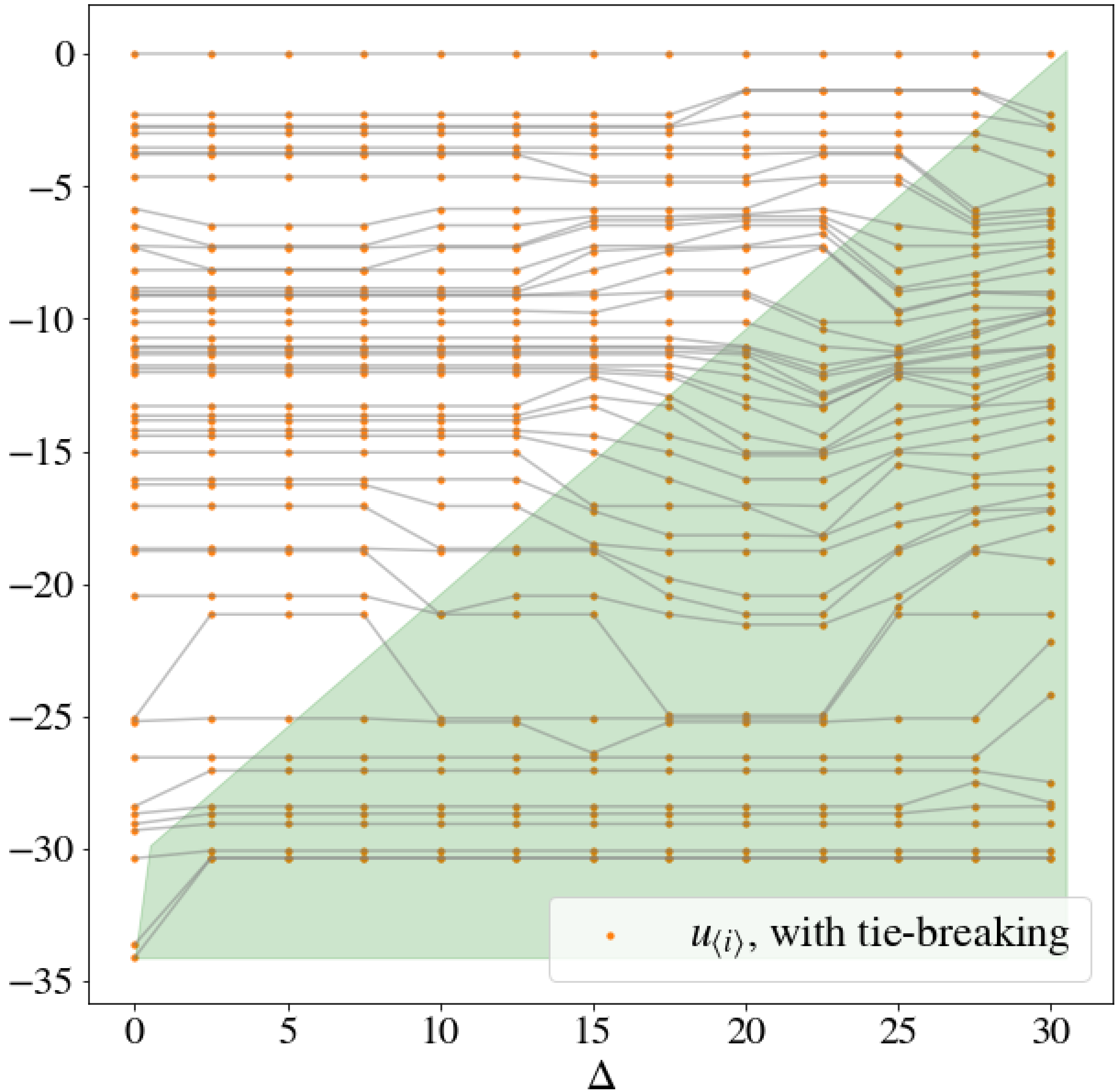
Threshold Methods – Disaster Preparedness Example

- Select earthquake shelter locations.
- Utility = negative distance of each neighborhood to nearest shelter, subject to limited budget.
- We will compare **2 utility-threshold SWFs**: utility + maximin and sequential utility + leximax.
- 50 neighborhoods, 50 potential shelter locations.
- Solution time = 1 to 18 seconds for each value of Δ.

Problem due to Mostajabdaveh, Gutjahr & Salman 2019

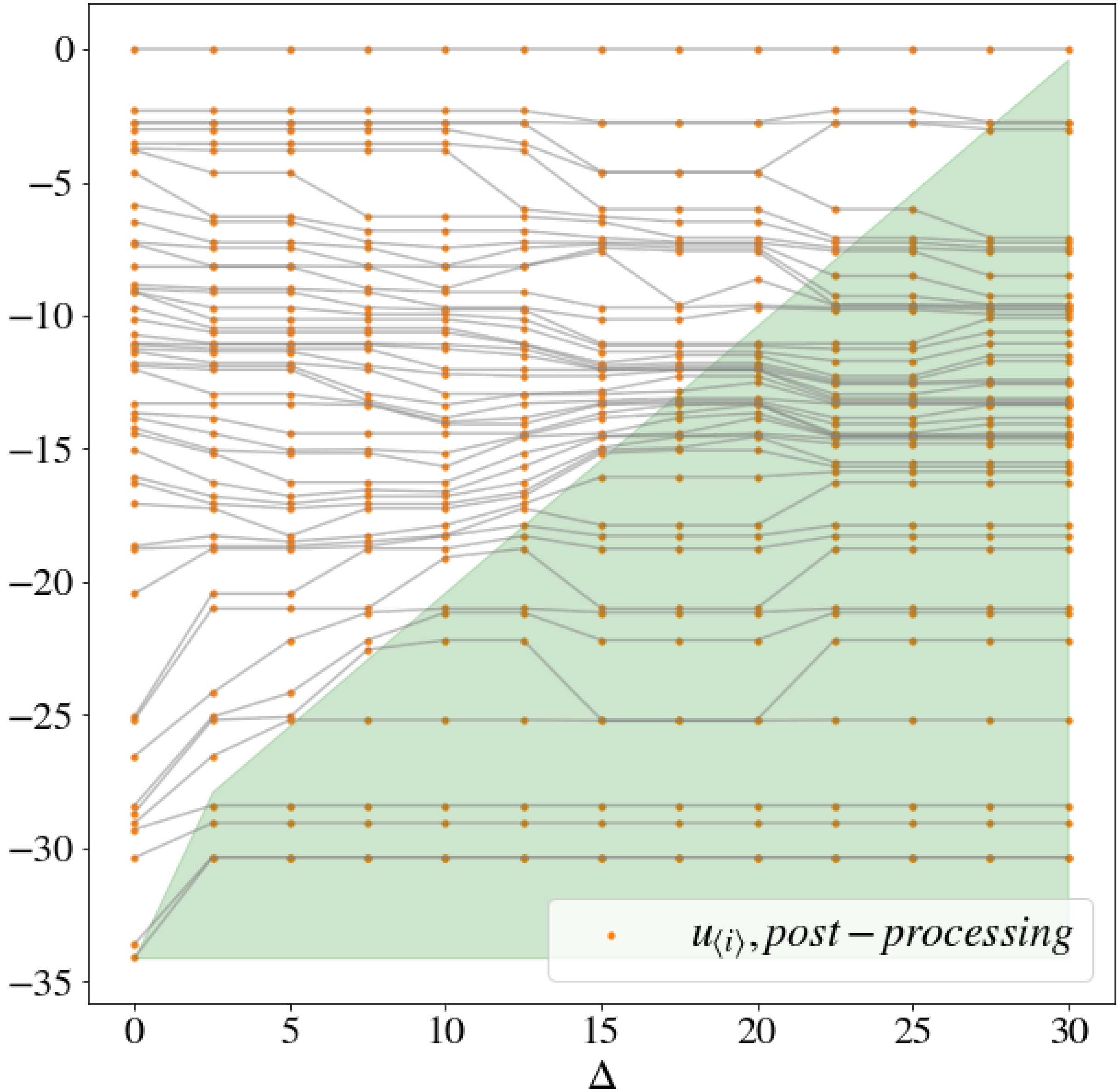
Threshold SWF

Utility + maximin



Threshold
SWF

Utility +
leximax



Statistical Fairness Metrics

Criterion	*P-D?*	*C-M?*	*Linear?*	*Discrete?*
Demographic parity			yes	no
Equalized odds			yes	no
Predictive rate parity			no	yes

Statistical Fairness Metrics

- Widely discussed in AI.
- Intended to measure bias against a subgroup.
- Most are based on statistical measures of classification error.
- Utility vector $\boldsymbol{u}$ is now vector δ of yes-no decisions.
- For example: mortgage loans, job interviews, parole.

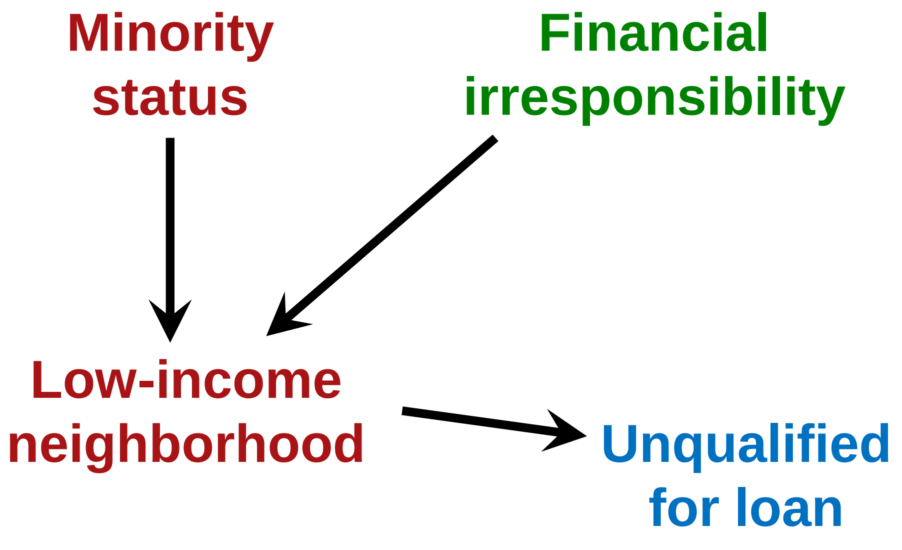
Rationale

- Unjustified bias against certain groups generally seen as inherently unfair.
- Bias may also incur legal problems.

Statistical Fairness Metrics

Example of implicit bias – Mortgage loans

- Financially irresponsible individuals may live in a **low-income neighborhood**.
- Members of a **minority group** may also live in the neighborhood due to historical discrimination.
- Minority status is **not** part of mortgage applicant's profile.
- But AI predictor sees the **correlation** between minority status and past defaults.
- Minority individual is **denied** a mortgage, even though financial irresponsibility is not the **cause** of past defaults.



Statistical Fairness Metrics

Notation

- TP = number of true positives (correct yes's)
- FP = number of false positives (incorrect yes's).
- TN = number of true negatives (correct no's).
- FN = number of false negatives (incorrect no's).

Basic model

- Maximize **accuracy**, perhaps $$\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$$

  …subject to a **bound** on a bias SWF.
- Bias measured by **comparing various statistics** across 2 groups (a protected group and everyone else).

Statistical Fairness Metrics

Demographic parity

- Compare $\frac{\mathrm{TP}+\mathrm{FP}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$ across 2 groups

$$W(\boldsymbol{\delta}) = 1 - |B(\boldsymbol{\delta})|, \text{ where } B(\boldsymbol{\delta}) = \frac{1}{|N|}\sum_{i\in N}\delta_i - \frac{1}{|N'|}\sum_{i\in N'}\delta_i$$

Majority group

Minority group

Rationale

- Equality of outcomes.

Possible problem

- Can discriminate against a minority group that is more qualified than majority group.

Dwork et al. 2012

Statistical Fairness Metrics

Equalized odds

Equality of opportunity

- Compare $\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$ and $\frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}}$ across 2 groups

$$B(\boldsymbol{\delta}) = \frac{\sum_{i\in N} a_i\delta_i}{\sum_{i\in N} a_i} - \frac{\sum_{i\in N'} a_i\delta_i}{\sum_{i\in N'} a_i}$$ and $$B(\boldsymbol{\delta}) = \frac{\sum_{i\in N}(1-a_i)\delta_i}{\sum_{i\in N}(1-a_i)} - \frac{\sum_{i\in N'}(1-a_i)\delta_i}{\sum_{i\in N'}(1-a_i)}$$

Rationale

- Compares fraction of **qualified** (or unqualified) persons selected.

Possible problem

- Considers only **yes** (or only **no**) decisions.

Hardt et al. 2016

Statistical Fairness Metrics

Predictive rate parity

- Compare $\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$ across 2 groups.

$$B(\boldsymbol{\delta}) = \frac{\sum_{i \in N} a_i \delta_i}{\sum_{\imath \in N} \delta_i} - \frac{\sum_{i \in N'} a_i \delta_i}{\sum_{\imath \in N'} \delta_i}$$

Rationale

- Compares what fraction of selected individuals **should** have been selected.

Dieterich et al. 2016

Problem

- Poses very difficult nonconvex discrete optimization problem.
- Unclear what justifies the computational burden.

Statistical Fairness Metrics

Matthews correlation coefficient

Rationale

- Most comprehensive measure of classification accuracy.

Problem

- Poses intractable nonconvex, discrete optimization problem.

Matthews 1975,
Chicco & Jurman 2020

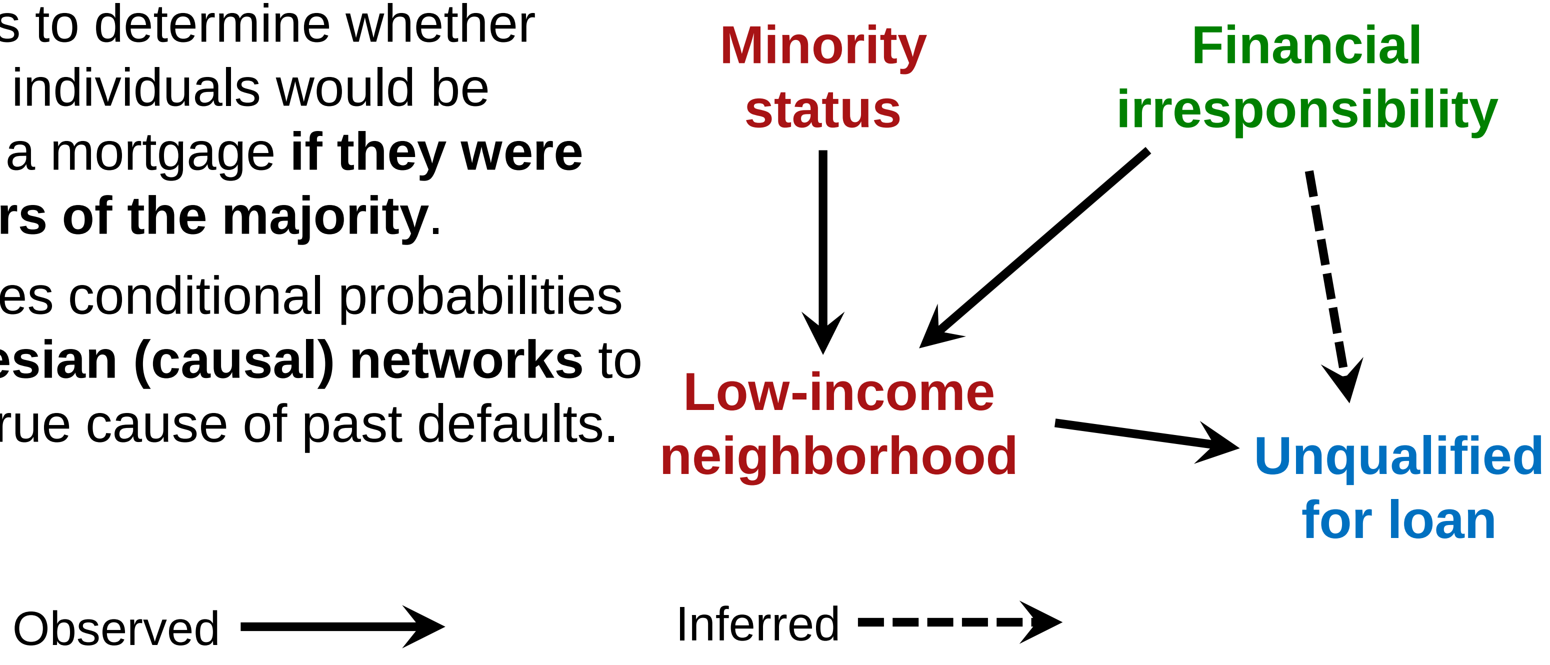
Statistical Fairness Metrics

Counterfactual fairness

Kusner et al. 2017, Russell et al. 2017

Rationale

- Attempts to determine whether minority individuals would be granted a mortgage **if they were members of the majority**.
- Computes conditional probabilities on **Bayesian (causal) networks** to isolate true cause of past defaults.



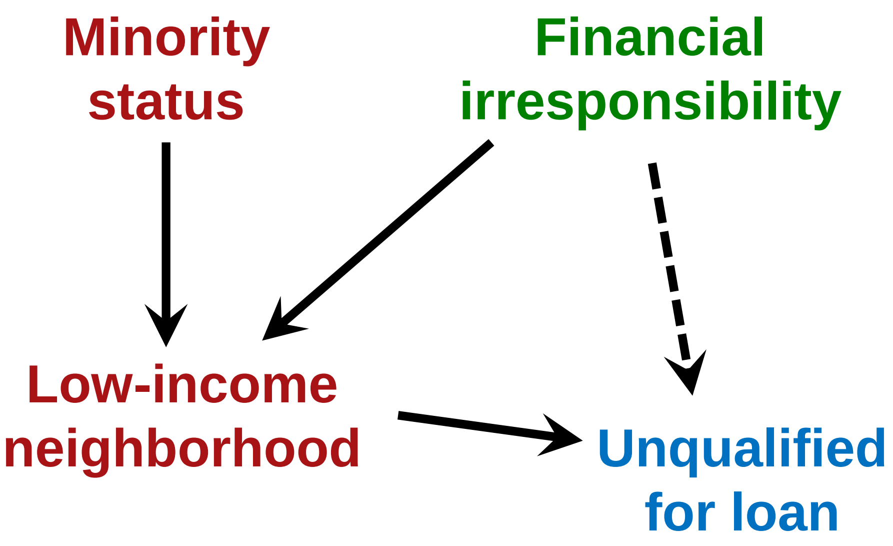
Statistical Fairness Metrics

Counterfactual fairness

Kusner et al. 2017, Russell et al. 2017

Problems

- Hard to find enough data to calibrate a large Bayesian network.
- Unclear how to formulate this in an optimization model.



Statistical Fairness Metrics

General problems of fairness metrics

- Yes-no outcomes (δ) provide a **limited perspective** on utility consequences (***u***)***.***
- No consensus on **which bias metric** $B(\delta)$, if any, is suitable for a given context. Bias metrics were developed to measure predictive accuracy, not fairness.
- No principle for **balancing** equity and efficiency.
- No clear principle for **selecting protected groups** (*N*), unless one simply selects those protected by law.