An introduction to ranking rules

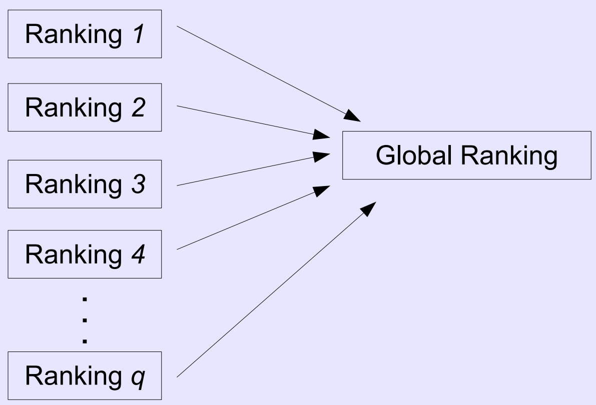
Claude Lamboray

claude.lamboray@statec.etat.lu

1. Definition of the ranking problem

Definition of a ranking rule

A ranking rule is a procedure which aggregates the initial rankings into a global ranking which „best“ combines the information contained in the initial rankings.



A first example: Borda's rule

8	a c b e d
7	e b c d a
4	d c b e a
4	b d e c a
2	c d b e a
POINTS :	***1 2 3 4 5***

Total score for a: 8*1 + 7*5 + 4*5 + 4*5 + 2*5 = 93
Total score for b: 8*3 + 7*2 + 4*3 + 4*1 + 2*3 = 60
Total score for c: 8*2 + 7*3 + 4*2 + 4*4 + 2*1 = 63
Total score for d: 8*5 + 7*4 + 4*1 + 4*2 + 2*2 = 84
Total score for e: 8*4 + 7*1 + 4*4 + 4*3 + 2*4 = 75

Global ranking

b > c > e > d > a

Notations (1)

A linear order O_k is an ordered list of *n* alternatives $\{a_1,a_2,...,a_n\}$ with no ex-aequos.

A liner order *O* can be modelled as a *n* times *n* matrix where

$$O_{ij} = \begin{cases} 1 & \text{if } a_i > a_j \text{ or } i=j \\ 0 & \text{otherwise} \end{cases}$$

If $O_{ij}=1$ and $O_{jk}=1$ then $O_{ik}=1$ (transitive)

$O_{ij}=1$ or $O_{ji}=1$ (complete)

If $O_{ij}=1$ and $O_{ji}=1$ then $i=j$ (antisymmetric)

Notations (2)

A profile $u = (O_1,O_2, \ldots, O_q)$ is a list of *q* linear orders. This list is the input of a ranking rule.

u → f(u)

The output f(u) of a ranking rule can be:

- one or several linear orders
- one weak order (with ex-aequos)

The ranking problem is different from the choice problem which associates with any profile a subset of best alternatives.

2. Rank based rules

Alternative-to-rank matrix

The alternative-to-rank matrix Q_{ij} counts the number of times the alternative a_i is ranked at position j

Q_{ij} = {# rankings: a_i is ranked at the j^{th} position}

8	a c b e d
7	e b c d a
4	d c b e a
4	b d e c a
2	c d b e a

Initial profile

	1	2	3	4	5
a	8	0	0	0	17
b	4	7	14	0	0
c	2	12	7	4	0
d	4	6	0	7	8
e	7	0	4	14	0

Alternative-to-rank matrix

Borda's rule (1)

1. A Borda score is computed for each alternative

$$Score(a_i)=\sum_{j=1}^{n} Q_{ij}\, j$$

2. The alternatives are ranked from the lowest to the largest according to the Borda scores

A generalization of the Borda rule is to use any increasing set of weights representing the ranks:

Let $w_1 < w_2 < \ldots < w_n$, then the scores are defined as follows:

$$Score(a_i)=\sum_{j=1}^{n} Q_{ij}\, w_j$$

Borda's rule (2)

	1	2	3	4	5	**s1**	**s2**
a	8	0	0	0	17	93	144
b	4	7	14	0	0	60	88
c	2	12	7	4	0	63	85
d	4	6	0	7	8	84	122
e	7	0	4	14	0	75	111
w1	1	2	3	4	5		
w2	1	2	5	6	8		

Method 1 (Borda) : b > c > e > d > a

Method 2 : c > b > e > d > a

Given the ordinal nature of the input data, there is no information on how to assign numerical values to the ranks !!

Other rank-based rules

Let r_{ik} *(i=1..n,k=1..q)* be the rank of alternative a_i in ranking *k*.

- Ordering according to the average rank (Borda's rule)

$$score(a_i)=\frac{1}{q}\sum_{k=1}^{q} r_{ik}$$

- Ordering according to the median rank

$$score(a_i)=med(r_{i1}, r_{i2}, \ldots, r_{iq})$$

- Minimizing a distance function (Cook and Seiford)

$$(\hat{c}_1,\ldots,\hat{c}_n)=argmin_c \sum_{k=1}^{q}\sum_{i=1}^{n} |r_{ik}-c_i|$$

3. Pairwise comparison based rules

Majority margins (1)

The majority margin B_{ij} counts the net advantage of an alternative a_i over an alternative a_j.

B_{ij} = {# rankings : $a_i > a_j$} - {# rankings : $a_j > a_i$}

If the profile consists of linear orders, then:

- $B_{ii} = 0$
- $B_{ij} + B_{ji} = 0$ (constant-sum property)

The majority relation M_{ij} checks if there is a majority of rankings which prefer alternative a_i over alternative a_j.

$$M_{ij} = \begin{cases} 1 & \text{if } B_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}$$

Majority margins (2)

Initial profile

1	a b c d
1	b c d a
1	c d a b
1	d a b c
1	d c b a

Majority Margins

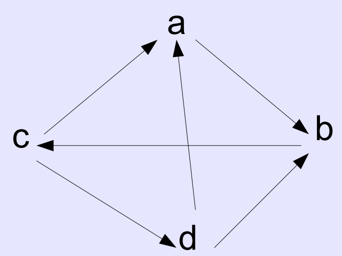
	a	b	c	d
a	0	1	-1	-3
b	-1	0	1	-1
c	1	-1	0	1
d	3	1	-1	0

Majority Relation

	a	b	c	d
a	0	1	0	0
b	0	0	1	0
c	1	0	0	1
d	1	1	0	0

Condorcet cycles

Unfortunately, the majority relation is not always a linear order.



	a	b	c	d
a	0	1	0	0
b	0	0	1	0
c	1	0	0	1
d	1	1	0	0

How should these cyles be broken ?

Copeland's rule (1)

1. A Copeland score is computed for each alternative:

$$Score(a_i)=\sum_{j=1}^{n} M_{ij}+\frac{1}{2}\left|\{j: M_{ij}=0 \wedge M_{ji}=0\}\right|$$

2. The alternatives are ranked from the highest to the lowest according to the Copeland scores

	a	b	c	d	Scores
a	0	1	0	0	1
b	0	0	1	0	1
c	1	0	0	1	2
d	1	1	0	0	2

Copeland ranking:

(c d) > (a b)

Copeland's rule (2)

The Copeland ranking can contain ex-aequos.

If the majority relation is a linear order, then the Copeland ranking corresponds to this linear order.

$O_c = M$

The Copeland ranking rule only depends on the majority relation M.

Kemeny ranking and Slater ranking (1)

The Slater ranking is the linear order which is closest to the majority relation

$$O_S = argmin_{O \in LO} d(M, O)$$

The Kemeny ranking is the linear order which is closest, in average, to the rankings of the input profile.

$$O_k = argmin_{O \in LO} \sum_{k=1}^{q} d(O, O_k)$$

d is the symmetric difference distance.

Kemeny ranking and Slater ranking (2)

Kemeny orders can equivalently be found by solving the following optimization problem:

$$max \sum_{i}^{n} \sum_{j}^{n} B_{ij} O_{ij}$$

s.t.

$$O_{ij}+O_{jk} \leq O_{ik}+1 \qquad \forall i, j, k$$

$$O_{ij}+O_{ji}=1 \qquad \forall i, j (i \neq j)$$

$$O_{ij} \in \{0,1\} \qquad \forall i, j$$

Kemeny ranking and Slater ranking (3)

	a	b	c	d
a	0	1	1	-3
b	-1	0	3	3
c	-1	-3	0	3
d	3	-3	-3	0

bcda	10	cdab	0
abcd	8	dbac	0
bacd	6	cbad	-2
bdac	6	dbca	-2
bcad	4	cdba	-2
bdca	4	acdb	-4
cbda	4	adbc	-4
abdc	2	dacb	-4
acbd	2	cadb	-6
dabc	2	dcab	-6
badc	0	dcba	-8
cabd	0	adcb	-10

Evaluation of a solution according to Kemeny's rule:

f(abcd) = B(a,b) + B(a,c) + B(a,d) + B(b,c)
+ B(b,d) + B(c,d)
= 1 + 1 + (-3) + 3 + 3 + 3
= 8

Kemeny ranking: b > c > d > a

Kemeny ranking and Slater ranking (4)

	a	b	c	d
a	0	1	1	0
b	0	0	1	1
c	0	0	0	1
d	1	0	0	0

abcd	5	bcad	3
abdc	4	cdab	3
acbd	4	dbac	3
bacd	4	bdca	3
dabc	4	cbda	3
bdac	4	adcb	2
bcda	4	cadb	2
acdb	3	dcab	2
adbc	3	cbad	2
badc	3	dbca	2
dacb	3	cdba	2
cabd	3	dcba	1

Evaluation of a solution according to Slater's rule:

f(abcd) = M(a,b) + M(a,c) + M(a,d) + M(b,c)
+ M(b,d) + M(c,d)
= 1 + 1 + 0 + 1 + 1 + 1
= 5

Slater ranking: a > b > c > d

Kemeny ranking and Slater ranking (5)

Slater rankings and Kemeny rankings are not necessarily unique.

Finding Slater rankings or Kemeny orders is an NP-complete problem.

If M (the majority relation) is a linear order, then:

$$O_S = O_K = M$$

The Slater ranking rule only depends on the majority relation M.

The Kemeny ranking rule only depends on the majority margins B.

Kohler's rule (1)

Kohler's rule can be formalized as follows:

At step r (r goes from 1 to n), do the following:

1. Compute for each row of B the smallest value B_{ij}
2. Select the row for which this minimum is maximal
3. Put the selected alternative at rank r
4. Delete the row and column corresponding to the selected alternative from B

Kohler's rule (2)

4	a b c d
4	d c a b
4	c a b d
5	d b c a
1	c b d a
3	b c a d
4	d a b c
2	c d a b
2	b a c d
1	a c d b

	a	b	c	d	min
a		8	-8	-2	-8
b	-8		6	-2	-8
c	8	-6		4	-6
d	2	2	-4		-4

	a	b	c	min
a		8	-8	-8
b	-8		6	-8
c	8	-6		-6

	a	b	min
a		8	8
b	-8		-8

Kohler's ranking : d > c > a > b

Kohler's rule (3)

A Kohler ranking is not necessarily unique.

Kohler's rule is an example of a „ranking by choosing“ procedure:

- A choice procedure determines the best alternative.
- The alternative is ranked first and removed from the set.
- The choice procedure is reapplied until all alternatives have been ranked

If the majority relation is a linear order, then the Kohler's ranking corresponds to this linear order.

$O_{KO} = M$

Kohler's rule only depends on the **order** of the majority margins.

Ranked Pairs rule (1)

The Ranked Pairs rule can be formalized as follows:

1. Rank the ordered pairs (a_i,a_j) according to their majority margins.

2. Consider the pairs in that order and do the following:

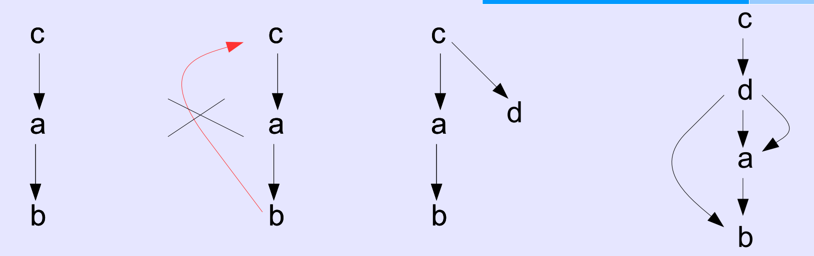
- If the ordered pair does not create a cycle with the pairs already blocked, block this pair.

- If the ordered pair creates a cycle with the pairs already blocked, skip this pair.

Ranked Pairs rule (2)

	a	b	c	d
a		8	-8	-2
b	-8		6	-2
c	8	-6		4
d	2	2	-4	

(c,a) (a,b)	8
(b,c)	6
(c,d)	4
(d,a) (d,b)	2
(a,d) (b,d)	-2
(d,c)	-4
(c,b)	-6
(a,c) (b,a)	-8



Ranked Pairs ranking : c > d > a > b

Ranked Pairs rule (3)

The Ranked Pairs does not necessarily lead to one unique ranking.

If the majority relation is a linear order, then the Ranked Pairs ranking corresponds to this linear order.

$O_{RP} = M$

The Ranked Pairs rule only depends on the **order** of the majority margins.

4. A classification of ranking rules

Condorcet consitency (1)

A ranking rule is Condorcet consistent if the following holds:

If the majority relation is a linear order, then this linear order is the unique solution of the ranking rule.

Condorcet-consistency

Kemeny
Slater
Copeland
Kohler
Ranked Pairs

No Condorcet-consistency

Borda

Condorcet consitency (2)

8	a c b e d
7	e b c d a
4	d c b e a
4	b d e c a
2	c d b e a

Borda ranking : b > c > e > d > a

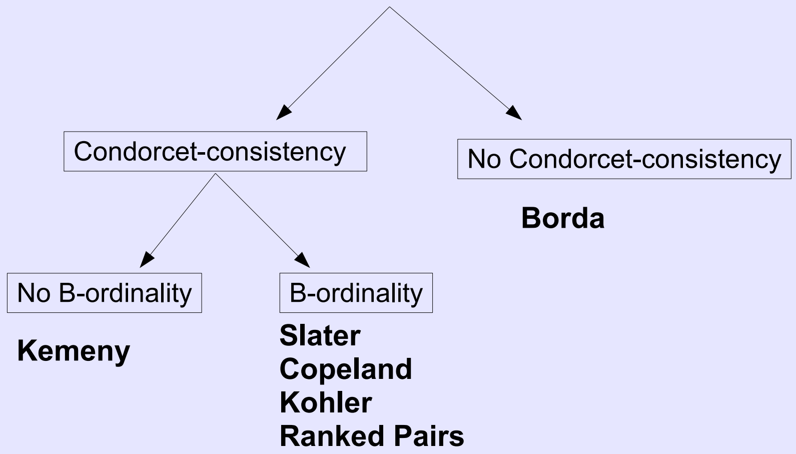
There is a majority of rankings which prefer:

- c over b (14 out of 25)
- c over e (14 out of 25)
- c over d (17 out of 25)
- c over a (17 out of 25)
- b over e (18 out of 25)
- b over d (19 out of 25)
- b over a (17 out of 25)
- e over d (15 out of 25)
- e over a (17 out of 25)
- d over a (17 out of 25)

Majority ranking: c > b > e > d > a

B-ordinality (1)

A ranking rule is B-ordinal if the result only depends on the order of the majority margins.



B-ordinality (2)

	a	b	c	d
a		1	1	-3
b	-1		3	3
c	-1	-3		3
d	3	-3	-3	

Kemeny ranking: b > c > d > a

	a	b	c	d
a		3	3	-5
b	-3		5	5
c	-3	-5		5
d	5	-5	-5	

Kemeny ranking: a > b > c > d

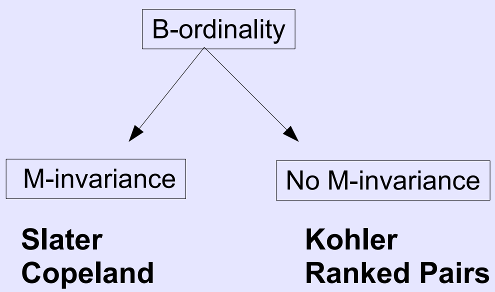
For both profiles, the order of the pairwise majority margins is the same:

(b,c), (b,d), (c,d), (d,a)	3	5
(a,b), (a,c)	1	3
(b,a), (c,a)	-1	-3
(c,b), (d,b), (d,c), (a,d)	-3	-5

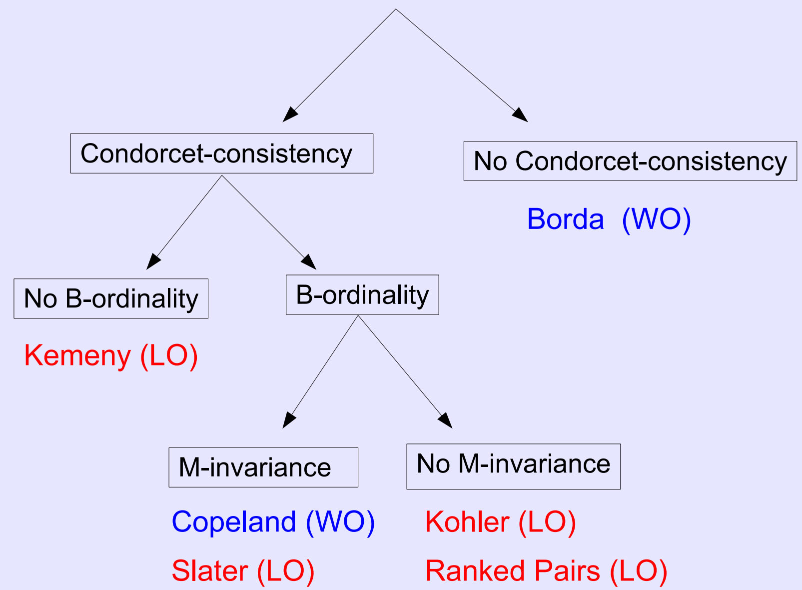
M-invariance

A ranking rule is M-invariant if the result only depends on the majority relation.

Remark: If a ranking rule is M-invariant then it is B-ordinal



Classification of ranking rules



WO : solution is a weak order
LO : solution is one or more linear orders

Which ranking rule should we use? (1)

There is no perfect ranking rule (cf Arrow's theorem)!

Some approaches to evaluate a ranking rule:

- Looking at the properties of a ranking rule (which properties are verified, how often are they violated,..)

- Axiomatic characterizations (the ranking rule is the only ranking rule which verifies a certain combination of properties)

Which ranking rule should we use? (2)

- Comparing the results of two ranking rules

  - the result of a ranking rule 1 can be the opposite of the result of ranking rule 2

  - how often do the results of two ranking rules coincide (and for what profiles)

- Computational complexity

- Transparency of the ranking rule

Generalizations

- More complex preference structures than linear orders as an input

- Other outputs than only a linear (or weak) order

- Introducing weights on the input criteria