Distance based aggregation rules

Mark C. Wilson
(joint with Benjamin Hadjibeyli, ENS Lyon)

Department of Computer Science
University of Auckland
www.cs.auckland.ac.nz/~mcw/

IMBS, UC Irvine, 2016-10-07

Speaker background

- PhD from Wisconsin (Mathematics), worked in Computer Science Department for 15 years. Main research is now in mathematical/computational social sciences.
- Frequent visitor to UCI: seminar talks 2013, 2015, 2016.
- Relevant interests: voting rules, electoral systems, matching algorithms, learning on networks, wisdom of crowds.
- Here until 14 October (SSPA 2117), happy to talk to anyone about research!

Too many voting rules, yet in some sense not enough

Too many properties

A unifying principle

- Think of a voting rule quite generally, as a mapping from a profile of preferences to an outcome.

A unifying principle

- Think of a voting rule quite generally, as a mapping from a profile of preferences to an outcome.
- Some subsets of the profile space yield an uncontroversial (consensus) outcome.

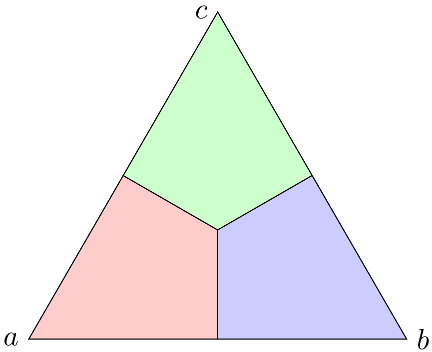
A unifying principle

- Think of a voting rule quite generally, as a mapping from a profile of preferences to an outcome.
- Some subsets of the profile space yield an uncontroversial (consensus) outcome.
- For other input profiles, we minimize their distance to a consensus set and choose the corresponding outcome.

A unifying principle

- Think of a voting rule quite generally, as a mapping from a profile of preferences to an outcome.
- Some subsets of the profile space yield an uncontroversial (consensus) outcome.
- For other input profiles, we minimize their distance to a consensus set and choose the corresponding outcome.
- This allows us to derive properties of the rule $\mathcal{R}(\mathcal{K}, d)$ from properties of the consensus $\mathcal{K}$ and distance d. Of course, we still need to agree on those!

Voronoi diagram



Commonly used inputs and outputs

- Assume we have n voters and m alternatives.

Commonly used inputs and outputs

- Assume we have n voters and m alternatives.
- Input preferences:

Commonly used inputs and outputs

- Assume we have n voters and m alternatives.
- Input preferences:
  - $\mathcal{L}$: linear orders of the form $c_1 \succ c_2 \succ \cdots \succ c_m$

Commonly used inputs and outputs

- Assume we have n voters and m alternatives.
- Input preferences:
  - $\mathcal{L}$: linear orders of the form $c_1 \succ c_2 \succ \cdots \succ c_m$
  - weak orders $c_1 \succeq c_2 \succeq \cdots \succeq c_m$

Commonly used inputs and outputs

- Assume we have n voters and m alternatives.
- Input preferences:
  - $\mathcal{L}$: linear orders of the form $c_1 \succ c_2 \succ \cdots \succ c_m$
  - weak orders $c_1 \succeq c_2 \succeq \cdots \succeq c_m$
  - pairwise preferences $c_i \succ c_j$ for each fixed i, j

Commonly used inputs and outputs

- Assume we have n voters and m alternatives.
- Input preferences:
  - $\mathcal{L}$: linear orders of the form $c_1 \succ c_2 \succ \cdots \succ c_m$
  - weak orders $c_1 \succeq c_2 \succeq \cdots \succeq c_m$
  - pairwise preferences $c_i \succ c_j$ for each fixed i, j
  - $\mathcal{T}$: top orders $c_i \succ c$ for all other c

Commonly used inputs and outputs

- Assume we have n voters and m alternatives.
- Input preferences:
  - $\mathcal{L}$: linear orders of the form $c_1 \succ c_2 \succ \cdots \succ c_m$
  - weak orders $c_1 \succeq c_2 \succeq \cdots \succeq c_m$
  - pairwise preferences $c_i \succ c_j$ for each fixed i, j
  - $\mathcal{T}$: top orders $c_i \succ c$ for all other c
  - dichotomous (approval) orders $\{c_1, \ldots, c_k\} \succ \{c_{k-1}, \ldots, c_m\}$

Commonly used inputs and outputs

- Assume we have n voters and m alternatives.
- Input preferences:
  - $\mathcal{L}$: linear orders of the form $c_1 \succ c_2 \succ \cdots \succ c_m$
  - weak orders $c_1 \succeq c_2 \succeq \cdots \succeq c_m$
  - pairwise preferences $c_i \succ c_j$ for each fixed i, j
  - $\mathcal{T}$: top orders $c_i \succ c$ for all other c
  - dichotomous (approval) orders $\{c_1, \ldots, c_k\} \succ \{c_{k-1}, \ldots, c_m\}$
  - k-approval orders (as above but k is fixed and given in advance).

Commonly used inputs and outputs

- Assume we have n voters and m alternatives.
- Input preferences:
  - $\mathcal{L}$: linear orders of the form $c_1 \succ c_2 \succ \cdots \succ c_m$
  - weak orders $c_1 \succeq c_2 \succeq \cdots \succeq c_m$
  - pairwise preferences $c_i \succ c_j$ for each fixed i, j
  - $\mathcal{T}$: top orders $c_i \succ c$ for all other c
  - dichotomous (approval) orders $\{c_1, \ldots, c_k\} \succ \{c_{k-1}, \ldots, c_m\}$
  - k-approval orders (as above but k is fixed and given in advance).
- Outcomes:

Commonly used inputs and outputs

- Assume we have n voters and m alternatives.
- Input preferences:
  - $\mathcal{L}$: linear orders of the form $c_1 \succ c_2 \succ \cdots \succ c_m$
  - weak orders $c_1 \succeq c_2 \succeq \cdots \succeq c_m$
  - pairwise preferences $c_i \succ c_j$ for each fixed i, j
  - $\mathcal{T}$: top orders $c_i \succ c$ for all other c
  - dichotomous (approval) orders $\{c_1, \ldots, c_k\} \succ \{c_{k-1}, \ldots, c_m\}$
  - k-approval orders (as above but k is fixed and given in advance).
- Outcomes:
  - top order (single winner, social choice function)

Commonly used inputs and outputs

- Assume we have n voters and m alternatives.
- Input preferences:
  - $\mathcal{L}$: linear orders of the form $c_1 \succ c_2 \succ \cdots \succ c_m$
  - weak orders $c_1 \succeq c_2 \succeq \cdots \succeq c_m$
  - pairwise preferences $c_i \succ c_j$ for each fixed i, j
  - $\mathcal{T}$: top orders $c_i \succ c$ for all other c
  - dichotomous (approval) orders $\{c_1, \ldots, c_k\} \succ \{c_{k-1}, \ldots, c_m\}$
  - k-approval orders (as above but k is fixed and given in advance).
- Outcomes:
  - top order (single winner, social choice function)
  - linear order (full ranking, social welfare function)

Commonly used inputs and outputs

- Assume we have n voters and m alternatives.
- Input preferences:
  - $\mathcal{L}$: linear orders of the form $c_1 \succ c_2 \succ \cdots \succ c_m$
  - weak orders $c_1 \succeq c_2 \succeq \cdots \succeq c_m$
  - pairwise preferences $c_i \succ c_j$ for each fixed i, j
  - $\mathcal{T}$: top orders $c_i \succ c$ for all other c
  - dichotomous (approval) orders $\{c_1, \ldots, c_k\} \succ \{c_{k-1}, \ldots, c_m\}$
  - k-approval orders (as above but k is fixed and given in advance).
- Outcomes:
  - top order (single winner, social choice function)
  - linear order (full ranking, social welfare function)
  - k-approval order (committee of size k)

Commonly used inputs and outputs

- Assume we have n voters and m alternatives.
- Input preferences:
  - $\mathcal{L}$: linear orders of the form $c_1 \succ c_2 \succ \cdots \succ c_m$
  - weak orders $c_1 \succeq c_2 \succeq \cdots \succeq c_m$
  - pairwise preferences $c_i \succ c_j$ for each fixed i, j
  - $\mathcal{T}$: top orders $c_i \succ c$ for all other c
  - dichotomous (approval) orders $\{c_1, \ldots, c_k\} \succ \{c_{k-1}, \ldots, c_m\}$
  - k-approval orders (as above but k is fixed and given in advance).
- Outcomes:
  - top order (single winner, social choice function)
  - linear order (full ranking, social welfare function)
  - k-approval order (committee of size k)
  - any valid input could also be an output, and vice versa.

Examples: consensus sets

- strong unanimity $\mathbf{S}$ - every voter has the same preference. Works for any inputs.

Examples: consensus sets

- strong unanimity $\mathbf{S}$ - every voter has the same preference. Works for any inputs.
- weak unanimity $\mathbf{W}$ - every voter has the same top choice. Works for linear orders, top orders.

Examples: consensus sets

- strong unanimity $\mathbf{S}$ - every voter has the same preference. Works for any inputs.
- weak unanimity $\mathbf{W}$ - every voter has the same top choice. Works for linear orders, top orders.
- Condorcet winner $\mathbf{C}$ - some candidate beats all others in pairwise comparisons. Works for linear and pairwise inputs.

Examples: consensus sets

- strong unanimity $\mathbf{S}$ - every voter has the same preference. Works for any inputs.
- weak unanimity $\mathbf{W}$ - every voter has the same top choice. Works for linear orders, top orders.
- Condorcet winner $\mathbf{C}$ - some candidate beats all others in pairwise comparisons. Works for linear and pairwise inputs.
- Condorcet order $\mathbf{C}_*$ - the pairwise majority relation is a total order. Works for linear and pairwise inputs.

Examples: distance

- Define a graph by stipulating edges between some pairs of inputs. Then define $d(E, E')$ to be the length of a shortest path in this graph. This includes:

distance	create an edge when we
Hamming d_H	change one preference order
Kemeny d_K	swap two candidates in a pref. order
insertion d_{ins}	add a preference order
deletion d_{del}	delete a preference order
tournament d_{RT}	reverse arrow in majority tournament

Examples: distance

- Define a graph by stipulating edges between some pairs of inputs. Then define $d(E, E')$ to be the length of a shortest path in this graph. This includes:

distance	create an edge when we
Hamming d_H	change one preference order
Kemeny d_K	swap two candidates in a pref. order
insertion d_{ins}	add a preference order
deletion d_{del}	delete a preference order
tournament d_{RT}	reverse arrow in majority tournament

- Votewise distances can be formed by using any distance d on individual orders and combining the components with a norm on $\mathbb{R}^n$. The most common is ℓ^1, yielding d^1.

A concrete example

- Inputs: linear orders. Output: a winner.

A concrete example

- Inputs: linear orders. Output: a winner.
- Consensus: weak unanimity (everyone agrees on the winner).

A concrete example

- Inputs: linear orders. Output: a winner.
- Consensus: weak unanimity (everyone agrees on the winner).
- Distance: Kemeny.

A concrete example

- Inputs: linear orders. Output: a winner.
- Consensus: weak unanimity (everyone agrees on the winner).
- Distance: Kemeny.
- Name of rule: Borda.

A concrete example

- Inputs: linear orders. Output: a winner.
- Consensus: weak unanimity (everyone agrees on the winner).
- Distance: Kemeny.
- Name of rule: Borda.
- For an input with 5 voters with preferences abc, acb, bac, bca, cab, the distance to the consensus sets where a, b, c wins are respectively $3, 5, 6$. The unique winner is a.

A concrete example

- Inputs: linear orders. Output: a winner.
- Consensus: weak unanimity (everyone agrees on the winner).
- Distance: Kemeny.
- Name of rule: Borda.
- For an input with 5 voters with preferences abc, acb, bac, bca, cab, the distance to the consensus sets where a, b, c wins are respectively $3, 5, 6$. The unique winner is a.
- Changing the distance to Hamming yields plurality rule. There is a tie between a and b.

Examples: some DR rules $\mathcal{R}(\mathcal{K}, d)$, input $\mathcal{L}$, output $\mathcal{L}/\mathcal{T}$

$\mathcal{K}$ d	d_H	d_K	d_{RT}	d_{ins}	d_{del}
$\mathbf{S}$	MR	Kemeny	Copeland	undef	MR
$\mathbf{W}$	plurality	Borda	Copeland	undef	plurality
$\mathbf{C}$	VRR	Dodgson	Copeland	maximin	Young
$\mathbf{C}_*$			Slater		

- Here
  - VRR = "voter replacement rule" (Elkind, Faliszewski, Slinko; Soc Choice Welf 2012)
  - MR = "modal ranking rule" (Caragiannis, Procaccia, Shah; AAAI 2014).

Mathematical notes

- The distance need not be a metric — in fact pseudometrics, quasimetrics and in general hemimetrics all arise in applications.

Mathematical notes

- The distance need not be a metric — in fact pseudometrics, quasimetrics and in general hemimetrics all arise in applications.
- The norm could be replaced by a seminorm (need not distinguish points).

Putting together nice components to make nice rules

- If $\mathcal{K}$ and d satisfy the following, then so does $\mathcal{R}(\mathcal{K}, d)$

Putting together nice components to make nice rules

- If $\mathcal{K}$ and d satisfy the following, then so does $\mathcal{R}(\mathcal{K}, d)$
  - anonymity (voter identities don't matter)

Putting together nice components to make nice rules

- If $\mathcal{K}$ and d satisfy the following, then so does $\mathcal{R}(\mathcal{K}, d)$
  - anonymity (voter identities don't matter)
  - neutrality (candidate identities don't matter)

Putting together nice components to make nice rules

- If $\mathcal{K}$ and d satisfy the following, then so does $\mathcal{R}(\mathcal{K}, d)$
  - anonymity (voter identities don't matter)
  - neutrality (candidate identities don't matter)
  - consistency (combining voter sets makes no difference if they agree on the result)

Putting together nice components to make nice rules

- If $\mathcal{K}$ and d satisfy the following, then so does $\mathcal{R}(\mathcal{K}, d)$
  - anonymity (voter identities don't matter)
  - neutrality (candidate identities don't matter)
  - consistency (combining voter sets makes no difference if they agree on the result)
  - continuity (small changes in inputs don't usually matter)

Putting together nice components to make nice rules

- If $\mathcal{K}$ and d satisfy the following, then so does $\mathcal{R}(\mathcal{K}, d)$
  - anonymity (voter identities don't matter)
  - neutrality (candidate identities don't matter)
  - consistency (combining voter sets makes no difference if they agree on the result)
  - continuity (small changes in inputs don't usually matter)
  - reversal symmetry (turning input orders upside down reverses the output)

Putting together nice components to make nice rules

- If $\mathcal{K}$ and d satisfy the following, then so does $\mathcal{R}(\mathcal{K}, d)$
  - anonymity (voter identities don't matter)
  - neutrality (candidate identities don't matter)
  - consistency (combining voter sets makes no difference if they agree on the result)
  - continuity (small changes in inputs don't usually matter)
  - reversal symmetry (turning input orders upside down reverses the output)
- Interestingly, homogeneity does not extend in this way. Monotonicity is also tricky.

Compressing the input

- Most rules we study are anonymous and homogeneous. Thus they depend only on the (probability) distribution of voter preferences.

Compressing the input

- Most rules we study are anonymous and homogeneous. Thus they depend only on the (probability) distribution of voter preferences.
- In this case we can compress the input. For example with m candidates and n voters there are $(m!)^n$ profiles, but only $\binom{n+m!-1}{n}$ equivalence classes under anonymity. The number of voters with each preference order is all we need to know for an anonymous rule, not who has which order.

Compressing the input

- Most rules we study are anonymous and homogeneous. Thus they depend only on the (probability) distribution of voter preferences.
- In this case we can compress the input. For example with m candidates and n voters there are $(m!)^n$ profiles, but only $\binom{n+m!-1}{n}$ equivalence classes under anonymity. The number of voters with each preference order is all we need to know for an anonymous rule, not who has which order.
- We can then describe consensus sets quite efficiently in the quotient space, which is a simplex of dimension $m! - 1$. For example, strong unanimity consensus $\leftrightarrow$ vertices of simplex.

Compressing the input

- Most rules we study are anonymous and homogeneous. Thus they depend only on the (probability) distribution of voter preferences.
- In this case we can compress the input. For example with m candidates and n voters there are $(m!)^n$ profiles, but only $\binom{n+m!-1}{n}$ equivalence classes under anonymity. The number of voters with each preference order is all we need to know for an anonymous rule, not who has which order.
- We can then describe consensus sets quite efficiently in the quotient space, which is a simplex of dimension $m! - 1$. For example, strong unanimity consensus $\leftrightarrow$ vertices of simplex.
- The law of conservation of difficulty applies: the distance can be harder to understand.

New results

- A votewise distance d on profiles corresponds to a Wasserstein distance d_W on anonymous and homogeneous profiles (preference distributions). This is related to the theory of optimal transportation.

New results

- A votewise distance d on profiles corresponds to a Wasserstein distance d_W on anonymous and homogeneous profiles (preference distributions). This is related to the theory of optimal transportation.
- If the votewise distance uses ℓ^1, then d_W is induced by a norm, so we have a Minkowski space.

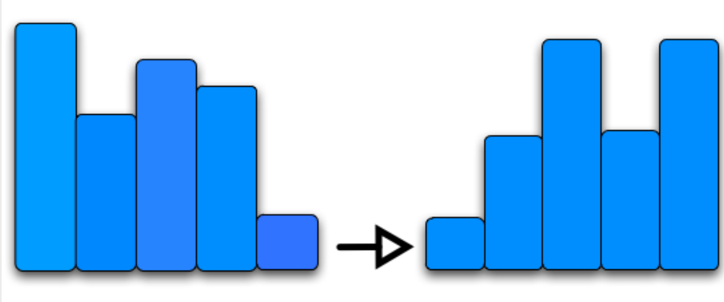
New results

- A votewise distance d on profiles corresponds to a Wasserstein distance d_W on anonymous and homogeneous profiles (preference distributions). This is related to the theory of optimal transportation.
- If the votewise distance uses ℓ^1, then d_W is induced by a norm, so we have a Minkowski space.
- This gives a geometric interpretation not seen before in the voting literature, and connects it to well-developed areas of mathematics.

New results

- A votewise distance d on profiles corresponds to a Wasserstein distance d_W on anonymous and homogeneous profiles (preference distributions). This is related to the theory of optimal transportation.
- If the votewise distance uses ℓ^1, then d_W is induced by a norm, so we have a Minkowski space.
- This gives a geometric interpretation not seen before in the voting literature, and connects it to well-developed areas of mathematics.
- All rules using the strong unanimity consensus and neutral votewise distances are anonymous, homogeneous, neutral, consistent, and continuous, and have tied sets lying in hyperplanes. If the distance satisfies reversal symmetry, so does the rule.

Optimal transportation picture



How to make this change with minimal effort?

The strange case of ℓ^1

- In the context of distances, ℓ^1 norm corresponds to adding the contribution to the distance (or error) from each voter. It is very natural, and no votewise rule in the literature uses anything else, including ℓ^2, the Euclidean norm.

The strange case of ℓ^1

- In the context of distances, ℓ^1 norm corresponds to adding the contribution to the distance (or error) from each voter. It is very natural, and no votewise rule in the literature uses anything else, including ℓ^2, the Euclidean norm.
- However the geometry of the Minkowski space ℓ^1 is much less nice than ℓ^2. For example, Voronoi cells are not convex, bisectors can be large (see below).

The strange case of ℓ^1

- In the context of distances, ℓ^1 norm corresponds to adding the contribution to the distance (or error) from each voter. It is very natural, and no votewise rule in the literature uses anything else, including ℓ^2, the Euclidean norm.
- However the geometry of the Minkowski space ℓ^1 is much less nice than ℓ^2. For example, Voronoi cells are not convex, bisectors can be large (see below).
- Despite all these possibly negative consequences for decisiveness, we don't see rules with such bad behaviour. Why?

The strange case of ℓ^1

- In the context of distances, ℓ^1 norm corresponds to adding the contribution to the distance (or error) from each voter. It is very natural, and no votewise rule in the literature uses anything else, including ℓ^2, the Euclidean norm.
- However the geometry of the Minkowski space ℓ^1 is much less nice than ℓ^2. For example, Voronoi cells are not convex, bisectors can be large (see below).
- Despite all these possibly negative consequences for decisiveness, we don't see rules with such bad behaviour. Why?
- The answer lies in subtle choices of consensus and distance. We have recently shown how such rules can be very indecisive if the choice is not made well.

Voronoi diagram

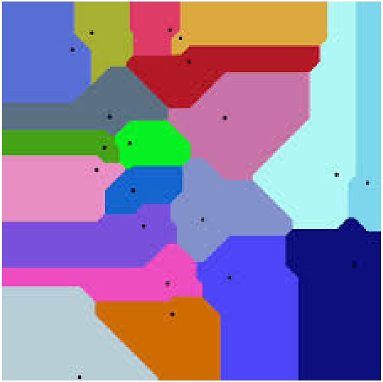


Figure: Voronoi diagram for points in ℓ^1. Source: Wikipedia.

ℓ^1 large bisector

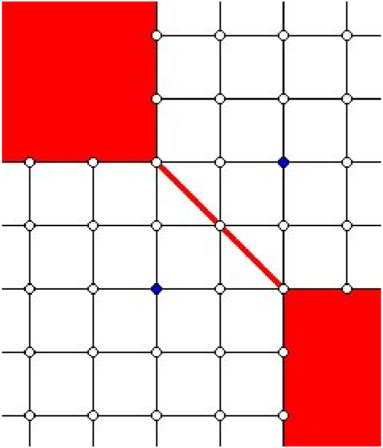
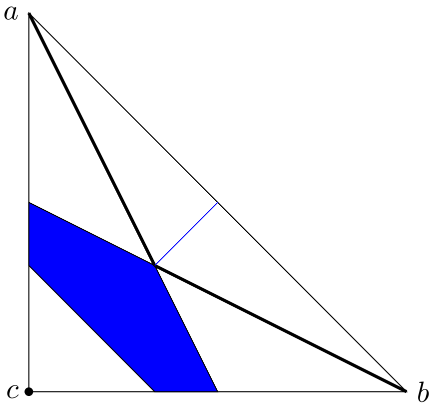


Figure: Large bisector in ℓ^1. Source: http://www.ams.org/samplings/feature-column/fcarc-taxi.

Bad things can happen even with ℓ^2



Such a rule would be very indecisive.

Consensus sets

- Based on our geometric analysis, in order to create voting rules with good behaviour, we should require our consensus to be:

Consensus sets

- Based on our geometric analysis, in order to create voting rules with good behaviour, we should require our consensus to be:
  - anonymous

Consensus sets

- Based on our geometric analysis, in order to create voting rules with good behaviour, we should require our consensus to be:
  - anonymous
  - consistent

Consensus sets

- Based on our geometric analysis, in order to create voting rules with good behaviour, we should require our consensus to be:
  - anonymous
  - consistent
  - neutral

Consensus sets

- Based on our geometric analysis, in order to create voting rules with good behaviour, we should require our consensus to be:
  - anonymous
  - consistent
  - neutral
  - touching the boundary of the simplex

Consensus sets

- Based on our geometric analysis, in order to create voting rules with good behaviour, we should require our consensus to be:
  - anonymous
  - consistent
  - neutral
  - touching the boundary of the simplex
  - separated

Consensus sets

- Based on our geometric analysis, in order to create voting rules with good behaviour, we should require our consensus to be:
  - anonymous
  - consistent
  - neutral
  - touching the boundary of the simplex
  - separated
- The last condition rules out the Condorcet consensus.

A family of nice rules

- Suppose that d is a neutral distance (like d_K, d_H) on rankings, and $\mathcal{K}$ is the strong or weak unanimity consensus.

A family of nice rules

- Suppose that d is a neutral distance (like d_K, d_H) on rankings, and $\mathcal{K}$ is the strong or weak unanimity consensus.
- The rule that uses $\mathcal{K}$ and the votewise version of d with an ℓ^p norm is:

A family of nice rules

- Suppose that d is a neutral distance (like d_K, d_H) on rankings, and $\mathcal{K}$ is the strong or weak unanimity consensus.
- The rule that uses $\mathcal{K}$ and the votewise version of d with an ℓ^p norm is:
  - anonymous

A family of nice rules

- Suppose that d is a neutral distance (like d_K, d_H) on rankings, and $\mathcal{K}$ is the strong or weak unanimity consensus.
- The rule that uses $\mathcal{K}$ and the votewise version of d with an ℓ^p norm is:
  - anonymous
  - neutral

A family of nice rules

- Suppose that d is a neutral distance (like d_K, d_H) on rankings, and $\mathcal{K}$ is the strong or weak unanimity consensus.
- The rule that uses $\mathcal{K}$ and the votewise version of d with an ℓ^p norm is:
  - anonymous
  - neutral
  - consistent

A family of nice rules

- Suppose that d is a neutral distance (like d_K, d_H) on rankings, and $\mathcal{K}$ is the strong or weak unanimity consensus.
- The rule that uses $\mathcal{K}$ and the votewise version of d with an ℓ^p norm is:
  - anonymous
  - neutral
  - consistent
  - continuous

A family of nice rules

- Suppose that d is a neutral distance (like d_K, d_H) on rankings, and $\mathcal{K}$ is the strong or weak unanimity consensus.
- The rule that uses $\mathcal{K}$ and the votewise version of d with an ℓ^p norm is:
  - anonymous
  - neutral
  - consistent
  - continuous
- These rules have not been studied in detail. In the case of weak consensus they must be scoring rules by Young's characterization (1975), but this must be made explicit.

A family of nice rules

- Suppose that d is a neutral distance (like d_K, d_H) on rankings, and $\mathcal{K}$ is the strong or weak unanimity consensus.
- The rule that uses $\mathcal{K}$ and the votewise version of d with an ℓ^p norm is:
  - anonymous
  - neutral
  - consistent
  - continuous
- These rules have not been studied in detail. In the case of weak consensus they must be scoring rules by Young's characterization (1975), but this must be made explicit.
- For the strong unanimity consensus, the class includes Kemeny rule; such rules are all maximum likelihood estimators for various noise models.

Strong unanimity: maximum likelihood estimation

- When using the strong unanimity consensus, we can interpret the above procedure as maximizing the likelihood of a real underlying ranking, given the voters' observations.

Strong unanimity: maximum likelihood estimation

- When using the strong unanimity consensus, we can interpret the above procedure as maximizing the likelihood of a real underlying ranking, given the voters' observations.
- This idea of "wisdom of crowds" is related to Condorcet's Jury Theorem — we use voting as statistical estimation, not preference aggregation. There is renewed interest because of internet-based crowdsourcing applications.

Strong unanimity: maximum likelihood estimation

- When using the strong unanimity consensus, we can interpret the above procedure as maximizing the likelihood of a real underlying ranking, given the voters' observations.
- This idea of "wisdom of crowds" is related to Condorcet's Jury Theorem — we use voting as statistical estimation, not preference aggregation. There is renewed interest because of internet-based crowdsourcing applications.
- The key is that the distance can be interpreted as a noise model. The basic (Mallows) model for rankings boils down to

$$\Pr(\rho|P) = C \prod_{k=1}^{n} q^{-d(P_k, \rho_k^*)} = C q^{-d^1(P, \mathbf{S}_\rho)}.$$

Here ρ is an input, P a profile of inputs, and $0 < q < 1$. Thus the MLE is found by computing $\mathcal{R}(\mathbf{S}, d^1)$.

Example: MLE

- Suppose the profile is $P := \{abc, acb, cab, bca, bac\}$ and we use the Kemeny noise model (voters make small independent errors in ranking adjacent candidates).

Example: MLE

- Suppose the profile is $P := \{abc, acb, cab, bca, bac\}$ and we use the Kemeny noise model (voters make small independent errors in ranking adjacent candidates).
- The likelihood function for each of the 6 possible rankings is maximized when the Kemeny distance is minimized.

Example: MLE

- Suppose the profile is $P := \{abc, acb, cab, bca, bac\}$ and we use the Kemeny noise model (voters make small independent errors in ranking adjacent candidates).
- The likelihood function for each of the 6 possible rankings is maximized when the Kemeny distance is minimized.
- For example, $d(P, \mathbf{S}_{abc}) = 0 + 1 + 2 + 2 + 1 = 6$. The distances to acb, bac, bca, cab, cba are respectively $7, 8, 8, 7, 9$.

Example: MLE

- Suppose the profile is $P := \{abc, acb, cab, bca, bac\}$ and we use the Kemeny noise model (voters make small independent errors in ranking adjacent candidates).
- The likelihood function for each of the 6 possible rankings is maximized when the Kemeny distance is minimized.
- For example, $d(P, \mathbf{S}_{abc}) = 0 + 1 + 2 + 2 + 1 = 6$. The distances to acb, bac, bca, cab, cba are respectively $7, 8, 8, 7, 9$.
- Thus abc is the best estimate by Kemeny's rule.

Example: MLE

- Suppose the profile is $P := \{abc, acb, cab, bca, bac\}$ and we use the Kemeny noise model (voters make small independent errors in ranking adjacent candidates).
- The likelihood function for each of the 6 possible rankings is maximized when the Kemeny distance is minimized.
- For example, $d(P, \mathbf{S}_{abc}) = 0 + 1 + 2 + 2 + 1 = 6$. The distances to acb, bac, bca, cab, cba are respectively $7, 8, 8, 7, 9$.
- Thus abc is the best estimate by Kemeny's rule.
- If we use the Hamming noise model (voters make independent large errors, randomly choosing a wrong vote with probability q) then the MLE instead chooses the most common ranking from the input.

Strong unanimity: universal rule

- Zwicker (2014) used the characteristic function encoding of a binary relation to give a universal approach that yields Borda and Kemeny in special cases.

Strong unanimity: universal rule

- Zwicker (2014) used the characteristic function encoding of a binary relation to give a universal approach that yields Borda and Kemeny in special cases.
- It also yields other rules with an "inversion-counting" ℓ^1 flavour for different input-output pairs.

Strong unanimity: universal rule

- Zwicker (2014) used the characteristic function encoding of a binary relation to give a universal approach that yields Borda and Kemeny in special cases.
- It also yields other rules with an “inversion-counting” ℓ^1 flavour for different input-output pairs.
- The characteristic function encodes pairwise information in a matrix: the usual tournament matrix is just the sum of these over all voters.

Strong unanimity: universal rule

- Zwicker (2014) used the characteristic function encoding of a binary relation to give a universal approach that yields Borda and Kemeny in special cases.
- It also yields other rules with an "inversion-counting" ℓ^1 flavour for different input-output pairs.
- The characteristic function encodes pairwise information in a matrix: the usual tournament matrix is just the sum of these over all voters.
- Thus the rationalization of Kemeny's rule via ℓ^1 and the hypercube from Saari & Merlin (2000) is an isomorphic representation of the usual one via $\mathbf{S}$ and d_K.

Strong unanimity: universal rule

- Zwicker (2014) used the characteristic function encoding of a binary relation to give a universal approach that yields Borda and Kemeny in special cases.
- It also yields other rules with an "inversion-counting" ℓ^1 flavour for different input-output pairs.
- The characteristic function encodes pairwise information in a matrix: the usual tournament matrix is just the sum of these over all voters.
- Thus the rationalization of Kemeny's rule via ℓ^1 and the hypercube from Saari & Merlin (2000) is an isomorphic representation of the usual one via $\mathbf{S}$ and d_K.
- Presumably this gives a unified maximum likelihood interpretation.

Strong unanimity: universal rule

- Zwicker (2014) used the characteristic function encoding of a binary relation to give a universal approach that yields Borda and Kemeny in special cases.
- It also yields other rules with an "inversion-counting" ℓ^1 flavour for different input-output pairs.
- The characteristic function encodes pairwise information in a matrix: the usual tournament matrix is just the sum of these over all voters.
- Thus the rationalization of Kemeny's rule via ℓ^1 and the hypercube from Saari & Merlin (2000) is an isomorphic representation of the usual one via $\mathbf{S}$ and d_K.
- Presumably this gives a unified maximum likelihood interpretation.
- Are there other universal rules with different encodings?

Interesting questions

- What does the quotient distance corresponding to d_K look like on the simplex?

Interesting questions

- What does the quotient distance corresponding to d_K look like on the simplex?
- There are other geometric interpretations (using the permutahedron) of distance rationalization, which are obviously special cases of our approach under change of variables. But what about explicit formulae?

Interesting questions

- What does the quotient distance corresponding to d_K look like on the simplex?
- There are other geometric interpretations (using the permutahedron) of distance rationalization, which are obviously special cases of our approach under change of variables. But what about explicit formulae?
- There are many commonly used statistical distances on the simplex, which as far as I know have not been used for voting (the simplest case is total variation which corresponds to the ℓ^1-Hamming distance). Similarly, other distances on pairwise tournament matrices could be used.

Interesting questions

- What does the quotient distance corresponding to d_K look like on the simplex?
- There are other geometric interpretations (using the permutahedron) of distance rationalization, which are obviously special cases of our approach under change of variables. But what about explicit formulae?
- There are many commonly used statistical distances on the simplex, which as far as I know have not been used for voting (the simplest case is total variation which corresponds to the ℓ^1-Hamming distance). Similarly, other distances on pairwise tournament matrices could be used.
- How many of our results extend easily to general input/output cases?

Conclusion

- Distance rationalization offers a way to navigate the jungle of aggregation rules.

Conclusion

- Distance rationalization offers a way to navigate the jungle of aggregation rules.
- Very little has been done beyond linear orders. The idea of universal rules is quite attractive.

Conclusion

- Distance rationalization offers a way to navigate the jungle of aggregation rules.
- Very little has been done beyond linear orders. The idea of universal rules is quite attractive.
- There are many new rules waiting to be discovered (invented?). Using distances that are not votewise, and compressed input, allows for more complex and interesting ones.

Conclusion

- Distance rationalization offers a way to navigate the jungle of aggregation rules.
- Very little has been done beyond linear orders. The idea of universal rules is quite attractive.
- There are many new rules waiting to be discovered (invented?). Using distances that are not votewise, and compressed input, allows for more complex and interesting ones.
- Some of the "new" rules are actually old ones in disguise. We need to make these alternate representations explicit.

References

- B. Hadjibeyli, M. C. Wilson. Distance rationalization of social rules. Submitted to Social Choice and Welfare, 2016.
- B. Hadjibeyli, M. C. Wilson. Distance rationalization of anonymous and homogeneous social rules. Submitted to Social Choice and Welfare, 2016.