Soc Choice Welfare (2004) 23: 161–185
DOI: 10.1007/s00355-003-0242-x

Social Choice and Welfare

Maximum likelihood approach to vote aggregation with variable probabilities

Mohamed Drissi-Bakhkhat[1], **Michel Truchon**[2]

[1]Department of Economics, University Abdelmalek Essaadi, Tangier, Morocco (e-mail: mohamed.drissi@menara.ma)
[2]CIRPÉE, CIRANO and Département d'économique, Université Laval, Québec, Qc, Canada G1K 7P4 (e-mail: mtru@ecn.ulaval.ca)

Received: 23 November 2002/Accepted: 23 April 2003

Abstract. The Condorcet-Kemeny-Young statistical approach to vote aggregation is based on the assumption that voters have the same probability of comparing correctly two alternatives and that this probability is the same for any pair of alternatives. We relax the second part of this assumption by letting the probability of comparing correctly two alternatives be increasing with the distance between two alternatives in the allegedly true ranking. This leads to a rule in which the majority in favor of one alternative against another one is given a larger weight the larger the distance between the two alternatives in the true ranking, i.e., the larger the probability that the voters compare them correctly. This rule is not Condorcet consistent and does not satisfy local independence of irrelevant alternatives. Yet, it is anonymous, neutral, and paretian. It also appears that its performance in selecting the alternative most likely to be the best improves with the rate at which the probability increases.

1 Introduction

Condorcet (1785) showed that with the simple majority rule, a group of individuals will choose the better of two alternatives more frequently than a single individual. He also studied the relation between the number of voters and the probability of selecting the best alternative. The larger the number of voters, the larger this probability. This result, known as Condorcet's jury

We would like to thank Michel Le Breton for his encouragement to examine this question and for his comments, as well as Philippe De Donder, Jean-Yves Duclos, Stephen Gordon, Cyril Téjédo and an anonymous referee for their comments.

theorem, is obtained under the assumption that the voters have the same competence level, i.e. the same probability, above $\frac{1}{2}$, of choosing the right alternative.

Many authors have worked on relaxing this assumption. Nitzan and Paroush (1982) show that the maximum likelihood rule, when the voters have unequal competence, is a weighted majority rule, where the weight of each individual's vote is a function of the competence level of the voter. Shapley and Grofman (1984) study the case of correlated votes. They provide a necessary and sufficient condition for the weighted rule to be optimal. These results, along with other extensions of the jury theorem, are gathered in Grofman et al. (1983). More recent versions of the theorem are presented by Berg (1993, 1994) and Ladha (1992, 1993, 1995) for the case of correlated individual competencies and by Paroush (1998) for heterogeneous competencies.

Ben-Yashar and Nitzan (1997) present a more general approach in which the optimal rule maximizes the expected benefit resulting from the choice of an alternative. Ben-Yashar and Paroush (2001) and Rata (2002a) extend their approach to more than two alternatives but the objective remains the choice of a single alternative. In many situations, we need a complete ranking of all alternatives, i.e. an aggregation rule. This is the case in competitions for grants and scholarships, in judged sports such as figure skating, in projects ranking, etc.

A possible avenue suggested by Ben-Yashar and Paroush (2001) to deal with this more general problem is to let the choice set be the one of all rankings of the alternatives. However, this idea leaves open the question of how to model the competence of the voters or judges in choosing the correct order, and more specifically how to model the probability of an incorrect ranking. Another avenue, which we pursue here, is the binary approach to vote aggregation also initiated by Condorcet (1785). This method consists of ordering alternatives pairwise according to the majority rule. Condorcet showed that, if this procedure yields an order on the set of alternatives, this order is the most probable under a uniform competence parameter. This was one of the first applications of the maximum likelihood principle.

Condorcet was well aware that the binary relation resulting from his procedure may contain cycles. He proposed a method for breaking these cycles, but unfortunately, this method gives consistent results only for the case of three alternatives. Young (1988) shows that a correct application of the maximum likelihood principle leads to the selection of rankings that have the minimal total number of disagreements with those of the voters. In other words, these rankings minimize a "distance" proposed by Kemeny (1959) and for this reason, they are often given the name of Kemeny. When it exists, i.e., in the absence of cycles in the majority relation, the Condorcet ranking is the unique Kemeny ranking.

The Condorcet-Kemeny-Young approach is based on the assumption that voters have the same probability of comparing correctly two alternatives and that this probability is the same for any pair of alternatives. In this paper, we relax the second part of this assumption. We let the probability of comparing correctly two alternatives be increasing with the distance between two alternatives in the true ranking. This reflects the possibility that voters or judges may have a better chance of correctly ranking two alternatives when one is very good and the other very bad, than when facing two similar alternatives.

As we shall point out, there is some inconsistency in the binary approach: a judge's vote may be cyclical with positive probability. Nevertheless, we remain within that framework, which has received much attention in the literature, since our purpose is to verify the dependency of the results obtained so far on the assumption of constant probability across pairs of alternatives.

Our assumption leads to a rule in which the majority in favour of one alternative against another one receives a larger weight as the distance between the two alternatives in the true ranking increases.[1] This result is comparable to those of Nitzan and Paroush (1982) and of Grofman et al. (1983). They obtain a rule in which weights vary across voters, as a result of unequal competencies of the voters, while in our rule, weights may vary across pairs of alternatives because of different probabilities of ordering correctly alternatives in different pairs.

Our rule is not Condorcet consistent. It does not necessarily select the Condorcet ranking nor the Condorcet winner, when they exist. Actually, the selected ranking may change with the specification of the probabilities. Therefore, this rule is different from the Kemeny rule. However, it is anonymous, neutral, and paretian as is the Kemeny rule but contrary to the latter, it does not satisfy a property that Young and Levenglick (1978) call "local independence of irrelevant alternatives".

Condorcet also hinted that the Condorcet winner, or the top alternative in the Condorcet ranking, is not necessarily the most likely to be the best. Young (1988) confirms that indeed with a constant probability close to $\frac{1}{2}$, the alternative most likely to be the best is the Borda winner, while when the probability is close to 1, it is an alternative whose smallest majority against other alternatives is at least as large as the smallest majority of any other alternative. We extend his analysis to the case of variable probabilities.

Young's result implies that the Kemeny rule does not necessarily select the alternative most likely to be the best. A natural question that comes to mind is whether or not the rule obtained with variable probabilities does better than the Kemeny rule in this respect. It appears that this performance improves with the rate at which the probability increases.

The structure of the paper is the following. The notation, the basic assumptions, the voting procedure, and some useful concepts from social choice theory are presented in Sect. 2. The maximum likelihood approach for

[1] The criterion to choose an order remains the maximum likelihood principle. See Drissi-Bakhkhat (2002) for an expected loss approach, which could be compared to the expected benefit model of Ben-Yashar and Paroush (2001).

variable probabilities is described in Sect. 3. Some properties of the rule arising from the general specification of the probabilities are discussed in Sect. 4. We show that the most likely ranking depends on the distribution of the probabilities. In particular, the most likely ranking is not necessarily the Condorcet ranking. Starting with Sect. 5, we present a systematic analysis of the case of three alternatives for a particular class of increasing probability functions. The characterization of the most likely rankings and of the alternatives most likely to be the best are provided in this section and Sect. 6 respectively. This last section also deals with the performance of the choice rule resulting from the maximum likelihood approach in terms of the alternative most likely to be the best. A brief conclusion is presented as a final section.

2 The social choice problem

Let $X = \{a, b, c, \ldots\}$ be the set of *alternatives* or *candidates* to be ranked. The cardinality of X is m. We denote by $\mathcal{B}$ the set of complete and asymmetric binary relations on X and by $\mathcal{R}$ the subset of complete orders on X also called rankings. A complete order on X can be represented by a permutation of the elements of the vector $(1, 2, \ldots, m)$, denoted $r = (r_a, r_b, r_c, \ldots)$, where r_a is the rank of a, r_b the rank of b, and so on. Equivalently, an order can be represented by a sequence $s_1 s_2 s_3 \ldots$ where $s_1, s_2, s_3 \ldots$ are respectively the alternatives with ranks $1, 2, 3, \ldots$.

There is a set $I = \{1, 2, \ldots, n\}$ of *voters* or *judges*. Each is asked to compare the alternatives pair by pair, as in the Condorcet procedure. His or her vote is summarized in a matrix $N^i = \left[v^i_{st}\right]_{s,t \in X}$. For any pair of alternatives $(s, t) \in X^2$, $v^i_{st} = 1$ if voter i chooses s over t and $v^i_{st} = 0$ otherwise, and $v^i_{st} = 0$ if $s = t$. Alternatively, we can ask each voter i to fill in N^i according to the previous convention. Each N^i is an element of $\mathcal{B}$. Since only the aggregate information will ever be needed, we define a *poll* by

$$N = \sum_{i=1}^{n} N^i$$

and we let $\mathcal{N}$ be the set of possible polls on X.

Once the voters or judges have expressed their opinions in a poll, the problem is to aggregate these opinions in order to select a final ranking. We formalize this idea in the following definition and give examples of aggregation rules, before turning to the maximum likelihood approach.

Definition 1. *An aggregation rule is a correspondence $FR : \mathcal{N} \to \mathcal{R}$ that assigns to each poll N, a final ranking or a subset of final rankings $FR(N)$ of the alternatives.*

Consider now the correspondence $FR^M : \mathcal{N} \to \mathcal{B}$ defined by $sFR^M(N)t \Leftrightarrow v_{st} > v_{ts}$. The binary relation $FR^M(N)$ is the *majority relation* issued from the poll N. We assume that $FR^M(N)$ is complete, which is always the case when n is odd. However, it is not necessarily transitive: it may contain cycles.

Condorcet advocated the use of FR^M as an aggregation rule when $FR^M(N)$ is an order. He also proposed a method for breaking eventual cycles in $FR^M(N)$, which does not work when there are more than three alternatives. We shall return to this point later.

Definition 2. *Given a poll* N, if $FR^M(N)$ *is an order, we call this order the Condorcet ranking.*

Another well known example of an aggregation rule is due to Borda (1784). It is a scoring method with the vector of scores $(m-1, m-2, \ldots, 2, 1, 0)$. An alternative receives $m-1$ points if it is ranked first by a voter, $m-2$ if it is ranked second, ..., and 0 points if it is last. The scores of each alternative are then aggregated across voters and alternatives are ordered according to the sums of these scores. It can be checked that the sum of the scores for alternative s is equal to $\sum_{\tau \in X} v_{s\tau}$. This prompts the following definition.

Definition 3. *Given a poll* N, *we shall say that a ranking* $s_1 s_2 s_3 \ldots$ *such that:*

$$\sum_{\tau \in X} v_{s_1,\tau} \geq \sum_{\tau \in X} v_{s_2,\tau} \geq \sum_{\tau \in X} v_{s_3,\tau} \geq \cdots \tag{1}$$

is a Borda ranking. The Borda rule is the correspondence $FR^B : \mathscr{N} \to \mathscr{R}$ *that assigns to each poll* N, *the set of Borda rankings. The number* $\sum_{\tau \in X} v_{s\tau}$ *is called the Borda score of* s.

Remark 1. Strictly speaking, this definition is correct only if each individual vote in the poll is transitive, which we do not assume, to be consistent with the binary approach. This abuse of terminology will simplify the presentation throughout the paper.

Any aggregation rule FR induces a *social choice correspondence* that assigns to each poll N, the subset of alternatives that are top ranked in at least one ranking in $FR(N)$. Accordingly, we have the following concepts.

Definition 4. *A Borda winner for a poll* N *is the top ranked alternative in at least one Borda ranking.*

Definition 5. *The Condorcet winner for a poll* N, *if it exists, is the alternative* s *that satisfies* $v_{st} > v_{ts}\ \forall t \neq s$.

We also have the following type of winner even if we do not define a corresponding aggregation rule.

Definition 6. *A Kramer-Simpson winner for a poll* N *is an alternative* s *such that* $\max_\tau v_{\tau s} \leq \max_\tau v_{\tau u}\ \forall u \in X$ *or, equivalently,* $\min_\tau v_{s\tau} \geq \min_\tau v_{u\tau}\ \forall u \in X$.

Definition 7. *An aggregation rule $FR : \mathcal{N} \to \mathcal{R}$ has the Condorcet property if $FR(N) = FR^M(N)$ for every poll N such that $FR^M(N)$ is an order. Similarly, a social choice correspondence has the Condorcet property if it selects exclusively the Condorcet winner when it exists.*

It is well known that the Borda rule does not have the Condorcet property. It does not necessarily produce a ranking that has the Condorcet winner as the top alternative.

3 The maximum likelihood approach

The maximum likelihood approach to voting was initiated by Condorcet (1785). It starts with the assumption that there exists a true ranking $r \in \mathcal{R}$ on the set of alternatives. The true ranking, however, is not known. Experts, i.e. voters or judges, are then asked to provide their opinion as to what should be considered the true ranking. Their opinions are collected in N^i, $i = 1, \ldots, n$.

Assuming that every voter has the same probability (larger than $\frac{1}{2}$) of correctly comparing any two alternatives, Condorcet showed that if the binary relation $FR^M(N)$ is an order, then it is the most likely ranking. He also offered indications on how to break the cycles that $FR^M(N)$ might contain. Unfortunately, these indications yield consistent results only for the case of three alternatives. Young (1988) shows what a correct application of the maximum likelihood principle leads to.

It is this approach that we pursue here with a more flexible representation of the competence of the experts. The vote of expert i on a pair of alternatives (s, t) is a random variable $\bar{v}^i_{st} \in \{0, 1\}$, conditional on the true ranking r. Again, each voter has the same probability of ranking correctly two alternatives and this probability is the same for any two couples of alternatives $(s, t), (u, v) \in X^2$ such that $r_s - r_t = r_u - r_v$. However, this probability is a non-decreasing function of the distance between the two alternatives in the true ranking.

More precisely, let P^m be the class of non-decreasing functions $p : \{1, \ldots, m-1\} \to \left(\frac{1}{2}, 1\right)$. Given a ranking $r \in \mathcal{R}$, a function $p \in P^m$, and two alternatives $s, t \in X$ such that $r_s < r_t$, the conditional distributions of $\bar{v}^i_{st}$ and $\bar{v}^i_{ts}$ are defined by:

$$\Pr\left(\bar{v}^i_{st} = 1 \mid r\right) = \Pr\left(\bar{v}^i_{ts} = 0 \mid r\right) = p(r_t - r_s)$$
$$\Pr\left(\bar{v}^i_{ts} = 1 \mid r\right) = \Pr\left(\bar{v}^i_{st} = 0 \mid r\right) = 1 - p(r_t - r_s)$$

The probabilities are assumed to be strictly less than unity to avoid degenerate distributions. This distribution on pairs of alternatives induces a conditional probability distribution on the binary relations $N^i \in \mathcal{B}$:

$$\Pr\left(N^i \mid r\right) = \prod_{\substack{s,t \in X \\ r_s < r_t}} \Pr\left(\bar{v}^i_{st} = v^i_{st} \mid r\right)$$

Remark 2. At this stage, it is important to point out an important fact that is not always mentioned in the literature on the binary approach. Given the statistical independence of votes across pairs of alternatives, cyclical relations in $\mathscr{B}$ have a positive probability, even if r is an order. For example, given the true ranking abc on $\{a, b, c\}$, we have

$$\Pr\left(v_{ab}^i = 1 \mid abc\right) \Pr\left(v_{bc}^i = 1 \mid abc\right) \Pr\left(v_{ca}^i = 1 \mid abc\right) > 0$$

since $p(2) < 1$. There is thus a positive conditional probability that a judge's vote be cyclical. This is the reason for not assuming that individual votes N^i are transitive. Interestingly, making $p(2)$ larger than $p(1)$ contributes in diminishing the probability of observing a cycle.

We focus now on the aggregate N. The entries of N are random variables $\bar{v}_{st} \in \{0, 1, \ldots, n\}$. Since the probabilities $\Pr\left(\bar{v}_{st}^i = 1 \mid r\right)$ are independent and the same for all voters, each random variable $\bar{v}_{st} \equiv \sum_{i=1}^n \bar{v}_{st}^i$ has a binomial distribution defined by:

$$\forall \mu \in \{1, 2, \ldots, n\} : \Pr(\bar{v}_{st} = \mu \mid r) = \binom{n}{\mu} \Pr\left(\bar{v}_{st}^i = 1 \mid r\right)^{\mu} \Pr\left(\bar{v}_{st}^i = 0 \mid r\right)^{n-\mu}$$

for any $i \in I$. With the independence assumption across pairs of alternatives, the probability of a poll N is the product of the probabilities of its elements:

$$\Pr(N \mid r) = \prod_{\substack{s,t \in X \\ r_s < r_t}} \Pr(\bar{v}_{st} = v_{st} \mid r) \tag{2}$$

This is the *likelihood function* of poll N, given order r. As an illustration, consider the set $X = \{a, b, c\}$ and the ranking abc. We have:

$$\Pr(N \mid abc) = \kappa p(1)^{v_{ab}}(1 - p(1))^{v_{ba}} p(1)^{v_{bc}}(1 - p(1))^{v_{cb}} p(2)^{v_{ac}}(1 - p(2))^{v_{ca}}$$

with $\kappa = \frac{n!^3}{v_{ab}! v_{ba}! v_{bc}! v_{cb}! v_{ac}! v_{ca}!}$. Note that this term is independent of the function $p(\cdot)$.

In the maximum likelihood approach, we are interested in a ranking r^* (not necessarily unique) that maximizes the likelihood function $\Pr(N \mid r)$ of poll N. Equivalently, this ranking maximizes the posterior probability $\Pr(r \mid N)$, conditional on N. These posterior probabilities are obtained by applying Bayes' rule under the assumption of equal prior for rankings:

$$\begin{aligned} \Pr(r \mid N) &= \frac{\Pr(r \wedge N)}{\Pr(N)} = \frac{\Pr(r) \Pr(N \mid r)}{\Pr(N)} \\ &= \frac{\Pr(r) \Pr(N \mid r)}{\sum_{q \in \mathscr{R}} \Pr(q) \Pr(N \mid q)} = \frac{\Pr(N \mid r)}{\sum_{q \in \mathscr{R}} \Pr(N \mid q)} \end{aligned} \tag{3}$$

Under the assumption of a constant prior for rankings, the prior probability that an alternative s be ranked ahead of another alternative t is then exactly $\frac{1}{2}$, for any pair (s, t). It is clear that for the true ranking r, $\Pr(r \mid N) \to 1$ when $n \to \infty$. This is also an implication of the Condorcet *jury theorem*.

Note that $\Pr(r \mid N)$ is an increasing transformation of $\Pr(N \mid r)$. The following other transformation will prove useful in deriving some of the results. Given a probability function $p \in P^m$, let

$$L_p(k) = \ln\left(\frac{p(k)}{1 - p(k)}\right) \quad \forall k \in \{1, \ldots, m-1\} \tag{4}$$

and for every ranking r and every poll N, consider the logarithmic transformation of (2):

$$M_p(r; N) = \sum_{k=1}^{m-1} \left[L_p(k) \sum_{\substack{s,t \in X \\ r_t = r_s + k}} v_{st}(N) \right]$$

Then, the *most likely rankings* or the *most probable rankings* are the elements of the set:

$$r_p(N) = \arg\max_{r \in \mathscr{R}} M_p(r; N)$$

Clearly, r_p is another aggregation rule.

Definition 8. *Given a function $p \in P^m, r_p : \mathscr{N} \to \mathscr{R}$ is the aggregation rule that assigns to each poll N the set of most likely rankings, i.e. the set $r_p(N) \subset \mathscr{R}$.*

r_p induces a social choice correspondence that assigns to each poll N, the subset of alternatives that are top ranked in at least one ranking in $r_p(N)$.

Definition 9. $\Gamma_p : \mathscr{N} \to X$ *defined by:*

$$\Gamma_p(N) = \left\{ s \in X : \exists r \in r_p(N) : r_s = 1 \right\}$$

is the social choice correspondence induced by r_p.

The case where $p(k) = \bar{p}$ with $\bar{p} \in \left(\frac{1}{2}, 1\right)$ is the one studied by Young. He shows that a ranking that has the greatest posterior probability is a ranking with minimum "distance" from the poll. The distance here is the one proposed by Kemeny (1959). Hence, with constant probability, the most likely rankings are often called Kemeny orders. The distance is defined as follows. Let $\gamma_{st} : \mathscr{R}^2 \to \mathbb{R}$, be a function defined for every couple of alternatives (s, t) and every couple of weak orders $(\hat{r}, r)$ by:

$$\gamma_{st}(\hat{r}, r) = \begin{cases} 1 & \text{if } \hat{r}_s < \hat{r}_t \text{ and } r_s > r_t \\ \frac{1}{2} & \text{if } \hat{r}_s = \hat{r}_t \text{ and } r_s < r_t \\ 0 & \text{otherwise} \end{cases}$$

Then, define the function $c^K : \mathscr{R}^2 \to \mathbb{R}$ by:

$$c^K(\hat{r}, r) = \sum_{s \in X} \sum_{\substack{t \in X \\ t > s}} \gamma_{st}(\hat{r}, r) \tag{5}$$

Note that c^K is a metric on the set of weak orders on X. The Kemeny "distance" δ^K between a ranking (or transitive binary relation) and a poll is defined by:

$$\delta^K(r,N) = \sum_{i \in I} c^K(N^i, r)$$

Definition 10 The *Kemeny rule is the correspondence* $FR^K : \mathcal{N} \to \mathcal{R}$ *that assigns to each poll* N, *the subset* $FR^K(N) = \arg\min_{r \in \mathcal{R}} \delta^K(r,N)$. *The elements of* $FR^K(N)$ *are the Kemeny orders.*

Remark 3. When $p(k) = \bar{p}$ with $\bar{p} \in \left(\frac{1}{2}, 1\right)$, the most likely orders are the Kemeny orders, i.e. $r_p(N) = FR^K(N)$.

When $p(k) = \bar{p}$ with $\bar{p} \in \left(\frac{1}{2}, 1\right)$, $L_p(k)$ is a constant that we may as well replace by 1. Then, $M_p(r;N)$ becomes:

$$K(r;N) = \sum_{\substack{s,t \in X \\ r_s < r_t}} v_{st}(N)$$

The value of $K(r;N)$, called the *Kemeny score* of order r given the poll N, is the total number of agreements between order r and the individual orders making up profile N. Thus, with a constant probability, the maximum likelihood approach consists in finding an order r maximizing $K(r;N)$.

4 Properties of r_p and Γ_p

The question that we now address is whether the aggregation rule r_p and the social choice correspondence Γ_p give results that are significantly different from those of the Kemeny rule when the function p is significantly different from $p(k) = \bar{p}$. We know that the Kemeny rule has the Condorcet property. We show in the following proposition that this property is not generally preserved by r_p and Γ_p. Therefore, r_p and Γ_p may give results different from those of the Kemeny rule.

Proposition 1. *The rule* r_p *and the correspondence* Γ_p *do not have the Condorcet property for every probability function* p.

Proof. Consider the following poll N on $X = \{a,b,c,d\}$ with $n = 9$:

Alternatives	*a*	b	c	d
a	–	5	5	5
b	4	–	5	7
c	4	4	–	6
d	4	2	3	–

Note that *a* and *abcd* are the Condorcet winner and ranking respectively. The Borda ranking is *bacd*[2].

However, with $p(k) = \bar{p}^{1/k}$ and $\bar{p} < 0.7048$, the most likely ranking is *bacd*. ■

In the previous example, *abcd* is the most likely ranking for $\bar{p} > 0.7049$. How is it that $r_p(N)$ is the Condorcet ranking and $\Gamma_p(N)$ the Condorcet winner only if the probability $\bar{p}$ is sufficiently high? Why is *bacd* more likely than *abcd* for $\bar{p}$ sufficiently small? Note that the Kemeny score $K(abcd; N)$ is the sum of the numbers above the diagonal of the matrix N while $K(bacd; N)$ is the sum of the numbers above the diagonal of the following matrix, obtained by permuting the rows and columns corresponding to *a* and *b* in the previous matrix:

Alternatives	*b*	a	c	d
b	–	4	5	7
a	5	–	5	5
c	4	4	–	6
d	2	4	3	–

Since the first sum (33) is larger than the second (32), *abcd* beats *bacd* under the Kemeny rule. By a continuity argument, this is also the case under r_p with the increasing function $p(k) = \bar{p}^{1/k}$ and with $\bar{p}$ sufficiently close to 1. Note that a larger $\bar{p}$ means a lower increase rate of the probability with respect to k, and hence a probability function closer to a constant.

With an increasing function p, the farther the numbers from the diagonal the larger the weights they receive. With $\bar{p} = 0.51$, the numbers just above the diagonal are given the weight 0.04 while those that are 2 and 3 positions away from the diagonal are given the weights 0.9156 and 1.3798 respectively. With the latter weight, the $v_{bd} = 7$, which replaces the $v_{ad} = 5$ of the first table, more than compensates for the replacements of the sum $v_{ac} + v_{bd} = 12$ by $v_{bc} + v_{ad} = 10$ and of the sum $v_{ab} + v_{bc} + v_{cd} = 16$ by $v_{ba} + v_{ac} + v_{cd} = 15$. This is why *bacd* beats *abcd*.

With this example, one can also see the logic behind these variable weights. With $\bar{p} = 0.51$, the voters have roughly the same probability of ordering correctly or incorrectly two adjacent alternatives in a ranking. Thus, we should not give too much importance to the majority that an alternative obtains against an adjacent one. This is not the case for alternatives that are farther apart in the true ranking. For instance, the majority of 7 votes for *b* against 2 for *d* in the above poll is given more importance when *bacd* is taken as the true ranking because the voters who rank *b* before *d* have a larger

[2]This is the true Borda ranking since there exists a profile of 9 transitive votes summing to N.

probability of voting correctly: $0.51^{1/3} \simeq 0.8$ compared to 0.51 for pairs of alternative just above the diagonal.

The rule r_p and the correspondence Γ_p being generally different from the Kemeny rule and correspondence, the question now is whether r_p and Γ_p verify properties, other than the Condorcet property, that are satisfied by the Kemeny rule and correspondence. The latter satisfy anonymity and neutrality. The Kemeny correspondence is also paretian. We shall see that these properties are still satisfied by the more general rule r_p and correspondence Γ_p. In the case of the Pareto principle, we shall establish the result for transitive votes. However, r_p no longer satisfies a condition that Young and Levenglick (1978) call "local independence of irrelevant alternatives" , which is satisfied by the Kemeny rule.

Anonymity or symmetry. Clearly, r_p and Γ_p are both *symmetric* for every $p \in P^m$: individual votes are treated identically by either function.

Neutrality. It is also obvious that r_p and Γ_p are *neutral*: the alternatives are treated identically. Names do not matter.

Weak Pareto principle. An alternative is a *weak Pareto optimum* if there is no other alternative that the voters unanimously prefer. The *weak Pareto principle* applied to Γ_p says that, for every poll N, $\Gamma_p(N)$ contains only weak Pareto optima:

$$\forall N \in \mathcal{N},\ \forall p \in P^m,\ \forall t \in \Gamma_p(N),\ \nexists s \in X : r_s^i < r_t^i\ \forall i \in I$$

Note that this principle does not imply that $\Gamma_p(N)$ contains all weak Pareto optima. Actually, $\Gamma_p(N)$ selects one (or more) of these optima. The next lemma will be useful to prove that Γ_p satisfies the weak Pareto principle.

Lemma 2. *Let $X = \{1, 2, \ldots, s, \ldots, m\}$ and N be a poll such that $v_{s1} = n$ and $v_{sh} \geq v_{1h}\ \forall h \neq 1, s$. Next, consider the rankings $r = (1, 2, \ldots, s, \ldots, m)$ and $\hat{r} = (s, 2, \ldots, 1, \ldots, m)$, where $\hat{r}$ is obtained by interchanging 1 and s in r. Then, $M_p(\hat{r}; N) > M_p(r; N)$.*

Proof. It is cumbersome but nonetheless straightforward to verify that:

$$\begin{aligned} M_p(\hat{r}; N) - M_p(r; N) &= L_p(s-1)(v_{s1} - v_{1s}) + \sum_{h=1}^{s-2} L_p(h)(v_{s-h,1} - v_{s-h,s}) \\ &\quad + \sum_{h=1}^{m-1} L_p(h)(v_{s,h+1} - v_{1,h+1}) \\ &\quad + \sum_{h=s}^{m-1} L_p(h+1-s)(v_{1,h+1} - v_{s,h+1}) \\ &= L_p(s-1)(v_{s1} - v_{1s}) + \sum_{h=1}^{s-2} L_p(h)(v_{s-h,1} - v_{s-h,s}) \end{aligned}$$

$$+\sum_{h=1}^{s-2} L_p(h)(v_{s,h+1} - v_{1,h+1})$$

$$+\sum_{h=s}^{m-1} \left[L_p(h) - L_p(h+1-s)\right](v_{s,h+1} - v_{1,h+1})$$

The terms of the last expression have either the form $(v_{sh} - v_{1h})$ or $(v_{h1} - v_{hs})$. Since $v_{sh} \geq v_{1h}$ is equivalent to $v_{h1} \geq v_{hs}\ \forall h \neq 1, s$, all the terms $(v_{sh} - v_{1h})$ and $(v_{h1} - v_{hs})$ are non-negative. Their coefficients also are non-negative. Furthermore, $L_p(s-1)(v_{s,1} - v_{1,s}) > 0$, hence $M_p(\hat{r}; N) - M_p(r; N) > 0$. ■

Proposition 3. *For every function $p \in P^m$, Γ_p satisfies the weak Pareto principle on the subset of polls resulting from transitive individual votes.*

Proof. Consider a poll N. Assume that alternative 1 is selected by $\Gamma_p(N)$ and that all voters prefer s to 1. Then, we have $v_{s1} = n$ and since all votes are transitive, we also have $v_{sh} \geq v_{1h}\ \forall h \neq 1$. Indeed, those preferring 1 to h must also prefer s to h since they prefer s to 1. Under the terms of Lemma 2, we also have $M_p(\hat{r}; N) > M_p(r; N)$, which excludes alternative 1 from $\Gamma_p(N)$, a contradiction. ■

Local independence of irrelevant alternatives. Arrow's (1951) independence of irrelevant alternatives says that, when aggregating individual rankings, only the way voters order alternatives within a pair should matter to arrive at a final ranking on the same pair. The Condorcet rule, which leads to the majority relation $FR^M(N)$, respects this condition, but it can produce cycles. On the other hand, r_p, and in particular the Kemeny rule, violates Arrow's independence condition. This violation is expected since r_p gives consistent rankings on the set of alternatives. Indeed, in a famous theorem, Arrow (1951) shows that there is no aggregation rule of transitive individual preferences that results in a transitive collective preference, that satisfies independence of irrelevant alternatives, and that is paretian, unless it is dictatorial or unless we restrict the set of admissible preferences. In light of the same theorem, cycles in $FR^M(N)$ are also to be expected.

This led Young and Levenglick (1978) to define a weaker condition that they call *local independence of irrelevant alternatives*. This condition requires that the ranking of any subset of successive alternatives (forming an interval) in the ranking produced by a rule should remain unchanged if we ignore alternatives that are outside this subset (interval). For example, with 5 alternatives, if a rule selects the ranking *abcde*, it should select the ranking *bcd* when applied to the subset formed by these three alternatives. Young and Levenglick show that the Kemeny rule is the only rule to satisfy this weaker independence condition along with some other desirable conditions.

The example in the proof of Proposition 1 shows that r_p does not satisfy this local independence condition for every choice of the probability function p. In that example, if we set $p(k) = 0.6^{1/k}$ and if we restrict the vote to $\{a, b, c\}$, the most likely ranking is *abc* while it is *bacd* on $\{a, b, c, d\}$. Thus, the interval *bac* of the most likely ranking is changed for *abc* when alternative d is ignored. This is a violation of the local independence condition.

Actually, d plays an important role in the determination of the most likely ranking on $\{a, b, c, d\}$. As it is defeated under the majority rule by each of the other alternatives, d cannot be the top ranked alternative in the most likely ranking regardless of the value of $\bar{p}$. However, it is the 7 votes for b over d compared to the 5 votes for a over d that makes *bacd* more likely than *abcd* when $\bar{p}$ is sufficiently close to $\frac{1}{2}$, i.e. when the relative weights of the majorities have a large dispersion.

5 The most likely rankings with three alternatives

Section 4 showed the importance of the choice of the function p in the selection of the most likely ranking. In particular, the Borda ranking can be selected as the most likely ranking over the Condorcet order and the Borda winner can be selected by Γ_p instead of the Condorcet winner. In this section, we return to this question in a more systematic way for the case where $X = \{a, b, c\}$, where the Kemeny ranking is unique and where p belongs to the class of logistic probability functions defined by:

$$p(k; \alpha, \beta) = \frac{e^{\alpha+\beta(k-1)}}{1 + e^{\alpha+\beta(k-1)}}, \text{ with } \alpha > 0 \text{ and } \beta \geq 0 \tag{6}$$

In this definition, k is again the distance between two alternatives and the parameter β controls the rate at which p increases with respect to k. The larger β, the higher this rate. With $\beta = 0$, the probability is constant. And the larger α, the larger the probability. Figure 1 shows how the probability varies with α when $\beta = 0$.

With no loss of generality, we assume that *abc* is the Kemeny ranking. We denote by $\mathcal{N}^{3Ku}$ the subset of polls on $\{a, b, c\}$ for which *abc* is the unique Kemeny ranking. The next lemma will allow us to partition $\mathcal{N}^{3Ku}$ into three subsets. As we shall see, the most likely ranking depends not only on the values of the parameters α and β but also on which of these three subsets the poll belongs to.

Lemma 4. *For every poll $N \in \mathcal{N}^{3Ku}$:*

1. $v_{ab} > \frac{n}{2} > v_{ba}$;
2. $v_{bc} > \frac{n}{2} > v_{cb}$;
3. $\min\{v_{ab}, v_{bc}\} > v_{ca}$.
4. *Moreover, abc is the Condorcet ranking if and only if $v_{ac} > \frac{n}{2} > v_{ca}$.*

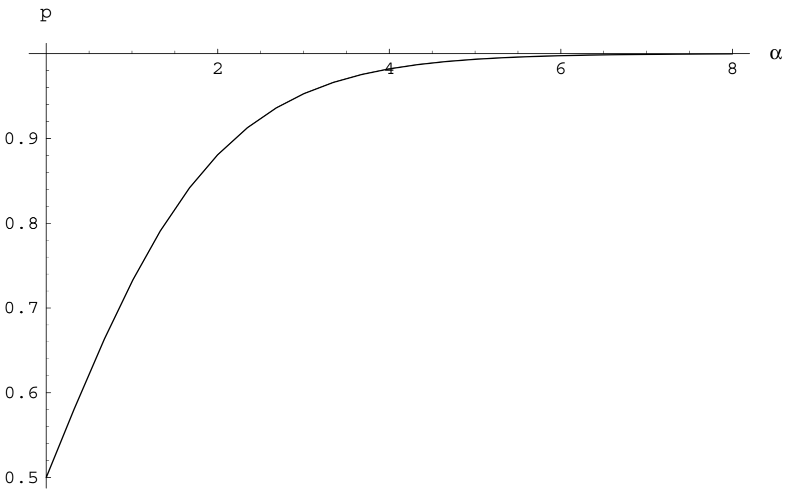

Fig. 1. $p\ (\cdot;\ \alpha,\ 0)$

Proof.

1. Since $K(abc;N) = v_{ab} + v_{bc} + v_{ac} > v_{ba} + v_{ac} + v_{bc} = K(bac;N)$, we immediately have $v_{ab} > v_{ba}$. The fraction $\frac{n}{2}$ is included between these two numbers since $v_{ab} + v_{ba} = n$.
2. Since $K(abc;N) > K(acb;N)$, we immediately have $v_{bc} > v_{cb}$.
3. Since $K(abc;N) = v_{ab} + v_{bc} + v_{ac} > v_{bc} + v_{ba} + v_{ca} = K(bca;N)$, we have $v_{ab} + n - v_{ca} > v_{ca} + n - v_{ab}$, and thus $v_{ab} > v_{ca}$. Similarly, since $K(abc;N) = v_{ab} + v_{bc} + v_{ac} > v_{ca} + v_{ab} + v_{cb} = K(cab;N)$, we have $v_{bc} + n - v_{ca} > v_{ca} + n - v_{bc}$, and thus $v_{bc} > v_{ca}$.
4. By definition and by 1 and 2. ■

In light of Lemma 4, only three entries of the matrix $N \in \mathcal{N}^{3Ku}$ can qualify as being possibly the largest: v_{ab}, v_{bc}, v_{ac}. To avoid future complications, we assume that these three cases are mutually exclusive. Consequently, we can partition $\mathcal{N}^{3Ku}$ into three non-empty subsets $\mathcal{N}_1$, $\mathcal{N}_2$, and $\mathcal{N}_3$:

- $\mathcal{N}_1$ gathers all the polls for which $\max\{v_{st} : s, t \in X\} = \{v_{ac}\}$;
- $\mathcal{N}_2$ gathers those for which $\max\{v_{st} : s, t \in X\} = \{v_{ab}\}$;
- $\mathcal{N}_3$ gathers those for which $\max\{v_{st} : s, t \in X\} = \{v_{bc}\}$.

The results of this section are summarized in Table 1. In $\mathcal{N}_1$, *abc* is the Condorcet ranking and it is also the most likely ranking for any values of the parameters α and β. In $\mathcal{N}_2$ and in $\mathcal{N}_3$, *abc* is still the most likely for values of

Table 1. Summary of the results

Type of poll	$\mathcal{N}_1$	$\mathcal{N}_2$	$\mathcal{N}_3$
$\max\{v_{st} : s, t \in X\}$	v_{ac}	v_{ab}	v_{bc}
r_p	$abc\ \forall\alpha, \beta$	abc if $\alpha > \breve{\alpha}(\beta; N)$ acb if $\alpha < \breve{\alpha}(\beta; N)$	abc if $\alpha > \hat{\alpha}(\beta; N)$ bac if $\alpha < \hat{\alpha}(\beta; N)$
Γ_p	$a\ \forall\alpha, \beta$	$a\ \forall\alpha, \beta$	a if $\alpha > \tilde{\alpha}(\beta; N)$ b if $\alpha < \tilde{\alpha}(\beta; N)$

α sufficiently large, given the value of β. For sufficiently small values of α, *acb* is the most likely in $\mathcal{N}_2$ and *bac* is the one in $\mathcal{N}_3$. Actually, with α sufficiently small, the top alternative in the most likely ranking is the alternative that has the largest majority in all binary comparisons while the alternative that is defeated with the largest majority by the top alternative is at the bottom of the most likely ranking.

In $\mathcal{N}_2$ as in $\mathcal{N}_1$, since a is the top alternative of the most likely ranking, be it *abc* or *acb*, it is the alternative selected by Γ_p. However, in $\mathcal{N}_3$, b can be selected if α is sufficiently small.

When is the Borda ranking the most likely and when does Γ_p select the Borda winner? The next proposition and remark provide some answers to this question. Another proposition establishes that in $\mathcal{N}^{3Ku}$, a is the unique Kramer-Simpson winner.

Proposition 5. *For every poll* $N \in \mathcal{N}_1 \cup \mathcal{N}_2$, *a is the unique Borda winner. For every poll* $N \in \mathcal{N}_1$, *abc is the unique Borda ranking.*

Proof. In $\mathcal{N}_1$, we have $v_{ac} > v_{bc}$. According to Lemma 4, we also have $v_{ab} > v_{ba}$. Therefore, $v_{ab} + v_{ac} > v_{ba} + v_{bc}$. In $\mathcal{N}_2$, we have $v_{ab} > v_{ca}$ and therefore $v_{ac} > v_{ba}$. We also have $v_{ab} > v_{bc}$. Therefore, $v_{ab} + v_{ac} > v_{ba} + v_{bc}$. In $\mathcal{N}_1$, we have $v_{ac} > v_{ab}$ and therefore $v_{ba} > v_{ca}$. From Lemma 4, we also have $v_{bc} > v_{cb}$. Thus, $v_{ba} + v_{bc} > v_{ca} + v_{cb}$, which means that *abc* is the Borda ranking. ■

Remark 4. In $\mathcal{N}_2$, the Borda ranking might be *acb* as well as *abc*. In $\mathcal{N}_3$, we can verify that $v_{ba} + v_{bc} > v_{ca} + v_{cb}$. However, $v_{ab} + v_{ac}$ can be larger, equal or less than $v_{ba} + v_{bc}$. Therefore, the Borda ranking can be *bac* or *abc*. In summary, in $\mathcal{N}_1$, the most likely ranking is always the Borda ranking. It is not always the case in $\mathcal{N}_2$ and $\mathcal{N}_3$ since the most likely ranking depends on the value of the parameters α and β. As for Γ_p, it always selects the Borda winner in $\mathcal{N}_1$ and $\mathcal{N}_2$ *but* not in $\mathcal{N}_3$.

Proposition 6. *For every poll* $N \in \mathcal{N}^{3Ku}$, *a is the unique Kramer-Simpson winner.*

Proof. It is well known and easy to verify that the Condorcet winner, when it exists, is also the unique Kramer-Simpson winner. Thus, let us focus on the polls for which there is no Condorcet winner. These are the polls in $\mathcal{N}_2$ and $\mathcal{N}_3$ such that $v_{ca} > v_{ac}$.

In $\mathcal{N}_2$, we have $v_{ab} > v_{bc} > v_{ca} > v_{ac} > v_{cb} > v_{ba}$ according to Lemma 4. Therefore, $v_{ab} > v_{ac}$, $v_{bc} > v_{ba}$, and $v_{ca} > v_{cb}$. Since $v_{ac} > v_{cb} > v_{ba}$, a is the alternative for which the smallest majority is the largest. Thus, it is the unique Kramer-Simpson winner.

In $\mathcal{N}_3$, we have $v_{bc} > v_{ab} > v_{ca} > v_{ac} > v_{ba} > v_{cb}$. Therefore, $v_{ab} > v_{ac}$, $v_{bc} > v_{ba}$, and $v_{ca} > v_{cb}$. Since $v_{ac} > v_{ba} > v_{cb}$, a is again the unique Kramer-Simpson winner. ■

Remark 5. For the probability function p defined in (6), we have:

$$\frac{p(k;\alpha,\beta)}{1-p(k;\alpha,\beta)} = e^{\alpha+\beta(k-1)}$$

Thus, the function L_p defined in (4) takes the form:

$$L(k;\alpha,\beta) = L_p(k) = \alpha + \beta(k-1)$$

In the case of three alternatives and for order *stu*, we therefore have:

$$M(stu;N,\alpha,\beta) = M_p(stu;N) = \alpha(v_{st}+v_{tu}) + (\alpha+\beta)v_{su} = \alpha(v_{st}+v_{tu}+v_{su}) + \beta v_{su}$$

Note that $M(stu;N,\alpha,0) = \alpha K(stu;N)$.

Lemma 7. *For every $\alpha > 0$, every $\beta \geq 0$ and every poll $N \in \mathcal{N}^{3Ku}$:*

1. $M(abc;N,\alpha,\beta) > M(bca;N,\alpha,\beta)$;
2. $M(abc;N,\alpha,\beta) > M(cab;N,\alpha,\beta)$;
3. $M(acb;N,\alpha,\beta) > M(cba;N,\alpha,\beta)$.

Proof.

1. From Lemma 4, we have $v_{ab} > v_{ca}$. Therefore, $v_{ac} > v_{ba}$ and $v_{ab} + v_{bc} > v_{bc} + v_{ca}$. This implies $M(abc;N,\alpha,\beta) > M(bca;N,\alpha,\beta)$.
2. Similarly, we have $v_{bc} > v_{ca}$, which implies $v_{ac} > v_{cb}$ and $v_{ab} + v_{bc} > v_{ca} + v_{ab}$. Thus, $M(abc;N,\alpha,\beta) > M(cab;N,\alpha,\beta)$.
3. Finally, we have $v_{ab} > v_{ca}$, which implies $v_{ac} > v_{ba}$ and $v_{ac} + v_{cb} > v_{cb} + v_{ba}$. Thus, $M(acb;N,\alpha,\beta) > M(cba;N,\alpha,\beta)$. ■

Corollary 8. *For every $\alpha > 0$, every $\beta \geq 0$, and every poll $N \in \mathcal{N}^{3Ku}$, the most likely ranking is either acb, abc, or bac.*

The most likely ranking in $\mathcal{N}_1$. For the polls of $\mathcal{N}_1$, things are simple: *abc* is both the Condorcet ranking and the Borda ranking. It is also the most likely ranking for all values of the parameters α and β.

Proposition 9. *For every $\alpha > 0$, every $\beta \geq 0$, and every poll $N \in \mathcal{N}_1$, the most likely ranking is abc.*

Proof. We know that *abc* is the most likely ranking for $\beta = 0$, i.e. $M(abc; N, \alpha, 0) > M(stu; N, \alpha, 0)\ \forall stu \neq abc$. Since $v_{ac} = \max\{v_{st} : s, t \in X\}$, $M(abc; N, \alpha, \beta)$ increases more rapidly than $M(stu; N, \alpha, \beta)$ when β increases, preserving the inequality $M(abc; N, \alpha, \beta) > M(stu; N, \alpha, \beta)$. ■

The most likely rankings in $\mathcal{N}_2$. For the polls of $\mathcal{N}_2$, things are a little more complex than in $\mathcal{N}_1$. For β sufficiently small, *abc* is still the most likely ranking but, for sufficiently large values of β, *acb* becomes the most likely in spite of the fact that $\max\{v_{st} : s, t \in X\} = \{v_{ab}\}$. However, $\Gamma_p(N) = \{a\}$ regardless of the value of β.

Proposition 10. *For every $\alpha > 0$, every $\beta \geq 0$, and every poll $N \in \mathcal{N}_2$, the most likely ranking is either abc or acb.*

Proof. By definition of $\mathcal{N}_2$, we have $v_{ab} > v_{bc}$. Therefore, $v_{cb} > v_{ba}$ and $v_{ac} + v_{cb} > v_{ba} + v_{ac}$. It follows that $M(acb; N, \alpha, \beta) > M(bac; N, \alpha, \beta)$. This eliminates *bac* as potentially the most likely ranking in $\mathcal{N}_2$. The rankings *bca*, *cab*, and *cba* were eliminated in Corollary 8. ■

Proposition 11. *Given a poll $N \in \mathcal{N}_2$, consider the linear function $\breve{\alpha}(\cdot; N) : \mathbb{R}_+ \to \mathbb{R}_+$ defined by:*

$$\breve{\alpha}(\beta; N) = \beta \frac{v_{ab} - v_{ac}}{v_{bc} - v_{cb}}$$

For every $\beta \geq 0$, we have:

$$\Pr(abc \mid N, \alpha, \beta) \gtreqless \Pr(acb \mid N, \alpha, \beta) \Leftrightarrow 0 < \alpha \gtreqless \breve{\alpha}(\beta; N)$$

Proof. $\breve{\alpha}(\beta; N)$ is the solution of $M(abc; N, \alpha, \beta) = M(acb; N, \alpha, \beta)$ with respect to α. The direction of the inequalities comes from $v_{ab} + v_{bc} + v_{ac} > v_{ac} + v_{cb} + v_{ab}$. ■

The most likely rankings in $\mathcal{N}_3$. The analysis of $\mathcal{N}_3$ is similar to that of $\mathcal{N}_2$. For β sufficiently small, *abc* is still the most likely ranking, but for sufficiently large values of β, it is *bac* that becomes the most likely. In contrast to what happens in $\mathcal{N}_2$, $\Gamma_p(N)$ changes with β.

Proposition 12. *For every $\alpha > 0$, every $\beta \geq 0$, and every poll $N \in \mathcal{N}_3$, the most likely ranking is either abc or bac.*

Proof. By definition of $\mathcal{N}_3$, we have $v_{bc} > v_{ab}$. Therefore, $v_{ba} > v_{cb}$ and $v_{ba} + v_{ac} > v_{ac} + v_{cb}$. It follows that $M(bac; N, \alpha, \beta) > M(acb; N, \alpha, \beta)$. Thus, *acb* is eliminated from the list of the potentially most likely rankings in $\mathcal{N}_3$. The rankings *bca*, *cab*, and *cba* were eliminated by Corollary 8. ■

Proposition 13. *Given a poll $N \in \mathcal{N}_3$, consider the function $\hat{\alpha}(\cdot; N) : \mathbb{R}_+ \to \mathbb{R}_+$ defined by:*

$$\hat{\alpha}(\beta; N) = \beta \frac{v_{bc} - v_{ac}}{v_{ab} - v_{ba}}$$

For every $\beta \geq 0$, *we have:*

$$\Pr(abc \mid N, \alpha, \beta) \gtreqless \Pr(bac \mid N, \alpha, \beta) \Leftrightarrow 0 < \alpha \gtreqless \alpha(\beta; N)$$

Proof. $\hat{\alpha}(\beta; N)$ is the solution of $M(abc; N, \alpha, \beta) = M(bac; N, \alpha, \beta)$ with respect to α. The direction of the inequalities comes from $v_{ab} + v_{bc} + v_{ac} > v_{ba} + v_{ac} + v_{bc}$. ■

Remark 6. It is easy to verify that for every $\beta > 0$, we have:

$\hat{\alpha}(\beta; N) < \beta$ if a is the Borda winner

$\hat{\alpha}(\beta; N) = \beta$ if a and b are Borda winners

$\hat{\alpha}(\beta; N) > \beta$ if b is the Borda winner

6 Alternatives most likely to be the best

Condorcet made a distinction between the most likely ranking and the alternative most likely to be the best. He was apparently aware that the alternative with the largest probability of being the best is not necessarily the top alternative in the most likely ranking. Young (1988) shows that indeed with a constant probability close to $\frac{1}{2}$, the alternative most likely to be the best is the Borda winner, which may be different from the top alternative in the most likely ranking. With a probability sufficiently large, the alternative most likely to be the best is the Kramer-Simpson winner. We extend Young's analysis to the more flexible formulation of the probabilities adopted in this paper, again for the case $m = 3$.

According to Young's approach, a is the best alternative if it is at the same time better than b and better than c. Denoting the relation "a is better than b" by $a \succ b$, the probability that a is the best alternative, conditional on the poll N, is, under the independence assumption, given by:

$$\Pr(a \mid N) = \Pr(a \succ b \mid N) \Pr(a \succ c \mid N)$$

The computation of the probabilities $\Pr(a \succ b \mid N)$ and $\Pr(a \succ c \mid N)$ can be done unambiguously with constant probabilities. However, this is not so with variable probabilities, since the distance between the alternatives in a given ranking is lost in the above formula. Thus, we need to develop an approach that captures the notion of distance between the alternatives.

Note that:

$$[a \succ b \wedge a \succ c] = [a \succ b \wedge a \succ c] \wedge [b \succ c \vee c \succ b] = [a \succ b \succ c] \vee [a \succ c \succ b]$$

In other words, a is at the same time better than b and better than c if and only if abc or acb is the most likely ranking. Thus, the probability that a is the best alternative, conditional on the poll N, is given by:

$$\Pr(a \mid N) = \Pr(abc \mid N) + \Pr(acb \mid N)$$

We obtain similar expressions for alternatives b and c.[3] With this approach, the computation of the probabilities is clear and consistent, whether the probability is constant or increasing.

We establish a first general result.

Lemma 14. *For every $\alpha > 0$, every $\beta \geq 0$ and every poll $N \in \mathcal{N}^{3Ku}$, the alternative most likely to be the best is either a or b.*

Proof. By Lemma 7, $\Pr(abc \mid N, \alpha, \beta) > \Pr(cab \mid N, \alpha, \beta)$ and $\Pr(acb \mid N, \alpha, \beta) > \Pr(cba \mid N, \alpha, \beta)$. Therefore, $\Pr(a \mid N, \alpha, \beta) > \Pr(c \mid N, \alpha, \beta)$. ■

Which of a or b is the alternative most likely to be the best depends once again on the values of the parameters α and β and on the type of poll. The last row of Table 1 summarizes the results of this section.

The alternative most likely to be the best in $\mathcal{N}_1$ and $\mathcal{N}_2$. For the polls of $\mathcal{N}_1$ and $\mathcal{N}_2$, the alternative most likely to be the best is always a. Let us recall that a is at the same time the Kemeny, the Kramer-Simpson and the Borda winner. Thus, the alternative most likely to be the best is given by $\Gamma_p(N)$.

Proposition 15. *For every $\alpha > 0$, every $\beta \geq 0$, and every poll $N \in \mathcal{N}_1 \cup \mathcal{N}_2$, a is the alternative most likely to be the best.*

Proof. In $\mathcal{N}_1$, we have $\Pr(abc \mid N, \alpha, \beta) > \Pr(bac \mid N, \alpha, \beta)$ from Proposition 9. By definition of N_1, $v_{ac} > v_{bc}$ and according to Lemma 4, $v_{ab} > v_{ba}$. Therefore, $v_{cb} > v_{ca}$ and $v_{ac} + v_{cb} > v_{bc} + v_{ca}$. Then, we have $M(acb; N, \alpha, \beta) > M(bca; N, \alpha, \beta)$, i.e. $\Pr(acb \mid N, \alpha, \beta) > \Pr(bca \mid N, \alpha, \beta)$. Combining these two inequalities, we obtain:

$$\begin{aligned}\Pr(a \mid N, \alpha, \beta) &= \Pr(abc \mid N, \alpha, \beta) + \Pr(acb \mid N, \alpha, \beta) \\ &> \Pr(bac \mid N, \alpha, \beta) + \Pr(bca \mid N, \alpha, \beta) = \Pr(b \mid N, \alpha, \beta)\end{aligned}$$

In $\mathcal{N}_2$, we have $v_{ab} > v_{bc}$ and from Lemma 4 , $v_{ab} > v_{ba}$. Therefore, $v_{cb} > v_{ba}$ and $v_{ac} + v_{cb} > v_{ba} + v_{ac}$. Then, we have $M(acb; N, \alpha, \beta) > M(bac; N, \alpha, \beta)$, i.e.

[3] The three events
$[a \succ b \succ c] \vee [a \succ c \succ b]$, $[b \succ a \succ c] \vee [b \succ c \succ a]$, $[c \succ b \succ a] \vee [c \succ a \succ b]$
are not exhaustive. We also have the event $[a \succ b \succ c \succ a] \vee [a \succ c \succ b \succ a]$, i.e. the possibility that there is no best alternative or that the poll is the result of a cyclical relation rather than an order. This is why the probabilities of the three alternatives do not add up to 1 in Young's formulation. With uniform probabilities, we can define $\Pr(N \mid a \succ b \succ c \succ a)$ unambiguously. This is not so with increasing probabilities. For example, is a one position behind c or two positions ahead of c in the cycle $a \succ b \succ c \succ a$? Thus, we must ignore these events. This is why we wrote $\Pr(r \mid N) = \frac{\Pr(N|r)}{\sum_{q \in \mathcal{R}} \Pr(N|q)}$ rather than $\frac{\Pr(N|r)}{\sum_{q \in \mathcal{B}} \Pr(N|q)}$ in (3) of Sect 3.

$\Pr(acb \mid N, \alpha, \beta) > \Pr(bac \mid N, \alpha, \beta)$. According to Lemma 7, $\Pr(abc \mid N, \alpha, \beta) > \Pr(bca \mid N, \alpha, \beta)$. Consequently, $\Pr(a \mid N, \alpha, \beta) > \Pr(b \mid N, \alpha, \beta)$. ■

Alternatives most likely to be the best in $\mathcal{N}_3$. For the polls of $\mathcal{N}_3$, the Borda winner can be a or b. We show that there is a critical value of α (possibly 0), which depends on β, above which a is the alternative most likely to be the best and below which it is b. In particular, if this critical value is 0, this means that a is always the alternative most likely to be the best. This can happen only when a is the Borda winner. In preparation for the next proposition, we prove the following lemma.

Lemma 16. *For every $\beta > 0$ and every poll $N \in \mathcal{N}_3$,*

$$\begin{array}{ll} \Pr(a \mid N, \hat{\alpha}(\beta;N), \beta) > \Pr(b \mid N, \hat{\alpha}(\beta;N), \beta) & \text{if } a \text{ is the Borda winner} \\ \Pr(a \mid N, \hat{\alpha}(\beta;N), \beta) = \Pr(b \mid N, \hat{\alpha}(\beta;N), \beta) & \text{if} a \text{ and } b \text{ are Borda winners} \\ \Pr(a \mid N, \hat{\alpha}(\beta;N), \beta) < \Pr(b \mid N, \hat{\alpha}(\beta;N), \beta) & \text{if} b \text{ is the Borda winner} \end{array}$$

Proof. Define $A = v_{ac} + v_{cb} + v_{ab}$ and $B = v_{bc} + v_{ca} + v_{ba}$. Note that in $\mathcal{N}_3$, if b is the Borda winner or if a and b are Borda winners, then $A < B$. Indeed, $v_{ab} + v_{ac} \leq v_{ba} + v_{bc}$ combined with $v_{cb} < v_{ca}$ give $A < B$. If a is the Borda winner, we can have $A < B$ as well as $A \geq B$. Next, note that the following statements are equivalent:

$$\begin{aligned} \Pr(a \mid N, \hat{\alpha}(\beta;N), \beta) &\gtreqless \Pr(b \mid N, \hat{\alpha}(\beta;N), \beta) \\ \Pr(acb \mid N, \hat{\alpha}(\beta;N), \beta) &\gtreqless \Pr(bca \mid N, \hat{\alpha}(\beta;N), \beta) \\ M(acb; N, \hat{\alpha}(\beta;N), \beta) &\gtreqless M(bca; N, \hat{\alpha}(\beta;N), \beta) \\ \hat{\alpha}(\beta;N)(A - B) &\gtreqless -\beta(v_{ab} - v_{ba}) \end{aligned}$$

If a is the Borda winner and if $A \geq B$, we immediately have:

$$\hat{\alpha}(\beta;N)(A - B) + \beta(v_{ab} - v_{ba}) > 0$$

If a is the Borda winner and if $A < B$, we have

$$\begin{aligned} &\hat{\alpha}(\beta;N)(A - B) + \beta(v_{ab} - v_{ba}) > \beta(A - B) + \beta(v_{ab} - v_{ba}) \\ &= 2v_{ac} + 2v_{ab} - 2v_{bc} - 2v_{ba} > 0 \end{aligned}$$

since $\hat{\alpha}(\beta;N) < \beta$ in this case. For the case where a and b are Borda winners, replace the inequalities by equalities in the last expression. Reverse the inequalities if b is the Borda winner. ■

Proposition 17. *For every $\beta \geq 0$ and every poll $N \in \mathcal{N}_3$, $\exists \tilde{\alpha}(\beta;N) \geq 0$:*

$$\Pr(a; N, \alpha, \beta) \gtreqless \Pr(b; N, \alpha, \beta) \Leftrightarrow 0 < \alpha \gtreqless \tilde{\alpha}(\beta;N)$$

Proof. Consider the equation:

$$\Pr(abc \mid N, \alpha, \beta) + \Pr(acb \mid N, \alpha, \beta) = \Pr(bac \mid N, \alpha, \beta) + \Pr(bca \mid N, \alpha, \beta) \tag{7}$$

The left-hand side is $\Pr(a \mid N, \alpha, \beta)$ and the right-hand side is $\Pr(b \mid N, \alpha, \beta)$. We look for the solution to this equation. Let us define the two terms C and D by:

$$C = \Pr(abc \mid N, \alpha, \beta) - \Pr(bac \mid N, \alpha, \beta)$$
$$D = \Pr(acb \mid N, \alpha, \beta) - \Pr(bca \mid N, \alpha, \beta)$$

The above equation can also be written as $C + D = 0$.

From Lemma 4 and Proposition 12, we know that:

$$v_{ab} + v_{bc} + v_{ac} > v_{ba} + v_{ac} + v_{bc} > \max\{v_{ac} + v_{cb} + v_{ab},\quad v_{bc} + v_{ca} + v_{ba}\}$$

Thus, as α increases, $M(abc; N, \alpha, \beta)$ increases more rapidly than $M(bac; N, \alpha, \beta)$, and the latter more rapidly than both $M(acb; N, \alpha, \beta)$ and $M(bca; N, \alpha, \beta)$. It follows that $\Pr(abc \mid N, \alpha, \beta)$ increases monotonically with α. Thus, $\Pr(acb \mid N, \alpha, \beta)$ and $\Pr(bca \mid N, \alpha, \beta)$ can be made as small as necessary with α sufficiently large.

Suppose now that b is the Borda winner. By definition of $\hat{\alpha}(\beta; N)$ and by Lemma 16, we have $C = 0$ and $D < 0$. Increasing α from $\hat{\alpha}(\beta; N)$ increases $\Pr(abc \mid N, \alpha, \beta)$. The value of $\Pr(bac \mid N, \alpha, \beta)$ may also increase for a while but less than $\Pr(abc \mid N, \alpha, \beta)$. Thus, C increases monotonically with α. As for D, it may decrease initially (increase in absolute terms) but, for sufficiently large values of α, it will get close to 0. Since this behavior is continuous, there exists a unique number $\tilde{\alpha}(\beta; N) > \hat{\alpha}(\beta; N)$ such that $C + D = 0$, i.e. such that $\Pr(a; N, \alpha, \beta) = \Pr(b; N, \alpha, \beta)$. Moreover, $\alpha \lesseqgtr \tilde{\alpha}(\beta; N) \Rightarrow C + D \lesseqgtr 0 \Rightarrow \Pr(a; N, \alpha, \beta) \lesseqgtr \Pr(b; N, \alpha, \beta)$.

Suppose next that a is the Borda winner. We now have $C = 0$ and $D > 0$. By the same argument as above, starting from $\hat{\alpha}(\beta; N)$, C decreases monotonically with α. As for D, it may increase or decrease in the beginning but, for a sufficiently low value of α, it will be positive since v_{ab}, the coefficient of β in $\Pr(acb \mid N, \alpha, \beta)$, is larger than v_{ba}, the coefficient of β in $\Pr(bca \mid N, \alpha, \beta)$. Two cases can arise:

- A positive value of α, such that $C + D = 0$, is reached. This value is $\tilde{\alpha}(\beta; N)$. As when b is the Borda winner, $\alpha \lesseqgtr \tilde{\alpha}(\beta; N) \Rightarrow C + D \lesseqgtr 0 \Rightarrow \Pr(a; N, \alpha, \beta) \lesseqgtr \Pr(b; N, \alpha, \beta)$.
- $C + D > 0\ \ \forall \alpha \leq \hat{\alpha}(\beta; N)$. In this case, we set $\tilde{\alpha}(\beta; N) = 0$ and we have $\Pr(a; N, \alpha, \beta) > \Pr(b; N, \alpha, \beta)\ \ \forall \alpha \geq 0$. ■

Combining Lemma 16 and Proposition 17, we get the following corollary.

Corollary 18. *For every* $\beta > 0$ *and every poll* $N \in \mathcal{N}_3$,

$\hat{\alpha}(\beta; N) > \tilde{\alpha}(\beta; N)$ if a is the Borda winner

$\hat{\alpha}(\beta; N) = \tilde{\alpha}(\beta; N)$ if a and b are Borda winners

$\hat{\alpha}(\beta; N) < \tilde{\alpha}(\beta; N)$ if b is the Borda winner

Figures 2 and 3 illustrate two typical behaviors of $\tilde{\alpha}$ and $\hat{\alpha}$. The first one is obtained from the following poll borrowed from Condorcet (1785), and in which b is the Borda winner:

Alternatives	*a*	b	c
a	–	31	31
b	29	–	37
c	29	23	–

Young (1988) uses this poll to show that the Kemeny rule does not necessarily select the alternative most likely to be the best. This is the case here for $\alpha < \tilde{\alpha}(0) = 0.225$, i.e. for a probability of ordering correctly two alternatives less than 0.556. More interestingly, this figure shows that the performance of Γ_p in selecting the best alternative improves as β gets larger. Indeed, this performance is better the smaller the interval $(\hat{\alpha}(\beta; N),\ \tilde{\alpha}(\beta; N))$. It is only within this interval that Γ_p does not select the alternative most likely to be the best.

Figure 3 is drawn from the following poll, in which a is the Borda winner:

Alternatives	*a*	b	c
a	–	13	12
b	8	–	16
c	9	5	–

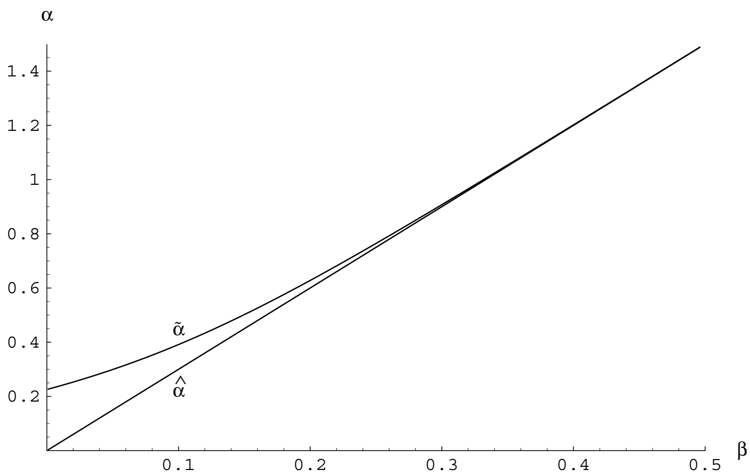

Fig. 2. Typical behaviour of $\tilde{\alpha}$ and $\hat{\alpha}$ when b is the Borda winner

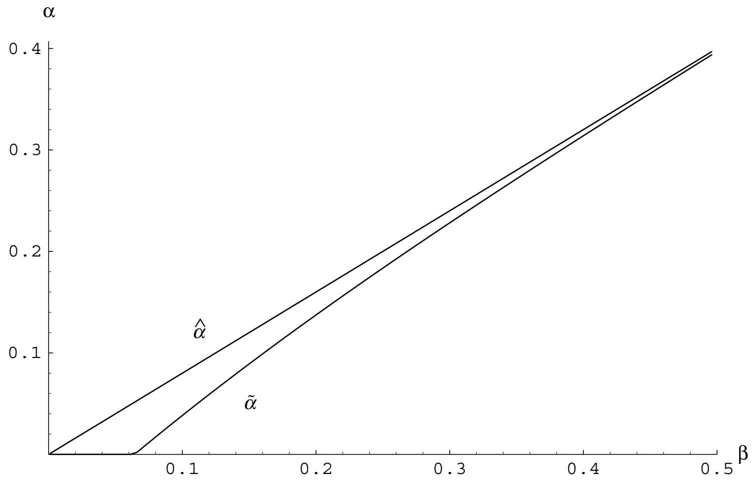

Fig. 3. Typical behaviour of $\tilde{\alpha}$ and $\hat{\alpha}$ when a is the Borda winner

In this case, $\tilde{\alpha}(\beta; N) = 0$ on some interval $[0, \overline{\beta}]$. This follows from Young's result, which, transposed to our context, says that for $\beta = 0$ and α sufficiently small, the alternative most likely to be the best is the Borda winner. As a result, the gap between $\hat{\alpha}(\beta; N)$ and $\overline{\alpha}(\beta; N)$ increases on $[0, \overline{\beta}]$. It decreases monotonically thereafter.

The behavior of $\tilde{\alpha}$ with respect to $\hat{\alpha}$ depicted in Figs. 2 and 3 seems to be quite general. We found numerically that it is typical of all polls in $\mathcal{N}_3$ with $3 \leq n \leq 12$. It is thus also typical of all replicas of these polls. The reason is that, as β gets larger, $\Pr(abc \mid N, \hat{\alpha}(\beta; N), \beta)$ and $\Pr(bac \mid N, \hat{\alpha}(\beta; N), \beta)$ both get larger and, because of this, the term D in the proof of Proposition 17 gets smaller, thus requiring a smaller departure from $\hat{\alpha}(\beta; N)$. However, given the form of the probability function used in this section, a rigorous proof of this fact appears to be out of reach.

7 Conclusion

Condorcet (1785) initiated the statistical approach to vote aggregation. Young (1988) clarified Condorcet's contribution and showed that a correct application of the maximum likelihood principle leads to the Kemeny rule. This approach is based on the assumption that judges or voters have the same probability of ranking correctly two alternatives and that this probability is the same for all pairs of alternatives.

In this paper, we relaxed the second part of this assumption, adopting the point of view that the probability of comparing correctly two alternatives is an increasing function of the distance between them in the true ranking. We showed that the aggregation rule that consists in selecting the most probable ranking under this more reasonable assumption, differs in many respects from the Kemeny rule. We also provided a systematic analysis of the case of three alternatives. Due to the complexity of the problem, it was hardly possible to go further with more alternatives.

The approach adopted here, as in the previous literature, relies on the assumption that votes are statistically independent from one pair of alternatives to the other. Thus, if a voter or a judge orders alternatives pair by pair, there is a positive probability that a cycle will emerge from the vote, even if this probability is conditional on an order. This is inherent to the binary approach, which Saari often criticizes in his writings. For example, Saari and Merlin (2000) make the following remark about the Kemeny rule **(KR)**:

> The unexpected, troubling fact is that *KR achieves its consistency by weakening the crucial assumption about the individual rationality of the voters*. Indeed, KR treats certain groups of preferences as though they come from non-existent voters with cyclic preferences.

As we saw, cyclic preferences have indeed a positive probability of showing up in the binary approach. With a probability function that is increasing with the distance between the two alternatives in a pair, the probability of a cycle is reduced but not completely eliminated. Despite this shortcoming, we remained within the binary framework since our purpose was to verify the robustness of the results obtained so far under the assumption of constant probability across pairs of alternatives.

The next step would consist in relaxing the independence assumption while retaining some of the ideas of this paper to model the competence of the judges. Another topic for further research is the strategic behavior, which has been completely evacuated from this paper. Austen-Smith and Banks (1996) have paved the way in analyzing information aggregation and voting behavior in environments in which individuals have private information about which of two alternatives is possibly the best. Rata (2002b) extends their analysis to the case of many alternatives but the aim of this aggregation remains the choice of a best alternative as opposed to ranking all alternatives. This last problem does not seem to have been studied yet.

References

Arrow KJ (1951) Social choice and individual values. Wiley, New York

Austen-Smith D, Banks JS (1996) Information aggregation, rationality, and the Condorcet jury theorem. Am Polit Sci Rev 90: 34–45

Ben-Yashar R, Nitzan S (1997) The optimal decision rule for fixed size committees in dichotomous choice situations. Int Econ Rev 38: 175–187

Ben-Yashar R, Paroush J (2001) Optimal decision rules for fixed-size committees in polychotomous choice situations. Soc Choice Welfare 18: 737–746

Berg S (1993) Condorcet's jury theorem, dependency among jurors. Soc Choice Welfare 10: 87–95

Berg S (1994) Evaluation of some weighted majority decision rules under dependent voting. Math Soc Sci 28: 71–83

Borda JC de (1784) Mémoire sur les élections au scrutin. Histoire de L'Académie Royale des Sciences
Condorcet Marquis de (1785) Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. De l'Imprimerie Royale, Paris
Drissi-Bakhkhat M (2002) A statistical approach to the aggregation of votes. PhD Thesis, Université Laval, Québec
Grofman B, Owen G, Feld S (1983) Thirteen theorems in search of the truth. Theory Decision 15: 261–278
Kemeny J (1959) Mathematics without numbers. Daedalus 88: 571–591
Ladha KK (1992) The Condorcet jury theorem, free speech and correlated votes. Am J Pol Sci 36: 617–634
Ladha KK (1993) Condorcet's jury theorem in light of De Finetti's theorem. Majority voting with correlated votes. Soc Choice Welfare 10: 69–86
Ladha KK (1995) Information pooling through majority-rule voting: Condorcet's jury theorem with correlated votes. J Econ Behav Organ 26: 353–372
Nitzan S, Paroush J (1982) Optimal decision rules in uncertain dichotomous choice situations. Int Econ Rev 23: 289–297
Paroush J (1998) Stay away from fair coins: A Condorcet jury theorem. Soc Choice Welfare 15: 15–20
Rata C (2002a) Optimal group decision rules: Plurality and its extensions. Manuscript, Universitat Autònoma de Barcelona
Rata C (2002b) Strategic aspects of information aggregation by plurality rule. Manuscript, Universitat Autònoma de Barcelona
Saari DG, Merlin VR (2000) A geometric examination of Kemeny's rule. Soc Choice Welfare 17: 403–438
Shapley LS, Grofman B (1984) Optimizing group judgemental accuracy in the presence of interdependencies. Publ Choice 43: 329–343
Young HP (1988) Condorcet's theory of voting. Am Polit Sci Rev 82: 1231–1244
Young HP, Levenglick A (1978) A consistent extension of Condorcet's election principle. SIAM J Appl Math 35: 285–300