Scoring rules for judgment aggregation

Franz Dietrich

Workshop on Judgment Aggregation
University of Luxembourg
1st February 2012

The working paper is downloadable from the author's homepage:
www.franzdietrich.net
(This paper contains more material than is presented today.)

Background

- The JA problem: How can or should we merge many individuals' yes/no judgments on some interconnected propositions?
- A very general problem!
  –> generalizes the classical preference aggregation problem
- **Leading example** (born in legal theory): three jurors in a court trial need to merge their yes/no judgments on three propositions:
  - p : the defendant has broken the contract;
  - q : the contract is legally valid;
  - r : the defendant is liable (r).
- According to a universally accepted legal doctrine, r (the 'conclusion') is true if and only if p and r (the two 'premises') are both true.

Background

- Propositionwise majority rule may generate inconsistent collective judgments:

	premise p	premise q	conclusion r ($\Leftrightarrow p \wedge q$)
Juror 1	Yes	Yes	Yes
Juror 2	Yes	No	No
Juror 3	No	Yes	No
Majority	Yes	Yes	No

Background

- There are numerous other possible 'agendas', i.e., kinds of interconnected propositions a group might face.
- E.g., for preference aggregation,
  - propositions take the form 'x is better than y' (for various alternatives x and y),
  - these propositions are interconnected through standard conditions such as transitivity.
  - Condorcet's classical *voting paradox* about cyclical majority preferences is nothing but another example of inconsistent majority judgments.

Where does the theory stand?

- **Early phase** (perhaps until the 2010 JA symposium in JET):
  - Dominated by lots of exciting impossibility findings
  - The generic finding: For MANY agendas of propositions, there are NO propositionwise aggregation rules satisfying mild extra conditions XYZ.
  - The exact meanings of 'MANY' and 'XYZ' differ across results.
- **Current phase** (illustrated by the 2011 JA workshop in Freudenstadt):
  - Constructing concrete JA rules!
  - An experimental, playful, 'fun' phase.
  - Much seems permitted: we can try out rules without already providing complete axiomatic foundations.

And tomorrow?

- **Future phase (?)**:
  - a return to axiomatics
  - characterizing concrete (classes of) rules, i.e., finding their necessary and sufficient properties.

And the day after tomorrow?

- **Far future phase:**
  - After all theoretical work has been completed, we will together go on the streets, proclaim the good news, and force the world to use our rules :-).

But now let's go back to the present, 'experimental' phase!

What rules are on the market right now?

Existing proposals for generating consistent collective judgments:

- Premise- and conclusion-based rules (e.g., Pettit 2001, List & Pettit 2002, Dietrich 2006, Dietrich and Mongin 2010)
- Sequential priority rules (e.g., List 2004, Dietrich and List 2007)
- Distance-based rules (e.g., Konieczny & Pino-Perez 2002, Pigozzi 2005, Miller & Osherson 2008, Eckert & Klamler 2009, Hartmann , Pigozzi & Sprenger 2010, Lang, Pigozzi, Slavkovik & van der Torre 2011, Duddy and Piggins 2011)
- An (attempt of a) Borda-type aggregation rule (Zwicker 2011)
- 'Condorcet admissible' aggregation (Nehring, Pivato and Puppe 2011).

A new proposal: scoring rules

- *Scoring rules*: they select collective judgments which 'score' highest in total.
  - inspired from classical scoring rules in preference aggregation theory, such as Borda rule (e.g., Smith 1973, Young 1975, Myerson 1995, Zwicker 2008, Pivato 2011))
- Conceptually, our scoring rules differ from the classical ones is that ours assign scores to propositions, not to alternatives.
  –> in a general JA problem, there are no 'alternatives'!
- Nonetheless, our scoring rules will turn out to generalize the classical ones.

The ‘scoring paradigm’

- The paradigm underlying our scoring rules – i.e., the maximization of total score of collective judgments – differs from the standard paradigms in JA
  - such as the premise-, conclusion- or distance-based paradigms.
- Nonetheless, several existing rules (e.g., the Hamming rule) can be re-modelled as particular scoring rules, and can thus be ‘rationalized’ in terms of the maximization of total scores.

Goal

Goal: Explore various plausible ways to define scores – several 'scorings'.

- Some scorings lead to ('reconstruct') existing aggregation rules.
- Other scorings lead to new rules
- ... such as a Borda rule for JA.
  - 'Generalizing Borda' to JA has been a long-lasting open problem.
  - Zwicker (2011) made an interesting (incomplete) proposal (and Conal Duddy and Ashley Piggins also have work in progress about this).
  - Surprisingly, Zwicker's and my proposal are distinctively different.

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

Part 3: Set scoring rules

Part 4: Concluding remarks

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

Part 3: Set scoring rules

Part 4: Concluding remarks

The agenda

- Consider a set of n (≥ 2) individuals, denoted $N = \{1, ..., n\}$.
- An **agenda** of propositions on which judgments are needed. Formally, the agenda is an arbitrary set X (whose elements we call 'propositions') such that
  - X is closed under negation: for every proposition p in X there is a specified proposition denoted $\neg p$ ('not p') in X, where of course $\neg p \neq p$ and $\neg\neg p = p$;
  - X is endowed with logical interconnections: there is a specification of which subsets of X are 'consistent' (i.e., formally, there is a system $\mathcal{C}$ of subsets called 'consistent').

The agenda (cont.)

N.B.:

- This notion of an 'agenda' is very general. It might contain:
  - syntactic propositions (logical sentences), or
  - semantic propositions (modelled for instance as sets of worlds), or
  - arbitrary attributes that an agent may or may not possess.
- It is often natural to regard the agenda X as a subset of a logic $\mathbf{L}$ from which it inherits the negation operator and the logical interconnections.[1]

[1] This logic is general: it could for instance be standard propositional logic, standard predicate logic, or various modal or conditional logics (see Dietrich 2007).

Judgment sets

- A set $A \subseteq X$ (a 'judgment set') is
  - *complete* if it contains a member of each pair $p, \neg p \in X$,
  - *(fully) rational* if it is complete and consistent.
- $\mathcal{D}$ is the set of all rational judgment sets.

Regularity assumptions

- As usual, I assume the consistency notion is 'regular'.
- That is, the system of consistent sets takes the form $\mathcal{C} = \{C \subseteq A : A \in \mathcal{D}\} \neq \varnothing$, so that consistent sets are subsets of complete and consistent sets.
- (Equivalently, the consistency notion satisfies three weak conditions hold.[2])
- So, to specify all logical interconnections, it suffices to specify $\mathcal{D}$.

- Also, let X be finite.

[2](C1) No set $\{p, \neg p\}$ is consistent ('self-entailment'). (C2) Subsets of consistent sets are consistent ('monotonicity'). (C3) $\varnothing$ is consistent and each consistent set can be extended to a complete and consistent set ('completability'). See Dietrich (2007).

Notation

- A judgment set $A \subseteq X$ is often abbreviated by concatenating its members in any order (so, $p\neg q\neg r$ is short for $\{p, \neg q, \neg r\}$).
- The negation-closure of a set $Y \subseteq X$ is denoted

$$Y^{\pm} \equiv \{p, \neg p : p \in Y\}.$$

Example 1: the 'doctrinal paradox' agenda

- This agenda is

$$X = \{p, q, r\}^{\pm},$$

- where logical interconnections are defined relative to the external constraint $r \leftrightarrow (p \wedge q)$. So,

$$\mathcal{D} = \{pqr, p\neg q\neg r, \neg pq\neg r, \neg p\neg q\neg r\}.$$

Example 2: the preference agenda

- For an arbitrary, finite set of alternatives K, the *preference agenda* is defined as

$$X = X_K = \{xPy : x, y \in K, x \neq y\},$$

- where the negation of a proposition xPy is of course $\neg xPy = yPx$,
- and where logical interconnections are defined relative to the usual conditions of transitivity, asymmetry and connectedness, which define a *strict linear order*.
- Formally, to each binary relation $\succ$ over K uniquely corresponds a judgment set, denoted $A_\succ = \{xPy \in X : x \succ y\}$, and the set of all rational judgment sets is

$$\mathcal{D} = \{A_\succ : \ \succ \text{ is a strict linear order over } K\}.$$

Aggregation rules

- A *(multi-valued) aggregation rule* is a correspondence F which to every profile of 'individual' judgment sets $(A_1, ..., A_n)$ (from some domain, usually $\mathcal{D}^n$) assigns a set $F(A_1, ..., A_n)$ of 'collective' judgment sets.
- Typically, the output $F(A_1, ..., A_n)$ is a singleton set $\{C\}$, in which case we identify this set with C and write $F(A_1, ..., A_n) = C$.
- If $F(A_1, ..., A_n)$ contains more than one judgment set, there is a 'tie' between these judgment sets.
- An aggregation rule is called *single-valued* or *tie-free* if it always generates a single judgment set.

Aggregation rules

- A standard (single-valued) aggregation rule is *majority rule*, given by

$$F(A_1, ..., A_n) = \{p \in X : |\{i : p \in A_i\}| > n/2\}.$$

- It generates inconsistent collective judgment sets for many agendas and profiles.
- If both individual and collective judgment sets are rational (i.e., in $\mathcal{D}$), the aggregation rule defines a correspondences $\mathcal{D}^n \rightrightarrows \mathcal{D}$, and in the case of single-valuedness a function $\mathcal{D}^n \to \mathcal{D}$.

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

Part 3: Set scoring rules

Part 4: Concluding remarks

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

2.a: Definition

Part 3: Set scoring rules

Part 4: Concluding remarks

Definitions

- Scoring rules are particular aggregation rules, defined on the basis of a *scoring function*.
- A *scoring function* – or simply a *scoring* – is a function $s : X \times \mathcal{D} \to \mathbb{R}$ which to each proposition p and rational judgment set A assigns a number $s_A(p)$, called the *score* of p given A and measuring how p performs ('scores') from the perspective of holding judgment set A.
- For instance, *simple scoring* is given by:

$$s_A(p) = \begin{cases} 1 & \text{if } p \in A \\ 0 & \text{if } p \notin A, \end{cases} \tag{1}$$

- This and many other scorings will be analysed.

Definitions

- A scoring s gives rise to an aggregation rule, called the *scoring rule w.r.t.* s and denoted F_s.
- Given a profile $(A_1, ..., A_n) \in \mathcal{D}^n$, this rule determines the collective judgments by selecting the rational judgment set(s) with the highest sum-total score across all judgments and all individuals:

$$\begin{aligned} F_s(A_1, ..., A_n) &= \text{judgment set(s) in } \mathcal{D} \text{ with highest total score} \\ &= \mathsf{argmax}_{C \in \mathcal{D}} \sum_{p \in C, i \in N} s_{A_i}(p). \end{aligned}$$

- By a 'scoring rule' simpliciter we of course mean an aggregation rule which is a scoring rule w.r.t. *some* scoring.

Definitions

- Different scorings s and s' can generate the same scoring rule $F_s = F_{s'}$, in which case they are called *equivalent*. For instance, s is equivalent to $s' = 2s$.

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

2.b: Simple scoring and the Hamming rule

Part 3: Set scoring rules

Part 4: Concluding remarks

Simple scoring illustrated

- For *simple* scoring (1), the scoring rule works as follows in the face of the 'doctrinal paradox' agenda and profile:

	Score of...									
Individual	p	$\neg p$	q	$\neg q$	r	$\neg r$	pqr	$p\neg q\neg r$	$\neg pq\neg r$	$\neg p\neg q\neg r$
1 (pqr)	1	0	1	0	1	0	3	1	1	0
2 ($p\neg q\neg r$)	1	0	0	1	0	1	1	3	1	2
3 ($\neg pq\neg r$)	0	1	1	0	0	1	1	1	3	2
Group	2	1	2	1	1	2	5*	5*	5*	4

- So, the scoring rule delivers a tie between the premise-based outcome pqr and the conclusion-based outcomes $p\neg q\neg r$ and $\neg pq\neg r$. Formally:

$$F(A_1, A_2, A_3) = \{pqr, p\neg q\neg r, \neg pq\neg r\}.$$

Distance-based rules

- Consider any *distance function* ('metric') d over $\mathcal{D}$.[3]
- The most common example is *Hamming distance* $d = d_{\text{Ham}}$, defined as follows:

$$\begin{aligned} d_{\text{Ham}}(A, B) &= \text{number of judgment reversals} \\ &\qquad \text{needed to transform } A \text{ into } B \qquad (2) \\ &= |A \backslash B| = |B \backslash A| = \frac{1}{2} |A \triangle B| . \end{aligned}$$

E.g., the Hamming-distance between pqr and $p\neg q\neg r$ (for our doctrinal paradox agenda) is 2.

[3]A *distance function* or *metric* over $\mathcal{D}$ is a function $d : \mathcal{D} \times \mathcal{D} \to [0, \infty)$ satisfying three conditions: for all $A, B, C \in \mathcal{D}$, (i) $d(A, B) = 0 \Leftrightarrow A = B$, (ii) $d(A, B) = d(B, A)$ ('symmetry'), and (iii) $d(A, C) \leq d(A, B) + d(B, C)$ ('triangle inequality').

Distance-based rules (cont.)

- The *distance-based rule* w.r.t. a distance d is the aggregation rule F_d which for any profile $(A_1, ..., A_n) \in \mathcal{D}^n$ returns:

$$F_d(A_1, ..., A_n) = \text{judgment set(s) in } \mathcal{D} \text{ with minimal sum-distance to the profile}$$
$$= \text{argmin}_{C \in \mathcal{D}} \sum_{i \in N} d(C, A_i).$$

Distance-based rules

- The most popular example, *Hamming rule* $F_{d_{\mathsf{Ham}}}$, can be characterized as a scoring rule:

Proposition 1 *The simple scoring rule is the Hamming rule.*

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

2.c: Classical scoring rules for preference aggregation

Part 3: Set scoring rules

Part 4: Concluding remarks

Classical scoring

- I now show that our scoring rules generalize the classical scoring rules of preference aggregation theory.
- Consider the preference agenda X for a given set of alternatives K of finite size k.
- Classical scoring rules (such as Borda rule) are defined by assigning scores to alternatives in K, not to propositions xPy in X.
- Given a strict linear order $\succ$ over K, each alternative $x \in K$ is assigned a score $SCO_{\succ}(x) \in \mathbb{R}$.
- The most popular example is of course *Borda scoring*, for which the highest ranked alternative in K scores k, the second-highest $k - 1$, ...

Classical scoring rules

- Given a profile $(\succ_1, ..., \succ_n)$ of individual strict linear orders, the collective ranks the alternatives $x \in X$ according to their sum-total score $\sum_{i \in N} SCO_{\succ_i}(x)$.
- To translate this into the JA formalism, we identify each strict linear order $\succ$ over K with the corresponds judgment set $A \in \mathcal{D}$. So, we write $SCO_A(x)$ instead of $SCO_{\succ}(x)$.
- Formally, I define a *classical scoring* as an arbitrary function $SCO : K \times \mathcal{D} \to \mathbb{R}$.
- What I call the *classical scoring rule* w.r.t. SCO is the JA rule $F \equiv F_{SCO}$ for the preference agenda which for every profile $(A_1, ..., A_n) \in \mathcal{D}^n$ returns:

$$F(A_1, ..., A_n) = \{C \in \mathcal{D} : C \text{ contains all } xPy \in X \\ \text{s.t. } \textstyle\sum_{i \in N} SCO_{A_i}(x) > \sum_{i \in N} SCO_{A_i}(y)\}.$$

Classical scoring and 'our' scoring

- Any given classical (alternative-based) scoring SCO induces a scoring s in our (proposition-based) sense.
- In fact, there are two canonical (and, as we will see, equivalent) ways to define s: one might define s either by

$$s_A(xPy) = SCO_A(x) - SCO_A(y), \tag{3}$$

or, if one would like the lowest achievable score to be zero, by

$$s_A(xPy) = \max\{SCO_A(x) - SCO_A(y), 0\} \tag{4}$$

Proposition 2 *In the case of the preference agenda (for any finite set of alternatives), every classical scoring rule is a scoring rule, namely one with respect to a scoring s derived from the classical scoring SCO via (3) or via (4).*

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

2.d: Reversal scoring and a Borda rule for judgment aggregation

Part 3: Set scoring rules

Part 4: Concluding remarks

Reversal scoring

- Given the agent's judgment set A, let us think of the score of a proposition $p \in X$ as a measure of how 'distant' the negation $\neg p$ is from A; so, p scores high if $\neg p$ is far from A, and low if $\neg p$ is contained in A.
- More precisely, denoting the judgment set arising from A by negating the propositions in a subset $R \subseteq A$ by $A_{\neg R} = (A \backslash R) \cup \{\neg r : r \in R\}$, so-called *reversal scoring* is defined by

$$\begin{aligned} s_A(p) &= \text{nb. of judgment reversals needed to reject } p \qquad (5) \\ &= \min_{R \subseteq A : A_{\neg R} \in \mathcal{D} \,\&\, p \notin A_{\neg R}} |R| = \min_{A' \in \mathcal{D} : p \notin A'} |A \backslash A'| \\ &= \min_{A' \in \mathcal{D} : p \notin A'} d_{\mathrm{Ham}}(A, A'). \qquad (6) \end{aligned}$$

- E.g., a rejected proposition $p \notin A$ scores zero, since A itself contains $\neg p$.

Reversal scoring

- Let's try out reversal scoring for our doctrinal paradox agenda and profile:

	Score of...									
Individual	p	$\neg p$	q	$\neg q$	r	$\neg r$	pqr	$p\neg q\neg r$	$\neg pq\neg r$	$\neg p\neg q\neg r$
1 (pqr)	2	0	2	0	2	0	6	2	2	0
2 ($p\neg q\neg r$)	1	0	0	2	0	2	1	5	2	4
3 ($\neg pq\neg r$)	0	2	1	0	0	2	1	2	5	4
Group	3	2	3	2	2	4	8	9*	9*	8

- E.g., individual 1's judgment set pqr leads to a score of 2 for p, since rejecting p requires negating not just p (as $\neg pqr$ is inconsistent), but also r (where $\neg pq\neg r$ is consistent).
- Notice: a tie between the conclusion-based judgment sets $p\neg q\neg r$ and $\neg pq\neg r$!

Reversal scoring and classical Borda scoring

- The remarkable feature of reversal scoring is its link to classical Borda scoring for the preference agenda:

Remark 1 *In the case of the preference agenda (for any finite set of alternatives), reversal scoring* s *is given by* (4) *with* SCO *defined as classical Borda scoring.*

(See why this is true?)

Reversal scoring *rule* and classical Borda *rule*

- Classical Borda rule is only defined for the preference agenda X, namely as the classical scoring rule w.r.t. Borda scoring SCO.
- Remark 1 and Proposition 2 imply:

Proposition 3 *The reversal scoring rule generalizes Borda rule, i.e., matches it in the case of the preference agenda (for any finite set of alternatives).*

Bill Zwicker's way to generalize Borda rule

- Zwicker (2011) takes an interesting, very different strategy to extending Borda rule.
- The motivation derives from a geometric characterization of Borda preference aggregation obtained by Zwicker (1991).
- Write the agenda as $X = \{p_1, \neg p_1, p_2, \neg p_2, ..., p_m, \neg p_m\}$.
- Each profile gives rise to a vector $\mathbf{v} \equiv (v_1, ..., v_m)$ in $\mathbb{R}^m$ whose j^{th} entry v_j is the *net support for* p_j.
- Zwicker writes the vector $\mathbf{v}$ as an orthogonal sum $\mathbf{v}_{\text{consistent}} + \mathbf{v}_{\text{inconsistent}}$.
- Intuitively, '$\mathbf{v}_{\text{consistent}}$' contains the profile's 'consistent component'.
- Zwicker's Borda-type rule accepts all p_j for which $\mathbf{v}_{\text{consistent},j} > 0$.
- Problem: the decomposition $\mathbf{v}_{\text{consistent}} + \mathbf{v}_{\text{inconsistent}}$ so far 'works' only for special agendas.

Bill Zwicker's way to generalize Borda rule

In summary, there seem to exist two quite different approaches to generalizing Borda:

- Zwicker's approach is geometric and seeks to filter out the profile's 'inconsistent component'.
- My approach
  - retains the principle of score-maximization inherent in Borda aggregation (with scoring now defined at the level of propositions, not alternatives)
  - uses information about someone's *strength* of accepting a proposition (as measured by the score), just as classical Borda rule uses information about *strength* of preference (as measured by classical scores of alternatives).

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

2.e: A generalization of reversal scoring

Part 3: Set scoring rules

Part 4: Concluding remarks

A generalization of reversal scoring

- Recall that reversal scoring s can be characterized in terms of Hamming distance:

$$
\begin{aligned}
s_A(p) &= \text{nb. of judgment reversals needed to reject } p \qquad (7)\\
&= \min_{A' \in \mathcal{D} : p \notin A'} d_{\mathsf{Ham}}(A, A').
\end{aligned}
$$

- More generally, for any given distance function d over $\mathcal{D}$, one might consider the scoring s defined by

$$
\begin{aligned}
s_A(p) &= \text{distance by which one must} \qquad (8)\\
&\quad \text{depart from } A \text{ to reject } p \qquad (9)\\
&= \min_{A' \in \mathcal{D} : p \notin A'} d(A, A').
\end{aligned}
$$

- This yields a whole class of scoring rules, all of which are variants of our judgment-theoretic Borda rule. In the special case of the preference agenda, we thus obtain new variants of classical Borda rule.

Plan

Score as 'logical entrenchment'

- We now consider scoring rules which explicitly exploit the logical structure of the agenda.
- Think of the score of a proposition p ($\in X$) given the judgment set A ($\in \mathcal{D}$) as the degree to which p is logically entrenched in the belief system A, i.e., as the 'strength' with which A entails p.
- We measure this strength by the number of ways in which p is entailed by A, where each 'way' is given by a particular judgment *subset* $S \subseteq A$ which entails p, i.e., for which $S \cup \{\neg p\}$ is inconsistent.
- There are different ways to formalise this idea, depending on precisely which of the judgment subsets that entail p are deemed relevant.

First (naive) attempt

- Let's count *each* judgment subset which entails p as a separate, full-fledged 'way' in which p is entailed.
- This leads to so-called *entailment scoring*, defined by:

$$\begin{aligned} s_A(p) &= \text{number of judgment subsets entailing } p \\ &= \left|\{S \subseteq A : S \text{ entails } p\}\right| . \end{aligned} \tag{10}$$

- Objection: lots of redundancies, i.e., 'multiple counting'.

Second attempt

- To respond to the redundancy objection, let's count two entailments of p as different only if they have no premise in common.
- Formally, define *disjoint-entailment scoring* by:

$$
\begin{aligned}
s_A(p) &= \text{nb. of } \textit{disjoint} \text{ judgment subsets entailing } p \qquad (11)\\
&= \max\{m : A \text{ has } m \text{ disjoint subsets each entailing } p\}.
\end{aligned}
$$

Second attempt: example

- For our doctrinal paradox profile, we get the following disjoint-entailment scores

Individual	Score of...									
	p	$\neg p$	q	$\neg q$	r	$\neg r$	pqr	$p\neg q\neg r$	$\neg pq\neg r$	$\neg p\neg q\neg r$
1 (pqr)	2	0	2	0	2	0	6	2	2	0
2 ($p\neg q\neg r$)	1	0	0	2	0	2	1	5	2	4
3 ($\neg pq\neg r$)	0	2	1	0	0	2	1	2	5	4
Group	3	2	3	2	2	4	8	9*	9*	8

- E.g., individual 2 has judgment set $p\neg q\neg r$, so that p sores 1 (it is entailed by $\{p\}$ but by no other disjoint judgment subset), $\neg q$ scores 2 (it is disjointly entailed by $\{\neg q\}$ and $\{p, \neg r\}$), $\neg r$ scores 2 (it is disjointly entailed by $\{\neg r\}$ and $\{\neg q\}$), and all rejected propositions score zero (they are not entailed by any judgment subsets).

Third attempt

- Our third and fourth attempts aim to avoid 'multiple counting' by counting only those entailments whose sets of premises are *minimal*
- ... with minimality understood either in the sense that no premises can be removed, or in the sense that no premises can be logically weakened.
- To begin with the first sense of minimality, I say that a set *minimally entails* p ($\in X$) if it entails p but no strict subset of it entails p, and I define *minimal-entailment scoring* by

$$s_A(p) = \text{nb. of judgment subsets } \textit{minimally} \text{ entailing } p \tag{12}$$

$$= \left|\{S \subseteq A : S \text{ minimally entails } p\}\right| .$$

Third attempt: example

- Consider again our doctrinal paradox agenda.
- For an individual with judgment set $p\neg q\neg r$,
  - p scores 1 (it is minimally entailed only by $\{p\}$),
  - $\neg q$ scores 2 (it is minimally entailed by $\{\neg q\}$ and by $\{p, \neg r\}$),
  - $\neg r$ scores 2 (it is minimally entailed by $\{\neg r\}$ and by $\{\neg q\}$),
  - all rejected propositions score zero (they are not minimally entailed by any judgment subsets).

Fourth attempt

- To warm up, consider the preference agenda with set of alternatives $K = \{x, y, z, w\}$, and the judgment set $A = \{xPy$, yPz, zPw, xPz, yPw, $xPw\}$ ($\in \mathcal{D}$).
- xPw is entailed by the subset $S = \{xPy, yPz, zPw\}$. This entailment is
  - minimal in the (set-theoretic) sense that we cannot *remove* premises,
  - non-minimal in the (logical) sense that we can *weaken* some of its premises: if we replace xPy and yPz in S by their logical implication xPz, then we obtain a weaker set of premises $S' = \{xPz, zPw\}$ which still entails xPw.

Fourth attempt (cont.)

- In general, a set of propositions is called *weaker* than another one (which is called *stronger*) if the second set entails each member of the first set, but not vice versa.
- A set S ($\subseteq X$) is defined to *irreducibly* (or *logically minimally*) entail p if S entails p, and moreover there is no subset $Y \subsetneq S$ which can be weakened (i.e., for which there is a weaker set $Y' \subseteq X$ such that $(S \backslash Y) \cup Y'$ still entails p).
- Each irreducible entailment is a minimal entailment, as is seen by taking $Y' = \varnothing$.[4]
- *Irreducible-entailment scoring* is of course defined by

$$\begin{aligned} s_A(p) &= \text{nb. of judgment subsets } \textit{irreducibly} \text{ entailing } p \\ &= \left|\{S \subseteq A : S \text{ irreducibly entails } p\}\right| . \end{aligned} \tag{13}$$

[4]Assuming X contains no tautology, i.e., no p such that $\{\neg p\}$ is inconsistent.

Entrenchment-based & reversal scoring

- All our entrenchment-based scorings except the first (naive) one match reversal scoring for our doctrinal paradox example!
- But for many other agendas these scorings all deviate from one another.
- As for the preference agenda:

Proposition 4 *Disjoint-entailment scoring (**??**) and irreducible-entailment scoring (13) match reversal scoring (5) in the case of the preference agenda (for any finite set of alternatives).*

Propositions 3 and 4 jointly have an immediate corollary.

Corollary 1 *The scoring rules w.r.t. scorings (**??**) and (13) both generalize Borda rule, i.e., match it in the case of the preference agenda (for any finite set of alternatives).*

Plan

More scoring rules mentioned

- The premise-based rule can be reconstructed as a scoring rule, in virtue of a scoring which assigns far higher scores to accepted premises than to accepted conclusions.
- The conclusion-based rule can be reconstructed as a scoring rule, in virtue of a scoring which assigns far higher scores to accepted conclusions than to accepted premises.

More scoring rules mentioned

- A quota rule with rational outputs can be reconstructed as a scoring rule.
- A quota rule with sometimes not rational (e.g., inconsistent and/or incomplete) outputs can be 'repaired' by a suitable scoring rule:
  - this scoring rule matches the quota rule whenever the quota rule has a rational output, while rendering the output rational otherwise.

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

Part 3: Set scoring rules

Part 4: Concluding remarks

Plan

Set scoring

- An interesting generalization of scoring rules is obtained by assigning scores directly to entire judgment sets rather than single propositions.
- A *set scoring function* – or simply *set scoring* – is a function $\sigma : \mathcal{D} \times \mathcal{D} \longrightarrow \mathbb{R}$ which to every pair of rational judgment sets C and A assigns a real number $\sigma_A(C)$, the *score* of C given A.
- The most elementary example, to be called *naive* set scoring, is given by

$$\sigma_A(C) = \begin{cases} 1 & \text{if } C = A \\ 0 & \text{if } C \neq A. \end{cases} \tag{14}$$

Set scoring rules

- Any set scoring σ gives rise to an aggregation rule F_σ, the *set scoring rule* (or *generalized scoring rule*) *w.r.t.* σ, which for each profile $(A_1, ..., A_n) \in \mathcal{D}^n$ selects the collective judgment set(s) C in $\mathcal{D}$ having maximal sum-total score across individuals:

$$F_\sigma(A_1, ..., A_n) = \mathsf{argmax}_{C \in \mathcal{D}} \sum_{i \in N} \sigma_{A_i}(C).$$

- An aggregation rule is a *set scoring rule* simpliciter if it is the set scoring rule w.r.t. to some set scoring σ.

Set scoring rules (cont.)

- Set scoring rules generalize ordinary scoring rules, since to any ordinary scoring s corresponds a set scoring σ, given by

$$\sigma_A(C) \equiv \sum_{p \in C} s_A(p),$$

and the ordinary scoring rule w.r.t. s coincides with the set scoring rule w.r.t. σ.

Plan

Naive set scoring and plurality voting

- *Plurality rule* is the aggregation rule F which for every profile $(A_1, ..., A_n) \in \mathcal{D}^n$ returns:

$$\begin{aligned} F(A_1, ..., A_n) &= \text{most frequently submitted judgment set(s)} \\ &= \mathsf{argmax}_{C \in \mathcal{D}} \left| \{ i : A_i = C \} \right| . \end{aligned}$$

- Though normatively questionable (since the internal structure of judgment sets is being ignored), this rule deserves our attention, if only because of its simplicity and the recognized importance of plurality voting in social choice theory more broadly.
- Plurality rule can be construed as a set scoring rule:

Remark 2 *The naive set scoring rule is plurality rule.*

Plan

Distance-based set scoring

- Set scoring rules generalize distance-based aggregation.
- Given an arbitrary distance function d over $\mathcal{D}$, consider what I call *distance-based* set scoring, defined by

$$\sigma_A(C) = -d(C, A). \tag{15}$$

- This renders sum-score-maximization equivalent to sum-distance-minimization:

Remark 3 *For every given distance function over $\mathcal{D}$, the distance-based set scoring rule is the distance-based rule.*

So, all distance-based rules can be modelled as set scoring rules (but not vice versa).

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

Part 3: Set scoring rules

3.d: More that can be done with set scoring

Part 4: Concluding remarks

Further set scoring rules

- Let's take the *epistemic* or *truth-tracking* approach to JA.
- The goal is to reach objectively true collective judgments.
- In a full probabilistic model of votes and the 'unknown truth', one may define:
  - the *maximum-likelihood rule*, which returns collective judgments whose truth would make the profile (the 'data') maximally likely;
  - the *maximum-posterior rule*, which returns the collective judgments whose posterior probability of truth given the profile is maximal.
- Under particular conditions, these rules can be modelled as particular scoring rules.

Plan

Where do we stand?

- Figure 1 summarizes where we stand by depicting different classes of rules (scoring rules, set scoring rules, and distance-based rules) and positioning several concrete rules.[5]

Figure 1: A map of judgment aggregation possibilities

[5]While the positions of most rules in Figure 1 have been established in the paper or follow easily, a few positions are of the order of conjectures. This is so for the placement of our Borda generalization *outside* the class of distance-based rules.

Two possible extensions

Two plausible generalizations of (set) scoring rules:

- Allow scoring to depend on the individual i!
  - This leads to non-anonymous rules.
- Maximize total score within a larger set than the set $\mathcal{D}$ of fully rational judgment sets (such as the set of consistent but possibly incomplete judgment sets)!
  - This leads to 'boundedly rational scoring rules'.