Measuring Diversity of Preferences

Vahid M. Hashemi
joint work with Ulle Endriss

ILLC, University of Amsterdam

March 20, 2015

Introduction

- Real world vs. synthetic preference profiles
- Diverse vs. consensus preferences
  - less diverse: better behavior?
    - fewer paradoxes
    - easier to reach an agreement
    - less disappointment

Example

Which one is more diverse?

	$2 : a \succ b \succ c$ $2 : b \succ c \succ a$ $2 : c \succ a \succ b$	$3 : a \succ b \succ c$ $3 : c \succ b \succ a$	$1 : a \succ b \succ c$ $1 : a \succ c \succ b$ $1 : b \succ a \succ c$ $1 : b \succ c \succ a$ $1 : c \succ a \succ b$ $1 : c \succ b \succ a$	$2 : a \succ b \succ c$ $2 : b \succ a \succ c$ $2 : a \succ c \succ b$

Example

Which one is more diverse?

	$2: a \succ b \succ c$ $2: b \succ c \succ a$ $2: c \succ a \succ b$	$3: a \succ b \succ c$ $3: c \succ b \succ a$	$1: a \succ b \succ c$ $1: a \succ c \succ b$ $1: b \succ a \succ c$ $1: b \succ c \succ a$ $1: c \succ a \succ b$ $1: c \succ b \succ a$	$2: a \succ b \succ c$ $2: b \succ a \succ c$ $2: a \succ c \succ b$
	3	2	6	3

Example

Which one is more diverse?

	$2 : a \succ b \succ c$ $2 : b \succ c \succ a$ $2 : c \succ a \succ b$	$3 : a \succ b \succ c$ $3 : c \succ b \succ a$	$1 : a \succ b \succ c$ $1 : a \succ c \succ b$ $1 : b \succ a \succ c$ $1 : b \succ c \succ a$ $1 : c \succ a \succ b$ $1 : c \succ b \succ a$	$2 : a \succ b \succ c$ $2 : b \succ a \succ c$ $2 : a \succ c \succ b$
	3	2	6	3
	6	6	6	5

Example

Which one is more diverse?

	$2 : a \succ b \succ c$ $2 : b \succ c \succ a$ $2 : c \succ a \succ b$	$3 : a \succ b \succ c$ $3 : c \succ b \succ a$	$1 : a \succ b \succ c$ $1 : a \succ c \succ b$ $1 : b \succ a \succ c$ $1 : b \succ c \succ a$ $1 : c \succ a \succ b$ $1 : c \succ b \succ a$	$2 : a \succ b \succ c$ $2 : b \succ a \succ c$ $2 : a \succ c \succ b$
	3	2	6	3
	6	6	6	5
	12	9	15	12

Example

Which one is more diverse?

	$2 : a \succ b \succ c$ $2 : b \succ c \succ a$ $2 : c \succ a \succ b$	$3 : a \succ b \succ c$ $3 : c \succ b \succ a$	$1 : a \succ b \succ c$ $1 : a \succ c \succ b$ $1 : b \succ a \succ c$ $1 : b \succ c \succ a$ $1 : c \succ a \succ b$ $1 : c \succ b \succ a$	$2 : a \succ b \succ c$ $2 : b \succ a \succ c$ $2 : a \succ c \succ b$
	3	2	6	3
	6	6	6	5
	12	9	15	12
	2	3	3	2

Example

Which one is more diverse?

	$2 : a \succ b \succ c$ $2 : b \succ c \succ a$ $2 : c \succ a \succ b$	$3 : a \succ b \succ c$ $3 : c \succ b \succ a$	$1 : a \succ b \succ c$ $1 : a \succ c \succ b$ $1 : b \succ a \succ c$ $1 : b \succ c \succ a$ $1 : c \succ a \succ b$ $1 : c \succ b \succ a$	$2 : a \succ b \succ c$ $2 : b \succ a \succ c$ $2 : a \succ c \succ b$
	3	2	6	3
	6	6	6	5
	12	9	15	12
	2	3	3	2
	$4(2+2+2) = 24$	$9 * 3 = 27$	$\frac{6}{2}(1+1+2+2+3) = 27$	$4(1+1+2) = 16$

Outline

Basic Definitions

Individuals $\mathcal{N} = \{1, 2, \ldots, n\}$, finite set of n individuals (voters)

Alternatives $\mathcal{X} = \{x_1, \ldots, x_m\}$, finite set of m alternatives (candidates)

Preferences Members of $\mathcal{L}(\mathcal{X})$ (the set of strict linear orders over $\mathcal{X}$)

Profile $\boldsymbol{R} = (R_1, \ldots, R_n) \in \mathcal{L}(\mathcal{X})^n$, vector of preference orders

Example

For $\mathcal{X} = \{a, b, c\}$ and 5 voters, a possible profile is:

$$\boldsymbol{R} = (abc, abc, acb, cab, cba)$$

PDO & PDI

Definition (Preference diversity index)

A **preference diversity index** (PDI) is a function $\Delta : \mathcal{L}(\mathcal{X})^n \to \mathbb{R}^+ \cup \{0\}$, mapping profiles to the nonnegative reals, that respects $\Delta(R, \ldots, R) = 0$ for any $R \in \mathcal{L}(\mathcal{X})$.

A PDI Δ is *normalised* if it maps any given profile to the interval $[0, 1]$, and the maximum of 1 is reached for at least one profile, i.e., $\max\{\Delta(\boldsymbol{R}) \mid \boldsymbol{R} \in \mathcal{L}(\mathcal{X})^n\} = 1$.

Definition (Preference diversity order)

A **preference diversity order** (PDO) is a weak order $\succcurlyeq$ declared on the space of preference profiles $\mathcal{L}(\mathcal{X})^n$ that respects $\boldsymbol{R} \succcurlyeq (R, \ldots, R)$ for all $\boldsymbol{R} \in \mathcal{L}(\mathcal{X})^n$ and all $R \in \mathcal{L}(\mathcal{X})$.

PDO & PDI

Definition (Preference diversity index)

A **preference diversity index** (PDI) is a function $\Delta : \mathcal{L}(\mathcal{X})^n \to \mathbb{R}^+ \cup \{0\}$, mapping profiles to the nonnegative reals, that respects $\Delta(R, \ldots, R) = 0$ for any $R \in \mathcal{L}(\mathcal{X})$.

A PDI Δ is *normalised* if it maps any given profile to the interval $[0, 1]$, and the maximum of 1 is reached for at least one profile, i.e., $\max\{\Delta(\boldsymbol{R}) \mid \boldsymbol{R} \in \mathcal{L}(\mathcal{X})^n\} = 1$.

Definition (Preference diversity order)

A **preference diversity order** (PDO) is a weak order $\succcurlyeq$ declared on the space of preference profiles $\mathcal{L}(\mathcal{X})^n$ that respects $\boldsymbol{R} \succcurlyeq (R, \ldots, R)$ for all $\boldsymbol{R} \in \mathcal{L}(\mathcal{X})^n$ and all $R \in \mathcal{L}(\mathcal{X})$.

Specific preference diversity indices

Definition (support-based PDI)

$\Delta^{\ell=k}_{\textit{supp}}(\boldsymbol{R})$: number of ordered k-tuples of alternatives occurring in at least one individual preference in profile $\boldsymbol{R}$.

$\Delta^{\ell=m}_{\textit{supp}}(\boldsymbol{R})$: *simple support-based PDI*, counts number of different preferences in $\boldsymbol{R}$.

Definition (distance-based PDI)

$\Delta^{\Phi,\delta}_{\textit{dist}}(\boldsymbol{R})$: aggregated (e.g., $\Phi = \Sigma$) distance (δ) between all pairs of individual preferences in profile $\boldsymbol{R}$.

Kendall tau distance: $K(R, R') = \frac{1}{2} \cdot |\{(x, y) \mid xRy \text{ and } yR'x\}|$

Definition (compromise-based PDI)

$\Delta^{\Phi,F}_{\textit{com}}(\boldsymbol{R})$: aggregated (e.g., $\Phi = \Sigma$) Kendall tau distance of individual preferences in $\boldsymbol{R}$ to a compromise preference $F(\boldsymbol{R})$ (e.g., F = Borda rule).

Example

$\Delta^{\ell=m}_{\textit{supp}}(abc, abc, acb, cab, cba) = 4$

$\Delta^{\Sigma,K}_{\textit{dist}}(abc, abc, acb, cab, cba) = 0 + 1 + 2 + 3 + 1 + 2 + 3 + 1 + 2 + 1 = 16$

$\Delta^{\Sigma,\textit{Borda}}_{\textit{com}}(abc, abc, acb, cab, cba) = \sum_{r \in \boldsymbol{R}} K(acb, r) = 1 + 1 + 0 + 1 + 2 = 5$

Specific preference diversity indices

Definition (support-based PDI)

$\Delta^{\ell=k}_{supp}(\boldsymbol{R})$: number of ordered k-tuples of alternatives occurring in at least one individual preference in profile $\boldsymbol{R}$.

$\Delta^{\ell=m}_{supp}(\boldsymbol{R})$: *simple support-based PDI*, counts number of different preferences in $\boldsymbol{R}$.

Definition (distance-based PDI)

$\Delta^{\Phi,\delta}_{dist}(\boldsymbol{R})$: aggregated (e.g., $\Phi = \Sigma$) distance (δ) between all pairs of individual preferences in profile $\boldsymbol{R}$.

Kendall tau distance: $K(R, R') = \frac{1}{2} \cdot |\{(x, y) \mid xRy \text{ and } yR'x\}|$

Definition (compromise-based PDI)

$\Delta^{\Phi,F}_{com}(\boldsymbol{R})$: aggregated (e.g., $\Phi = \Sigma$) Kendall tau distance of individual preferences in $\boldsymbol{R}$ to a compromise preference $F(\boldsymbol{R})$ (e.g., F = Borda rule).

Example

$\Delta^{\ell=m}_{supp}(abc, abc, acb, cab, cba) = 4$

$\Delta^{\Sigma,K}_{dist}(abc, abc, acb, cab, cba) = 0 + 1 + 2 + 3 + 1 + 2 + 3 + 1 + 2 + 1 = 16$

$\Delta^{\Sigma,Borda}_{com}(abc, abc, acb, cab, cba) = \sum_{r \in \boldsymbol{R}} K(acb, r) = 1 + 1 + 0 + 1 + 2 = 5$

Specific preference diversity indices

Definition (support-based PDI)

$\Delta^{\ell=k}_{supp}(\boldsymbol{R})$: number of ordered k-tuples of alternatives occurring in at least one individual preference in profile $\boldsymbol{R}$.

$\Delta^{\ell=m}_{supp}(\boldsymbol{R})$: *simple support-based PDI*, counts number of different preferences in $\boldsymbol{R}$.

Definition (distance-based PDI)

$\Delta^{\Phi,\delta}_{dist}(\boldsymbol{R})$: aggregated (e.g., $\Phi = \Sigma$) distance (δ) between all pairs of individual preferences in profile $\boldsymbol{R}$.

Kendall tau distance: $K(R, R') = \frac{1}{2} \cdot |\{(x, y) \mid xRy \text{ and } yR'x\}|$

Definition (compromise-based PDI)

$\Delta^{\Phi,F}_{com}(\boldsymbol{R})$: aggregated (e.g., $\Phi = \Sigma$) Kendall tau distance of individual preferences in $\boldsymbol{R}$ to a compromise preference $F(\boldsymbol{R})$ (e.g., F = Borda rule).

Example

$\Delta^{\ell=m}_{supp}(abc, abc, acb, cab, cba) = 4$

$\Delta^{\Sigma,K}_{dist}(abc, abc, acb, cab, cba) = 0 + 1 + 2 + 3 + 1 + 2 + 3 + 1 + 2 + 1 = 16$

$\Delta^{\Sigma,Borda}_{com}(abc, abc, acb, cab, cba) = \sum_{r \in \boldsymbol{R}} K(acb, r) = 1 + 1 + 0 + 1 + 2 = 5$

Specific preference diversity indices

Definition (support-based PDI)

$\Delta^{\ell=k}_{supp}(\boldsymbol{R})$: number of ordered k-tuples of alternatives occurring in at least one individual preference in profile $\boldsymbol{R}$.

$\Delta^{\ell=m}_{supp}(\boldsymbol{R})$: *simple support-based PDI*, counts number of different preferences in $\boldsymbol{R}$.

Definition (distance-based PDI)

$\Delta^{\Phi,\delta}_{dist}(\boldsymbol{R})$: aggregated (e.g., $\Phi = \Sigma$) distance (δ) between all pairs of individual preferences in profile $\boldsymbol{R}$.

Kendall tau distance: $K(R, R') = \frac{1}{2} \cdot |\{(x, y) \mid xRy \text{ and } yR'x\}|$

Definition (compromise-based PDI)

$\Delta^{\Phi,F}_{com}(\boldsymbol{R})$: aggregated (e.g., $\Phi = \Sigma$) Kendall tau distance of individual preferences in $\boldsymbol{R}$ to a compromise preference $F(\boldsymbol{R})$ (e.g., F = Borda rule).

Example

$\Delta^{\ell=m}_{supp}(abc, abc, acb, cab, cba) = 4$

$\Delta^{\Sigma,K}_{dist}(abc, abc, acb, cab, cba) = 0 + 1 + 2 + 3 + 1 + 2 + 3 + 1 + 2 + 1 = 16$

$\Delta^{\Sigma,Borda}_{com}(abc, abc, acb, cab, cba) = \sum_{r\in\boldsymbol{R}} K(acb, r) = 1 + 1 + 0 + 1 + 2 = 5$

Example

	2 : abc 2 : bca 2 : cab	3 : abc 3 : cba	1 : abc 1 : acb 1 : bac 1 : bca 1 : cab 1 : cba	2 : abc 2 : bac 2 : acb
$\Delta^{\ell=m}_{supp}$	3	2	6	3
$\Delta^{\ell=2}_{supp}$	6	6	6	5
$\Delta^{\Sigma,D}_{dist}$	12	9	15	12
$\Delta^{\Sigma,K}_{dist}$	24	27	27	16
$\Delta^{\Sigma,S}_{dist}$	24	18	24	16
$\Delta^{\max,K}_{dist}$	2	3	3	2

Axioms

Axioms are used to evaluate/categorize methods.

PDO's are easier to deal with analytically. The results will also apply to PDI's indirectly.

*A PDO $\succcurlyeq$ is **anonymous** if, for every permutation $\sigma : \mathcal{N} \to \mathcal{N}$, we have $(R_1, \ldots, R_n) \sim (R_{\sigma(1)}, \ldots, R_{\sigma(n)})$.*

*A PDO $\succcurlyeq$ is **neutral** if, for every permutation $\tau : \mathcal{X} \to \mathcal{X}$, we have $(R_1, \ldots, R_n) \sim (\tau(R_1), \ldots, \tau(R_n))$.*

*A PDO $\succcurlyeq$ is **strongly discernible** if no two profiles are equally diverse, unless due to anonymity and neutrality.*

*A PDO $\succcurlyeq$ is **weakly discernible** if $\boldsymbol{R}$ being unanimous and $\boldsymbol{R'}$ not being unanimous together imply $\boldsymbol{R'} \succ \boldsymbol{R}$.*

*A PDO $\succcurlyeq$ is **support-invariant** if* $\textsc{Supp}(\boldsymbol{R}) = \textsc{Supp}(\boldsymbol{R'})$ *implies* $\boldsymbol{R} \sim \boldsymbol{R'}$.

Support-invariance $\Longrightarrow$ anonymity.

*A PDO $\succcurlyeq$ is **independent** if it is the case that $\boldsymbol{R} \succcurlyeq \boldsymbol{R'}$ if and only if $\boldsymbol{R} \oplus R \succcurlyeq \boldsymbol{R'} \oplus R$ for every two profiles $\boldsymbol{R}, \boldsymbol{R'} \in \mathcal{L}(\mathcal{X})^n$ and every preference $R \notin \textsc{Supp}(\boldsymbol{R}) \cup \textsc{Supp}(\boldsymbol{R'})$.*

Theoretical results

Basic axioms are satisfied by most PDO's:

Fact

*Every PDO induced by a PDI of the form $\Delta^{\ell=k}_{supp}$, $\Delta^{\Phi,\delta}_{dist}$, or $\Delta^{\Phi,F}_{com}$ with $k \in \{1,\ldots,m\}$, $\Phi \in \{\Sigma, \max\}$, $\delta \in \{K, S, D\}$, and F being an anonymous and neutral social welfare function is **anonymous**, **neutral**, and **weakly discernible**.*

Other axioms lead to impossibilities or narrow characterisations:

Proposition

*For $m > 2$ and $n > m!$, no PDO can be both support-invariant and **strongly discernable**.*

Proposition

*A PDO is support-invariant, **independent**, and weakly discernible if and only if it is the simple support-based PDO.*

Theoretical results

Basic axioms are satisfied by most PDO's:

Fact

Every PDO induced by a PDI of the form $\Delta^{\ell=k}_{supp}$, $\Delta^{\Phi,\delta}_{dist}$, or $\Delta^{\Phi,F}_{com}$ with $k \in \{1,\ldots,m\}$, $\Phi \in \{\Sigma, \max\}$, $\delta \in \{K,S,D\}$, and F being an anonymous and neutral social welfare function is ***anonymous***, ***neutral***, *and* ***weakly discernible***.

Other axioms lead to impossibilities or narrow characterisations:

Proposition

For $m > 2$ and $n > m!$, no PDO can be both support-invariant and ***strongly discernable***.

Proposition

A PDO is support-invariant, ***independent***, *and weakly discernible if and only if it is the simple support-based PDO.*

Table of Results

	$\Delta_{supp}^{\ell=k}$	$\Delta_{dist}^{\Sigma,\delta}$	$\Delta_{dist}^{\max,\delta}$	$\Delta_{com}^{\Sigma,F}$	$\Delta_{com}^{\max,F}$
Anonymity	✓	✓	✓	✓	✓
Neutrality	✓	✓	✓	✓	✓
Strong discernibility	X	X	X	X	X
Weak discernibility	✓	✓	✓	✓	✓
Support-invariance	✓	X	✓	X	X
Nonlocality	$n \leqslant k!$	✓	X	✓	X
Independence	$k = m$	X	X	X	X
Monotonicity	✓	X	✓	X	X
Swap-monotonicity	✓	$\delta = K$	$\delta = K$	F is Arrovian	

Experimental analysis

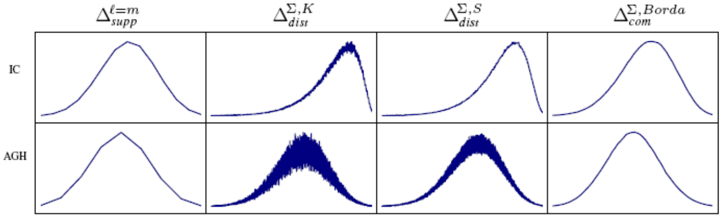
- Compare diversity of synthetic vs. real preference profiles
  - Impartial Culture assumption (IC): every possible profile is equally likely to occur
  - Course selection dataset (AGH): complete preferences of 153 students over 7 courses
- Relation between diversity and social choice-theoretic properties
  - Condorcet winner/cycle
  - agreement between voting rules
  - voter satisfaction

All profiles are preferences of 50 voters over 5 alternatives.

For each experiment we have drawn 1 million profiles from the relevant distribution.

Note that the number of all possible distinct profiles is: $(5!)^{50} > 10^{100}$

Diversity distribution across cultures



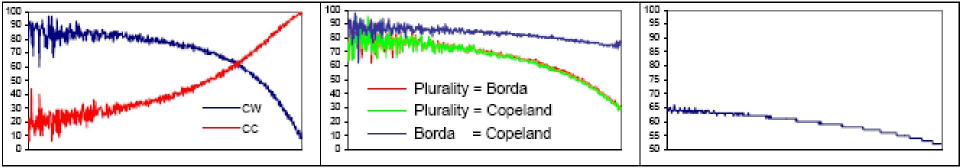
Preference diversity (x-axis) against frequency (y-axis) in IC and AGH. [$n = 50$, $m = 5$]

PDI	IC	AGH	PDI	IC	AGH	PDI	IC	AGH
$\Delta^{\ell=m}_{supp}$	**22**	**13**	$\Delta^{\Sigma,D}_{dist}$	34	244	$\Delta^{\Sigma,Bor}_{com}$	84	85
$\Delta^{\ell=2}_{supp}$	1	2	$\Delta^{\Sigma,S}_{dist}$	462	1170	$\Delta^{\Sigma,MG}_{com}$	94	88
$\Delta^{\ell=3}_{supp}$	4	12	$\Delta^{\Sigma,K}_{dist}$	**660**	**1561**	$\Delta^{\max,K}_{dist}$	2	3

Observed number of levels ($n = 50$, $m = 5$)

Impact on social choice-theoretic effects

Condorcet winners/cycles | Agreement between voting rules | Voter satisfaction



Diversity for $\Delta^{\Sigma,K}_{dist}$ / IC data (x-axis).

As diversity increases:

- the probability of encountering Condorcet cycles (winners) increases (decreases)
- average *degree of agreement* decreases
  - degree of agreement: $\frac{|W_1 \cap W_2|}{|W_1| \times |W_2|}$.
  - plurality rule has much more disagreement with other rules and it becomes worse as diversity increases
- average *voter satisfaction* decreases
  - voter satisfaction: number of alternatives below the (Borda) winner in the voter's preference
  - normalised to percent: average value is in the range of $50\% - 100\%$

Conclusion

- Preference diversity
  - Concept
  - Formal model
  - Axioms
  - Experiments
    - support our intuition/expectation

Future work

- Other options for measuring diversity
  - other distances and other aggregation operators (e.g., max-of-min)
  - for a given ℓ, maximum number of preferences with a common subpreference of length ℓ
  - for a given k, maximum length of a common subpreference of any k preferences
  - covering distance of the profile: how close a profile is to covering the full space of possibilities
  - measuring the distance from a single-peaked profile
- Normalization
  - Ratio
  - Percentile
  - Levels
- New axioms

Future work

- Distinguish (real data) profiles
  - Objective
  - Subjective
- Structure of profiles
  - Polarized/Divided
  - Central