Journal of Modern Applied Statistical Methods
May 2011, Vol. 10, No. 1, 349-361
Copyright © 2011 JMASM, Inc.
1538 – 9472/11/$95.00

The Likelihood of Choosing the Borda-Winner With Partial Preference Rankings of the Electorate

Ömer Eğecioğlu
University of California,
Santa Barbara, CA

Ayça Ebru Giritligil
Istanbul Bilgi University,
Istanbul, Turkey

Given that n voters report only the first r ($1 \leq r < m$) ranks of their linear preference rankings over m alternatives, the likelihood of implementing Borda outcome is investigated. The information contained in the first r ranks is aggregated through a Borda-like method, namely the r-Borda rule. Monte-Carlo simulations are run to detect changes in the likelihood of r-Borda winner(s) to coincide with the original Borda winner(s) as a function of m, n and r. The voters' preferences are generated through the Impartial Anonymous and Neutral Culture Model, where both the names of the alternatives and voters are immaterial. It is observed that, for a given r, the likelihood of choosing the Borda winner decreases down to zero independent of n as m increases within the computed range of parameter values, $1 \leq m$, $n \leq 30$. For $n = 30$, this likelihood is given as an approximating function of r and m through least square fit method.

Key words: Borda rule, r-Borda rule, impartial anonymous, neutral culture.

Introduction

A voting rule solves the collective decision problem where voters must jointly choose one among a number of possible candidates (alternatives) on the basis of reported ordinal preferences. The choice of a voting rule has been a major ethical question since the political philosophy of the Enlightenment. When only two alternatives are at stake, the ordinary majority voting is unambiguously regarded as the best method. For three and more alternatives, plurality voting at which each voter is asked to report exactly one alternative at her/his ballot and the alternative voted the most wins, has been historically the most popular voting rule. The two celebrated critiques of plurality voting, Borda (1781) and Condorcet (1785) noted that plurality voting may elect a poor candidate, namely, one that would lose in a simple pair-wise majority comparison to every other candidate, or one 'disliked' by the strict majority of voters.

Ömer Eğecioğlu is at the Department of Computer Science. Email: omer@cs.ucsb.edu. Ayça Ebru Giritligil is at BELIS, Murat Sertel Center for Advanced Economic Studies. Email: ayca.ebru@bilgi.edu.tr.

Borda and Condorcet are individually devised different rules to replace plurality voting. Borda introduced a scoring method; the Borda rule that assigns points to each candidate, increasing linearly with a candidate's ranking in a voter's opinion, and elects the alternative with the highest total score. Condorcet provided the voting principle which states that if a candidate defeats every other candidate in simple majority rule, then that candidate should be the winner in the election. These two approaches have generated most of the modern scholarly research in social choice literature.

As discussed by Niemi and Riker (1976), Fishburn (1984), Nurmi (1987) and Amy (2000), no voting rule is perfect in aggregating individual preferences into social decisions in a manner compatible with the fulfillment of a variety of positive and normative criteria. However, some procedures are clearly superior to others in satisfying these criteria. Saari (1987, 1989, 1990, 2001) show that the Borda rule is less susceptible than other positional scoring rules to some unsettling possibilities and paradoxes. Some of the theoretical and probabilistic results concerning the Borda rule are summarized in Brams and Fishburn (2002) and Pattanaik (2002). However, among its shortcomings, its vulnerability to strategic

manipulations and the practical difficulties of implementing it remain as the most criticized properties of the Borda rule. There are many studies theoretically and/or probabilistically considering the former issue. This article focuses on the latter which has not been studied in detail.

The implementation criterion is concerned with the complexity of information that a voting procedure requires voters to reveal concerning their preferences regarding the alternatives. Unlike non-ranked single-stage voting procedures (such as plurality voting, negative plurality voting and approval voting) and non-ranked multistage voting procedures (such as plurality with a run-off and plurality with successive elimination), the Borda rule is a ranked procedure. Asking the voters to provide their complete preference rankings over the set of all available alternatives is a difficult-to-fulfill requirement due to the associated complications both on the side of voters as well as administrators to collect the information.

This study investigates the likelihood of implementing Borda outcome when voters are asked to rank only a specified number of alternatives. The situations where n voters are required to report only the first r ($1 \leq r < m$) ranks of their linear (i.e., full or total) preferences over m alternatives are considered. It is assumed that the partial individual preferences are aggregated through a Borda-like method, namely the r-Borda rule. The r-Borda rule assigns strictly positive points to each alternative appearing in the first r-ranks of a voter's total preference, linearly increasing with its rank, and assigns zero points to those that are not among the top r-ranks in the voter's decision. The alternative(s) that receive(s) the highest score aggregated over the electorate's preferences is (are) chosen as the r-Borda winner(s).

In this study, Monte Carlo simulations are run to ascertain the information content of only the first r ranks of the electorate's preferences from the perspective of implementing the (original) Borda outcome. The way the r-Borda rule aggregates the voters' preferences is different than the aggregation methods implemented by well-known single- and multi-stage non-ranked procedures which permit truncated ballots. In other words, the present study does not aim to detect the likelihood of any of these rules to choose the Borda outcome.

The Borda-like aggregation of partially stated individual preferences is a popular method for sports and contests in real life. The Most Valuable Player of the National Basketball Association in the United States, the Eurovision Song Contest, The People's Remix Music Competition and the Formula 1 Car Race are well-known examples such cases. These contests require voters to rank a specified number of candidates. Each stated candidate is given a score depending on its rank in the preference ordering of a voter, and the candidate that receives the highhest total score over the electorate is elected as the winner. The number of candidates to be ranked and the scores to be assigned to the ranks differs from one contest to another.

Given that the Borda rule can choose more than one alternative as winners, in this study, two types of probabilities are computed for triples of m, n and r as the likelihood of choosing the Borda winner with partial individual preferences. The first type refers to the likelihood of the r-Borda rule choosing exactly the set of Borda winners. The second type of probability considers the likelihood of r-Borda winners to be included in the set of Borda winners. The changes in these values as a function of m, n and r are investigated by considering all possible values of these parameters in an appropriate range.

For Monte Carlo simulations, the voters' preferences are generated via the Impartial Anonymous and Neutral Culture Model (IANC). As introduced by Eğecioğlu and Giritligil (2011), IANC treats voters' preferences through a class of preference profiles, namely root profiles, where the names of both voters and alternatives are immaterial.

Contribution and Relation to Literature

To our knowledge, this study is the first attempt in the literature to analyze the extent of difficulty in implementing Borda outcome when voters are asked to rank only a specified number of alternatives and where the underlying model is as structurally general as is possible. The contribution of this computational work and its relation to literature can be discussed based on

two grounds: aggregation of truncated preferences and sampling voters' preferences.

Aggregation of Truncated Preferences

A positional scoring rule assigns a score vector $s = (s_1, s_2, \ldots, s_m)$ with $s_1 \geq s_2 \geq \ldots \geq s_m$ and $s_1 > s_m$ to a preference ranking over a set of m alternatives, and chooses the alternative(s) with the highest total score aggregated over the rankings of all voters. The Borda rule is defined by the scoring vector $s(B_m) = (m, m-1, \ldots, 1)$. That is, $s_i = m + 1 - i$ for all i, and the difference in scores $s_i - s_j$ is proportional to $j - i$ for all i and j. This article considers the situation where voters are asked to state only the first r ranks of their linear preferences over m alternatives and investigates the importance of the information revealed in the first r ranks of the electorate's preferences from the perspective of implementing the Borda outcome. This partial information is aggregated through r-Borda rule B_r which assigns alternative i the score $s_i(B_r) = r + 1 - i$ if $i \leq r$, and $s_i(B_r) = 0$ otherwise.

It should be noted that the r-Borda rule aggregates the partial preferences unlike constant scoring rules. A constant scoring rule asks each voter to indicate a given (and constant) number of alternatives. Each indicated alternative receives one point whereas all others get zero, and the alternative with the most votes is elected. Hence, for $m \geq 3$, the scoring vector imposed by the r-Borda rule is not the same as the one assigned by a constant scoring procedure unless $r = 1$ implying the scoring vector $(1, 0, \ldots, 0)$ which identifies the most popular constant scoring rule, namely the plurality rule. The probability of constant scoring voting rules to select the Borda outcome has been studied by Gehrlein (1981), Gehrlein and Lepelley (2000) and Vandercruyssen (1999). Gehrlein and Fishburn (1980) and Gehrlein, et al. (1982) provide the propensity of pairs of score vectors for a set A of alternatives and a non-empty proper subset of A to yield the same ranking over the subset for an arbitrary profile of linear orders on A.

The method adopted in this paper to aggregate truncated preferences is also different than the procedures that permit truncated ballots. Among these, approval voting has been widely considered in theoretical and practical grounds. Brams and Fishburn (2002) provide a summary of the theoretical debate between approval voting and the Borda rule. Approval voting requires each voter to indicate the alternatives that she/he approves. Each approved alternative by a voter receives one point and the alternative(s) with the highest point summed over all voters' preferences is (are) chosen as the winner(s). Note that, in approval voting, the number of alternatives to be indicated or ranked by the electorate is not given and thus, is not homogeneous across voters.

Another rule which permits voters to submit truncated preference rankings is majoritarian compromise since it needs at most the first half of the voters' rankings over the entire set of alternatives. Introduced by Sertel (1987), majoritarian compromise selects the candidate(s) that has (have) the support of the majority in the best degree possible. Clearly, both approval voting and majoritarian compromise aggregate the truncated preferences in a different fashion than the r-Borda rule adopted herein.

Consider a social planner who believes that Borda rule is the 'best' voting rule to aggregate individual preferences into a social choice. Due to the complications about requiring voters to state their total preference orderings, the planner can ask the electorate to report the first r ($1 \leq r < m$) ranks of their linear preference rankings instead of asking them to state only their first-best choices in the hope of increasing the probability of choosing the Borda winner. In such a case, it seems natural to aggregate the reported partial rankings via a Borda-like procedure for the sake of preserving some consistency in the aggregation method.

Sampling Voters' Preferences

An immense literature has been devoted to analyze the outcomes of various social choice rules through the use of computer simulations employing probability models to generate voters' preferences. The most commonly used probability models in the literature are Impartial Culture (IC) and Impartial Anonymous Culture (IAC) conditions. Introduced by Guilbaud (1952), IC is a multinomial equiprobable preference profiles model which assumes that each voter selects her/his preference according

to a uniform probability distribution. IAC, which was first introduced by Fishburn and Gehrlein (1978), also relies on an equiprobability assumption, but without taking the identity of the voters into account. Details about these assumptions and their use in the literature are presented in Berg and Lepelly (1994) and Gehrlein (1997).

In this paper, voters' preferences are sampled through IANC which is also an equiprobability assumption, however, neglecting the names of both alternatives and voters. IANC treats each 'root profile' (profile from which all preference profiles can be generated through renaming the alternatives and voters) equally probable. Because the number of root profiles is small relative to the number of profiles that can be generated for m alternatives and n voters, IANC enables the researchers to obtain accurate probabilities even for large parameter values. Based on Eğecioğlu and Giritligil (2011), it is known that the probabilities computed using the IAC and IANC models coincide only in the vanishingly small likelihood of $m!$ and n being relatively prime.

In the social choice theory literature, anonymity and neutrality are among the most important ethical axioms which a voting rule is expected to fulfill. Anonymity requires all voters to be treated equally whereas neutrality calls for equal treatment of alternatives. A large group of voting rules including all scoring rules and pair-wise majority relation rules fulfill these two axioms. The outcomes of anonymous and neutral voting rules are invariant under group symmetries of voter and alternative names. Among the probability models used for sampling voters' preferences, IC assumes no set of symmetries whereas IAC takes into account only the symmetry of voter names. On the other hand, IANC takes into account the symmetries of both voters' and alternatives' names. This paper is the first study in the literature that adopts IANC to sample electorate's preferences.

Through the preference sampling method developed by Falmagne and Regenwetter (1996), Regenwetter and Grofman (1998) analyzed seven three-candidate elections conducted under approval voting and constructed a distribution of preference rankings from subset choice data to compare the results with potential winners of the Borda and Condorcet rules. Based on the method of generalized spectral analysis introduced by Lawson et al. (2006), Brams et al. (2006) compare the results of The Public Choice Society presidential elections in 2006, which was run through approval voting, with the possible outcomes that would have been obtained if plurality, Condorcet, Borda or a single transferable vote had been adopted.

Both Regenwetter and Grofman (1998) and Brams, et al. (2006) start with partial information on voter preferences and assign probabilities to each alternative to be the Borda winner, and based on these probabilities, check whether the possible Borda winner(s) coincide(s) with the actually elected alternative(s). The present study, on the other hand, generates the full orderings of the electorate over the set of alternatives and then considers the first r ranks of the preference profiles. The approaches of the former studies and the present one are clearly different from each other methodologically.

Preliminaries: Preference Profiles and the Borda rule

A preference on a set A means any function p: $A \rightarrow 2^A$ which assigns to every $a \in A$ a subset (lower contour set) $p(A) \subseteq A$ such that, at all $a, b \in A$:

(1) $b \in p(a)$ or $a \in p(b)$: completeness;
(2) $p(b) \subset p(a)$ whenever $b \in p(a)$: transitivity
(3) $b \in p(a)$ and $a \in p(b)$ only if $a = b$: anti-symmetry.

Such a preference clearly corresponds to a linear (or total) order on A.

$\boldsymbol{p}(A)$ denotes the set of all preferences on A, any positive integer n means $[n] = \{1, 2,. ,n\}$, and a preference profile for a society of n voters on a set A means any family $P_{m,n} = (p_i)_{i\in[n]} \in \boldsymbol{p}(A)^{[n]}$ of preferences p_i on A indexed by voters $i \in [n]$. Let $card(p_i(a))$ be the cardinality of the lower counter set of $a \in A$ for the voter $i \in [n]$. Note that the cardinality of the lower counter set of the top- and bottom- ranked alternatives are m and 1, respectively.

The Borda score of $a \in A$ for $i \in [n]$ is defined as,

$$B_i^a = card(p_i(a)),$$

and the set of Borda winners at each $P_{m,n} \in \boldsymbol{p}(A)^{[n]}$ is determined by setting

$$B(P_{m,n}) = \arg\max_{a \in A} \sum_{i \in [n]} B_i^a .$$

Thus, the Borda rule chooses the candidates who maximize the total Borda score aggregated over the set of all n voters.

Let $P_{m,n}^r$ denote the portion of a preference profile $P_{m,n}$ where only the first r ranks of the voters' preferences can be observed. Viewing the profile $P_{m,n}$ as a $m \times n$ matrix, $P_{m,n}^r$ corresponds to the $r \times n$ submatrix of $P_{m,n}$. Note that, if $r = m$ or $r = m - 1$, $B(P_{m,n})$ is detectable. However, if $r < m - 1$, the observable preference p_i^r of voter i corresponds to a partial strict ordering on A, which is transitive and anti-symmetric, however incomplete. Let $A_i^r \subseteq A$ be the set of alternatives appearing at p_i^r.

Let $card(p_i^r(a))$ be the cardinality of the observable lower counter set of $a \in A_i^r$ for $i \in [n]$. Note that $1 \leq card(p_i^r(a)) \leq r$. The Borda score of $a \in A$ for $i \in [n]$ is redefined as: $B_i^a = card(p_i^r(a))$, if $a \in A_i^r$, and 0 otherwise, and the set of r-Borda winners at any $P_{m,n}^r$ is given by:

$$B(P_{m,n}^r) = \arg\max_{a \in A} \sum_{i \in [n]} B_i^a .$$

In other words, if an alternative is among the first r-ranks in voter i's ranking, then its associated r-Borda score is equal to the cardinality of its lower counter set in p_i^r. If it is not among the top r-ranked alternatives, it receives a score of zero. The r-Borda rule chooses the alternative(s) with the highest r-Borda score aggregated over $P_{m,n}^r$ as winner(s). From this point on, $B(P_{m,n}^r)$ and $B(P_{m,n})$ will be denoted by B_r and B, respectively, for $1 \leq r \leq m$.

Root Profiles and IANC

Let $\Omega(m, n)$ denote the set of all preference profiles that can be generated for m alternatives and n voters. As shown in Eğecioğlu and Giritligil (2011), a product permutation group on the names of alternatives and of voters 'acts' on $\Omega(m, n)$, and splits it into a disjoint union of subsets called orbits, that is:

$$\Omega(m, n) = \theta_1 + \theta_2 + \cdots + \theta_\omega$$

where each θ_i is an anonymous and neutral equivalence class (ANEC). All preference profiles within an ANEC can be generated from each other through re-labeling the alternatives and/or the voters. That is, all preference profiles in any ANEC are 'equalivalent' in the sense that any anonymous and neutral voting rule (such as the Borda rule) yields the same outcome (under different names) for all of these profiles.

A root profile is any preference profile that represents an ANEC. That is, all other preference profiles within the same equivalence class can be generated from this root profile via permuting the names of the m alternatives and of the n voters. The collection of all root profiles for m alternatives and n voters is denoted by $R = R(m, n)$, and each element of this set represents an ANEC in $\Omega(m, n)$.

Consider a case with two alternatives, a and b, and three voters labeled v_1, v_2 and v_3 linearly ranking these alternatives. Note that, in this example, there are 2! preference rankings over alternatives (a being strictly preferred to b and b strictly preferred to a). Below are the $(2!)^3 = 8$ possible preference profiles that can be generated:

P_1:

v_1	v_2	v_3
a	a	a
b	b	b

P_2:

v_1	v_2	v_3
a	a	b
b	b	a

P_3:

v_1	v_2	v_3
a	b	a
b	a	b

P_4:

v_1	v_2	v_3
b	a	a
a	b	b

P_5:

v_1	v_2	v_3
b	b	a
a	a	b

P_6:

v_1	v_2	v_3
b	a	b
a	b	a

P_7:

v_1	v_2	v_3
a	b	b
b	a	a

P_8:

v_1	v_2	v_3
b	b	b
a	a	a

If the group of permutations on the names of the voters acts on the above set of preference profiles, it partitions it into four anonymous equivalence classes (AECs):

$$AEC_1 = \{P_1\},\ AEC_2 = \{P_2, P_3, P_4\},$$
$$AEC_3 = \{P_5,\ P_6,\ P_7\},\ AEC_4 = \{P_8\}.$$

Note that there are two possible permutations on the names of the alternatives: one is the identity permutation which leaves the names of the alternatives intact and the other is the permutation which re-labels *a* as *b* and *b* as *a*. If this group of permutations act on the set of AECs, two ANECs are obtained:

$$ANEC_1 = \{AEC_1, AEC_4\},$$
$$ANEC_2 = \{AEC_2,\ AEC_3\}.$$

The root representing $ANEC_1$ shows a preference structure at which all voters have the same preference ranking and the root representing $ANEC_2$ exhibits a structure where one of the preference rankings is adopted by two voters and the other is adopted by one voter.

IANC uses root profiles to represent voters' preferences through an application of the Dixon-Wilf algorithm which enables the root profiles to be generated from the uniform distribution for *m* alternatives and *n* voters. That is, each root profile is generated uniformly with probability 1/*card*(*R(m, n)*). The formula for *card*(*R(m, n)*) and the details of the application of the Dixon-Wilf algorithm are given by Eğecioğlu and Giritligil (2011).

Likelihood Measures: Types of Likelihood

Two types of probabilities are considered to measure the likelihood of implementing the Borda outcome with truncated preference orderings.

1. $\boldsymbol{Pr_1} = \boldsymbol{Pr_1}(m, n, r)$ refers to the likelihood of choosing the entire set of Borda winners when only the first *r* rows of a preference profile, $P^r_{m,n}$, are considered. In other words, $\boldsymbol{Pr_1}$ is the probability that $B_r = B$.

For a given preference profile $P_{m,n}$, consider the random variable:

$$\boldsymbol{f_1}(P^r_{m,n}) = 1, \text{ if } B_r = B \text{ and}$$
$$0, \text{ otherwise.}$$

Given the distribution of profiles to be generated for given *m*, *n* and *r*, the approximate $\boldsymbol{Pr_1}$ is computed through Monte Carlo integration based on the law of large numbers. The law of large numbers implies that the average of a random sample from a large population is likely to be close to the mean of the whole population. That is, $\boldsymbol{Pr_1}$ is defined through the random variables $\boldsymbol{f_1}$:

$$\boldsymbol{Pr_1} = \frac{1}{card(R(m,n))} \sum_{P_{m,n} \in R(m,n)} f_1(P^r_{m,n}).$$

2. $\boldsymbol{Pr_2} = \boldsymbol{Pr_2}(m, n, r)$ is the likelihood that an *r*-Borda winner is among the Borda winners. Thus, it is the likelihood of an element of B_r to be also an element of B. For a given $P_{m,n}$, consider the random variable:

$$\boldsymbol{f_2}(P^r_{m,n}) = \frac{card(B \cap B_r)}{card(B_r)}$$

(Note that if $\boldsymbol{f_1}(P^r_{m,n}) = 1$, then $B_r = B$, and consequently, $\boldsymbol{f_2}(P^r_{m,n}) = 1$.)

Then, through the above explanation on approximation,

$$\boldsymbol{Pr_2} = \frac{1}{card(R(m,n))} \sum_{P_{m,n} \in R(m,n)} f_2(P^r_{m,n}).$$

Below given some examples regarding the calculations of $f_1(P^r_{m,n})$ and $f_2(P^r_{m,n})$:

Example 1

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

Note that B = {b}. For r = 1, B_1 = {a}. Since {a} ≠ {b}, $f_1(P^1_{4,5}) = 0$. Also $f_2(P^1_{4,5}) = 0$ because {a} ∩ {b}= ∅. For r = 2, B_2 = {a}. So, $f_1(P^2_{4,5}) = 0$ and $f_2(P^2_{4,5}) = 0$.

Example 2

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

For the above profile, B= {b, c} and B_1 = {a, b}, so $f_1(P^1_{4,5}) = 0$. Since {b, c} ∩ {a, b} = {b}, $f_2(P^1_{4,5}) = 1/2$. For r = 2, B_2 = {a, b, c}, so the profile yields $f_1(P^2_{4,5}) = 0$ and $f_2(P^2_{4,5}) = 2/3$ (since {b, c} ∩ {a, b, c}={b, c}).

The tools provided by Eğecioğlu and Giritligil (2011) allow for the generation of roots profiles from *R(m,n)* with probability 1/*card*(*R(m, n)*). The actual probability can then be approximated as follows: generate a large number of root profiles from *R(m, n)* with uniform probability 1/*card*(*R(m, n)*), where each selection is independent of the others. If *S(m, n)* denote the set of these generated profiles, then the law of large numbers implies that

$$\boldsymbol{Pr_1} = \frac{1}{card(S(m,n))} \sum_{P_{m,n} \in S(m,n)} f_1(P^r_{m,n}) \qquad (1)$$

Similarly,

$$\boldsymbol{Pr_2} = \frac{1}{card(S(m,n))} \sum_{P_{m,n} \in S(m,n)} f_2(P^r_{m,n}) \qquad (2)$$

Note that for (1) and (2) to result in a valid Monte Carlo algorithm for the computation of $\boldsymbol{Pr_1}$ and $\boldsymbol{Pr_2}$ respectively, it is essential that each $P_{m,n}$ in *S(m,n)* be drawn from the uniform probability on *R(m, n)*.

Monte Carlo Experiments

At the heart of the Monte Carlo experiments of this study is the Mathematica program GenerateRoot[m, n] (the Mathematica notebook containing this function can be accessed online for experimentation: see Eğecioğlu, 2004). The program GenerateRoot [m, n] takes two integers m and n as input parameters and generates a root profile in a matrix form $m \times n$ as output. The preference profile generated each time by GenerateRoot [m, n] is guaranteed to be distributed over the *R(m, n)* roots uniformly. To be able to estimate the probabilities $\boldsymbol{Pr_1}$ and $\boldsymbol{Pr_2}$ through the formulations (1) and (2) by using the law of large numbers, the preference profiles generated must be uniform over the set of roots *R(m, n)*: GenerateRoot[m, n] does exactly that.

The design of the Monte Carlo experiments is as follows. One thousand root profiles are generated for each value of the parameters m, n under consideration. Thus, *card(S(m, n))* = 1,000. The ranges $1 \le m \le 30$ and $1 \le n \le 30$ for most of the Monte Carlo experiments carried out. The basic steps followed for the computation of $\boldsymbol{Pr_1}$ and $\boldsymbol{Pr_2}$ in the symbolic algebra package Mathematica are:

1. Generate the values of m and n themselves, $1 \le m, n \le 30$ iteratively by means of two nested loops.
2. For the given values of m and n, invoke the function GenerateRoot[m, n], which generates a preference profile $P_{m,n}$ from the uniform distribution on the set of root profiles *R(m, n)*.
3. Compute the set of Borda winners B for the profile $P_{m,n}$ returned.

4. For every value of r in the range $1 \leq r < m$, detect the set of r_Borda winners B_r by considering only the first r rows of the profile $P_{m,n}$.
5. For given m, n and r, compute the random variables $\boldsymbol{f_1}$ and $\boldsymbol{f_2}$ using the sets B and B_r detected.

Steps 2 through 5 are executed *card(S(m, n))* times. The approximations to $\boldsymbol{Pr_1}$ and $\boldsymbol{Pr_2}$ for given m, n and r are calculated afterwards by dividing the sum of the computed values of $\boldsymbol{f_1}$ and $\boldsymbol{f_2}$ in Step 5 by *card(S(m, n))*.

Experimental Results on Pr_1 Type Probabilities

The computed $\boldsymbol{Pr_1}$ type probabilities are shown in Table 1 for $1 \leq m \leq 30$, $n = 30$ and $1 \leq r \leq m$. The rows are indexed by m and the columns are indexed by r. For instance, when $m = 5$, $\Pr[B_1=B] = 0.51$, $\Pr[B_2=B] = 0.671$ and $\Pr[B_3=B] = 0.818$ for $r = 1$, $r = 2$ and $r = 3$, respectively.

Figure 1 is a three-dimensional plot of the computed $\boldsymbol{Pr_1}$ type probabilities. $\boldsymbol{Pr_1}$ appears to be independent of n especially as the value of n increases. A close observation shows that for n fixed at 30, $\boldsymbol{Pr_1}$ approaches to zero as m gets large for a fixed r, and the behavior is roughly as $(1+r) / m$. A least-squares fit model was carried out for $1 \leq m \leq 30$, by considering the family of functions of the form

$$f(m, r) = c(1+ r)/ m$$

where c is a constant. The best approximating function in the least-squares sense was found to be

$$\boldsymbol{Pr_1} \sim f(m, r) = r/m + 1.4/m \qquad (3)$$

Figure 2 is a three-dimensional plot of the values of the approximating function (3). Comparing Figure 2 with the plot of the actual probabilities shown in Figure 1, (3) is observed to be a fine approximation of $\boldsymbol{Pr_1}$.

Table 1: For $1 \leq m \leq 30$, $n = 30$ and $1 \leq r \leq m$, the Probability that the set of r-Borda Winners is Equal to the Set of Actual Borda Winners

1.														
1.	1.													
0.799	1.	1.												
0.595	0.8	1.	1.											
0.51	0.671	0.818	1.	1.										
0.444	0.582	0.704	0.832	1.	1.									
0.363	0.521	0.622	0.738	0.843	1.	1.								
0.344	0.478	0.579	0.691	0.775	0.873	1.	1.							
0.274	0.404	0.512	0.603	0.7	0.79	0.878	1.	1.						
0.259	0.38	0.471	0.576	0.664	0.741	0.823	0.909	1.	1.					
0.226	0.362	0.459	0.548	0.629	0.695	0.774	0.838	0.91	1.	1.				
0.202	0.308	0.386	0.489	0.558	0.629	0.701	0.772	0.849	0.904	1.	1.			
0.198	0.312	0.387	0.461	0.531	0.598	0.661	0.721	0.786	0.845	0.925	1.	1.		
0.208	0.296	0.358	0.44	0.512	0.581	0.632	0.689	0.769	0.824	0.885	0.93	1.	1.	
0.184	0.269	0.352	0.413	0.476	0.546	0.606	0.671	0.734	0.782	0.821	0.867	0.922	1.	1.
0.16	0.245	0.338	0.387	0.437	0.498	0.572	0.626	0.701	0.737	0.794	0.861	0.903	0.954	1.
0.175	0.259	0.33	0.387	0.443	0.498	0.551	0.599	0.643	0.691	0.752	0.805	0.851	0.893	0.946
0.131	0.224	0.286	0.334	0.394	0.449	0.504	0.559	0.617	0.671	0.716	0.762	0.822	0.866	0.901
0.133	0.209	0.271	0.332	0.389	0.449	0.51	0.551	0.607	0.655	0.709	0.754	0.791	0.837	0.878
0.117	0.204	0.259	0.326	0.388	0.436	0.475	0.513	0.556	0.613	0.648	0.7	0.748	0.785	0.832
0.124	0.198	0.259	0.309	0.365	0.416	0.456	0.5	0.543	0.596	0.631	0.662	0.714	0.754	0.801
0.116	0.188	0.247	0.298	0.332	0.378	0.435	0.471	0.508	0.562	0.604	0.655	0.691	0.73	0.77
0.089	0.17	0.222	0.269	0.319	0.361	0.39	0.448	0.495	0.532	0.573	0.611	0.657	0.709	0.753
0.096	0.181	0.211	0.251	0.302	0.341	0.387	0.437	0.48	0.517	0.554	0.601	0.644	0.688	0.733
0.107	0.17	0.216	0.263	0.302	0.329	0.38	0.422	0.459	0.506	0.549	0.596	0.635	0.669	0.704
0.079	0.138	0.183	0.22	0.271	0.305	0.346	0.38	0.423	0.454	0.497	0.543	0.573	0.602	0.635
0.086	0.159	0.203	0.248	0.283	0.312	0.353	0.383	0.419	0.455	0.497	0.534	0.566	0.598	0.63
0.078	0.141	0.186	0.232	0.274	0.299	0.337	0.369	0.402	0.438	0.465	0.509	0.543	0.574	0.609
0.076	0.145	0.182	0.229	0.27	0.303	0.339	0.383	0.416	0.444	0.468	0.507	0.54	0.583	0.603
0.07	0.124	0.173	0.214	0.246	0.291	0.319	0.34	0.374	0.418	0.451	0.491	0.518	0.547	0.572

Figure 1: The Probability that *r*-Borda Winners are Identical to the Actual Borda Winners

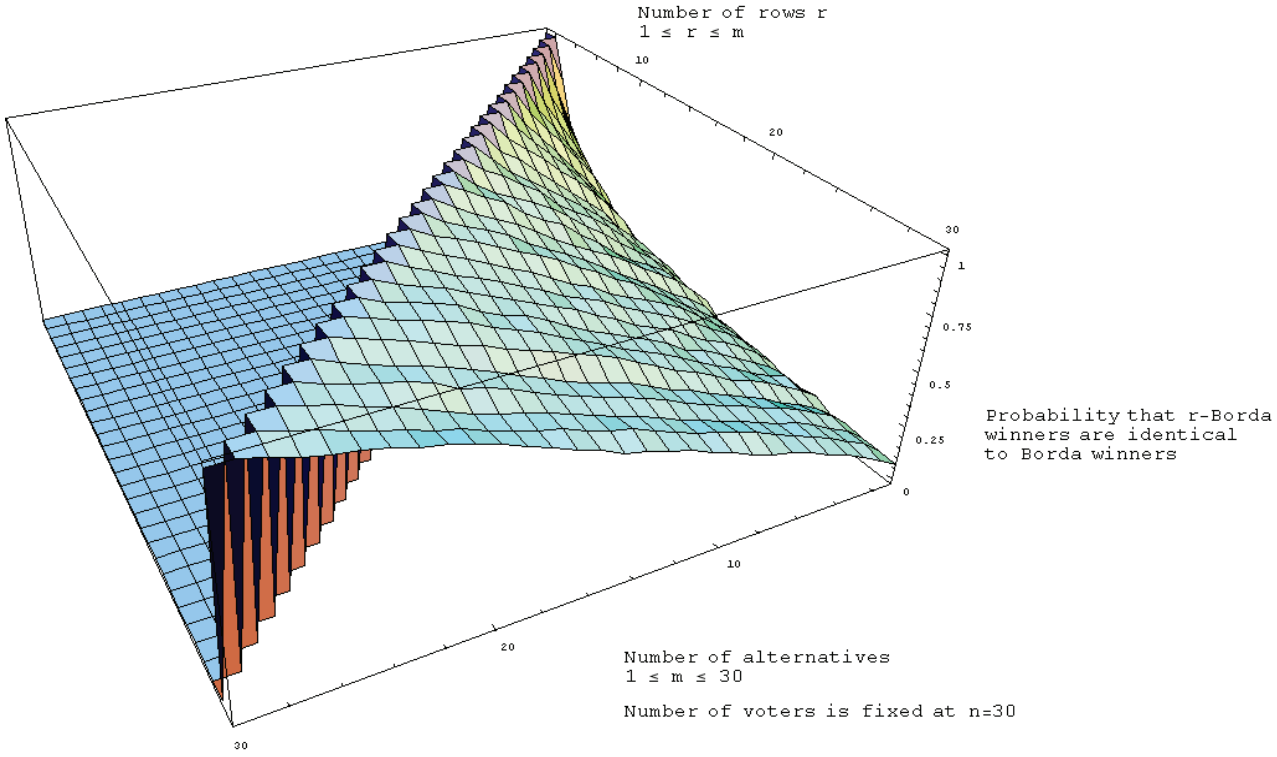
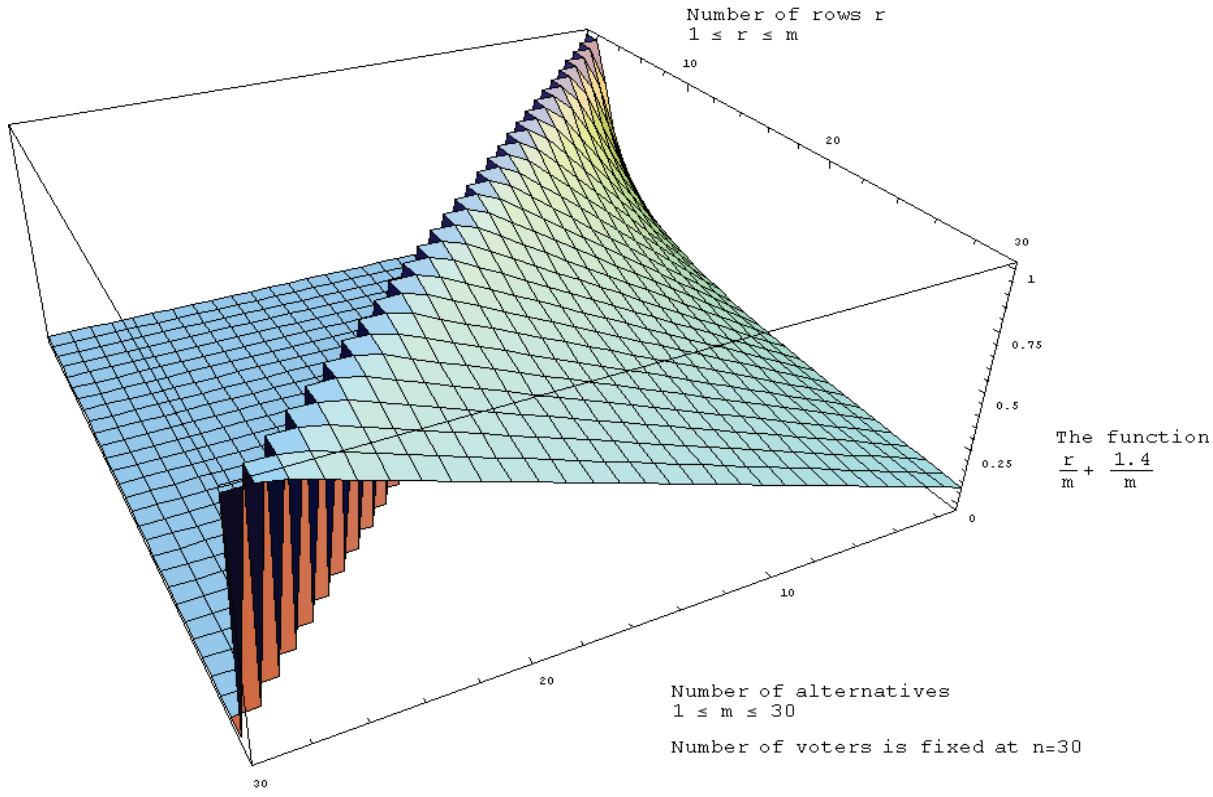


Figure 2: Approximating Function of the Probability that *r*-Borda Winners are Identical to Actual Borda Winners



To summarize, for large values of n, the likelihood of choosing all Borda winners by considering only the first r rows of a preference profile is independent of n and increases as the ratio r/m increases. It is impossible, however, to guarantee the exact Borda outcome unless r is set to be equal to $m-1$ or m.

Experimental Results on Pr_2 Type Probabilities

Table 2 shows the computed $\boldsymbol{Pr_2}$ type probabilities for $1 \leq m \leq 30$, $n = 30$ and $1 \leq r \leq m$. Again, the rows are indexed by m and the columns are indexed by r. Figure 3 is a three-dimensional plot of the computed $\boldsymbol{Pr_2}$ type probabilities.

It is observed that, as in the case of the $\boldsymbol{Pr_1}$ type probabilities, for n fixed at 30, $\boldsymbol{Pr_2}$ approaches to zero as m gets large for a fixed r and the behavior is roughly as $(1 + r)/m$. The properties of the analytic approximations can be employed to surmise that

$$\boldsymbol{Pr_2} \sim f(m, r) = r/m + 2.1/m \qquad (4)$$

The plot of this function is given in Figure 4.

Comparing Figure 4 with the plot of the actual probabilities shown in Figure 3, (4) is observed to be a fine approximation of $\boldsymbol{Pr_2}$.

The results show that, for large values of n, the likelihood that an r-Borda winner is one of the actual Borda winners is independent of n and increases as the r/m ratio increases. Results from this study show that, for $r = m - 2$, $\boldsymbol{Pr_2}$ approaches to 1 as m increases. Given this computational data, it can also be conjectured that, for any fixed k, $\boldsymbol{Pr_2}$ approaches to 1 for $r = m - k$, however, the rate of convergence decreases for larger k.

Table 2: For $1 \leq m \leq 30$, $n = 30$ and $1 \leq r \leq m$, the Probability that the Set r-Borda Winners to be a Subset of the Set of Actual Borda Winners

1.														
1.	1.													
0.844	1.	1.												
0.685	0.87	1.	1.											
0.607	0.759	0.895	1.	1.										
0.537	0.663	0.788	0.907	1.	1.									
0.474	0.618	0.706	0.827	0.919	1.	1.								
0.429	0.545	0.643	0.754	0.837	0.928	1.	1.							
0.371	0.475	0.582	0.666	0.761	0.849	0.936	1.	1.						
0.352	0.444	0.533	0.63	0.715	0.796	0.877	0.954	1.	1.					
0.313	0.428	0.519	0.602	0.681	0.754	0.827	0.893	0.96	1.	1.				
0.285	0.362	0.44	0.54	0.604	0.678	0.749	0.82	0.893	0.954	1.	1.			
0.278	0.372	0.445	0.512	0.579	0.647	0.712	0.774	0.84	0.906	0.971	1.	1.		
0.278	0.36	0.422	0.489	0.563	0.626	0.678	0.743	0.812	0.871	0.932	0.971	1.	1.	
0.252	0.318	0.395	0.462	0.522	0.592	0.655	0.716	0.773	0.82	0.863	0.913	0.967	1.	1.
0.225	0.301	0.386	0.437	0.483	0.541	0.616	0.676	0.74	0.784	0.839	0.904	0.944	0.982	1.
0.239	0.301	0.37	0.418	0.476	0.529	0.582	0.631	0.681	0.73	0.785	0.842	0.891	0.931	0.976
0.187	0.264	0.328	0.371	0.428	0.483	0.535	0.586	0.644	0.698	0.745	0.794	0.854	0.898	0.928
0.196	0.257	0.309	0.365	0.423	0.48	0.542	0.584	0.635	0.684	0.736	0.779	0.822	0.875	0.909
0.18	0.247	0.299	0.36	0.42	0.469	0.506	0.547	0.588	0.639	0.679	0.728	0.776	0.819	0.862
0.192	0.25	0.303	0.355	0.406	0.446	0.485	0.529	0.576	0.627	0.669	0.699	0.75	0.797	0.839
0.173	0.231	0.281	0.323	0.36	0.409	0.465	0.503	0.541	0.591	0.634	0.683	0.724	0.766	0.807
0.151	0.211	0.259	0.302	0.347	0.385	0.423	0.475	0.522	0.558	0.601	0.645	0.687	0.745	0.782
0.154	0.222	0.252	0.291	0.334	0.371	0.424	0.47	0.51	0.551	0.593	0.635	0.681	0.719	0.765
0.154	0.212	0.247	0.293	0.325	0.354	0.4	0.446	0.48	0.531	0.573	0.617	0.66	0.692	0.728
0.127	0.172	0.213	0.256	0.301	0.333	0.377	0.409	0.447	0.486	0.53	0.576	0.604	0.64	0.673
0.135	0.195	0.233	0.271	0.305	0.338	0.38	0.408	0.45	0.484	0.523	0.557	0.593	0.623	0.656
0.129	0.179	0.22	0.261	0.295	0.32	0.362	0.393	0.423	0.459	0.49	0.534	0.57	0.602	0.631
0.132	0.184	0.219	0.259	0.298	0.335	0.371	0.411	0.443	0.472	0.499	0.539	0.579	0.616	0.638
0.127	0.168	0.207	0.251	0.279	0.322	0.349	0.368	0.406	0.448	0.481	0.52	0.549	0.576	0.603

Figure 3: Probability that an *r*-Borda Winner is a Borda Winner

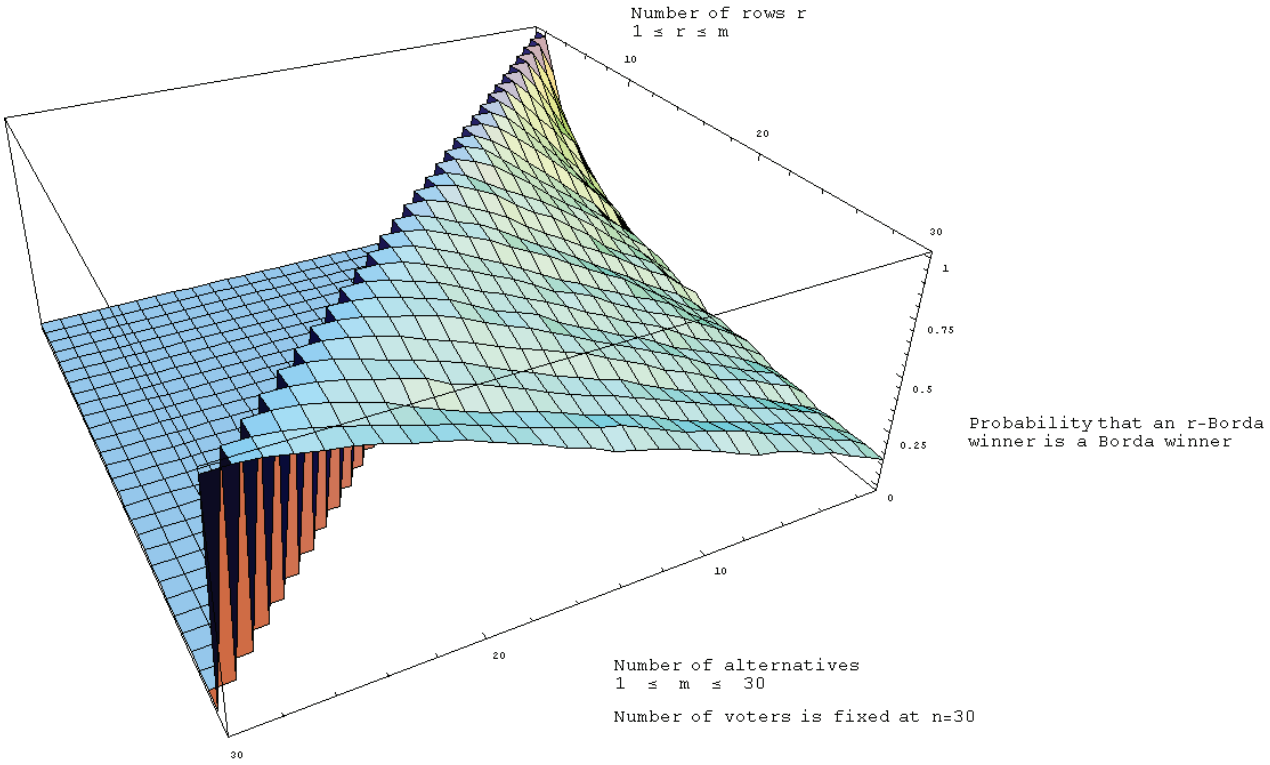
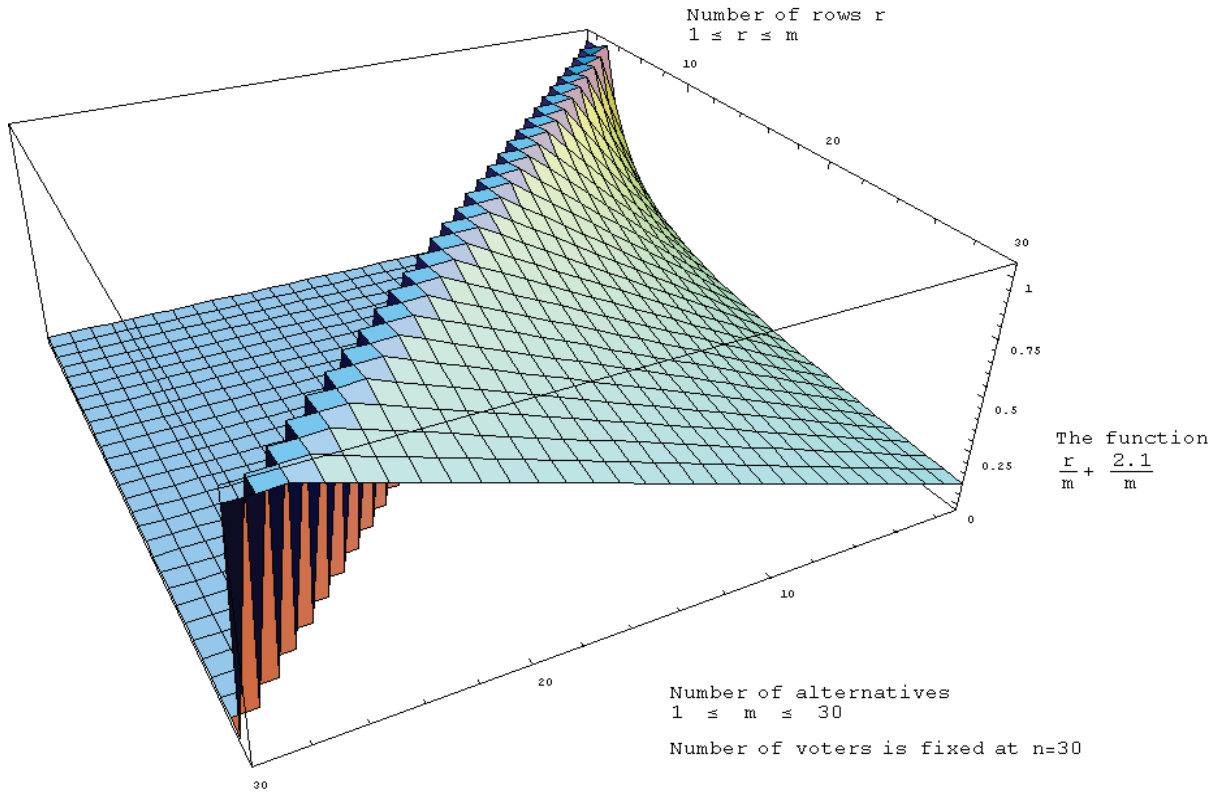


Figure 4: Approximating Function of the Probability that an *r*-Borda Winner is a Borda Winner



Conclusion

The Borda rule is one of the most studied voting procedures in the social choice theory literature. However, despite its well-known superiorities concerning the fulfillment of important positive and normative it is very difficult to be implemented in practice since it requires voters to rank all the alternatives at stake.

This computational study investigates the likelihood of implementing Borda outcome when n voters are asked to report only the first r ($1 \leq r < m$) ranks of their linear preferences over m alternatives. The truncated individual preferences are aggregated through a Borda-like method called the r-Borda rule.

The voters' preferences are sampled via IANC model which is an equiprobability assumption neglecting the names of both alternatives and voters.

The results of the Monte Carlo simulations indicate that, for large values of n, the likelihood of choosing exactly the set of Borda winners by considering only the first r ranks of voter preference orderings is independent of n, and approaches to zero as m gets large for a fixed r. Through the least square fit method, it is shown that, for any m, it is impossible to guarantee the exact Borda outcome with partial rankings over the alternatives.

It is observed that the likelihood that an r-Borda winner to be among the Borda winners is also independent of n and approaches to zero as m gets large for a fixed r. Our results show that for $r = m - k$, k being fixed, this probability approaches to one as m increases.

Some immediate directions exist for further research on this topic. First, although the r-Borda rule, as an equal-distance scoring method, is an intuitive way of aggregating the truncated preferences, computational studies can be designed to compare the success of assigning different score vectors to the reported ranks from the perspective of implementing the Borda outcome. Second, given truncated preferences of voters, the likelihood of implementing other well-known ranked rules can be investigated. However, especially in the case of pair-wise majority rules (such as the Condorcet rule), it should be noted that the methods used for aggregating truncated preferences are not as straightforward or intuitive as in the case of scoring methods. Hence, a similar study for such rules calls for theoretical and computational research.

References

Amy, D. C. (2000). *Behind the ballot box: A citizen's guide to voting systems*. Westport, CT: Praeger.

Berg, S., & Lepelley, D. (1994). On probability models in voting theory. *Statistica Neerlandica*, *48*,133-146.

Borda, J. C. (1781). Mémoire sur les élections au scrutin. *Histoire de l'Académie Royale des Sciences*, Paris.

Brams, S., & Fisburn, P. C. (2002). Voting procedures. In *Handbook of social choice and welfare*, *Vol.* 1, K. J. Arrow, A. K. Sen & K. Suzumura (Eds.), Amsterdam: Elsevier Science, 175-236.

Brams, S., Hansen, M. W., & Orrison, M. E. (2006). Dead heat: The 2006 public choice society election. *Public Choice*, *128*, 361-366.

Condorcet, M. (1785). Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. Paris.

Doğan, O. and Giritligil, A.E. (2011). Anonymous and neutral social choice: an existence result on resoluteness", Istanbul Bilgi University Murat Sertel Center for Advanced Economic Studies Working Papers, 2011:2.

Eğecioğlu, Ö. (2004). GenerateRoot[m,n], mathematica function. Available online as a Mathematica notebook *randomvote.nb* in the directory http://www.cs.ucsb.edu/~omer/.

Eğecioğlu, Ö., & Giritligil, A. E. (2011). The impartial, anonymous and neutral culture model: a probability model for sampling public preference structures, *Journal of Mathematical Sociology*, forthcoming.

Falmagne, J., & Regenwetter, M. (1996). A random utility model for approval voting. *Journal of Mathematical Psychology*, *40*, 152-159.

Fishburn, P. C., Gehrlein, W. (1978). Condorcet paradox and anonymous preference profiles, *Public Choice*, *26*, 1-18.

Fishburn, P. C. (1984). Discrete mathematics in voting and group choice. *SIAM Journal of Algebraic and Discrete Methods*, *5*, 263-275.

Gehrlein, W. V., & Fishburn, P. C. (1980). Robustness of positional scoring over subsets of alternatives. *Applied Mathematics and Optimization*, *6*, 241-255.

Gehrlein, W. V. (1981). Single stage election procedures for large electorates. *Journal of Mathematical Economics*, *8*, 263-275.

Gehrlein, W. V., Gopinath, B., Lagarias, J. C., & Fishburn, P. C. (1982). Optimal pairs of score vectors for positional scoring rules. *Applied Mathematics and Optimization*, *8*, 309-324.

Gehrlein, W. V. (1997). Condorcet's paradox, the Condorcet efficiency of the voting rules. *Mathematica Japonica*, *45*, 173-199.

Gehrlein, W. V., & Lepelley, D. (2000). The probability that all weighted scoring rules elect the same winner. *Economic Letters*, *66*, 191-197.

Guilbaud, G. T. (1952). Les théories de l'intéret genéral et le problémelogique de l'agrégation. *Economie Appliquée*, *5*, 501-584.

Lawson, B. L., Orrison, M., & Uminsky, D. T. (2006). Spectral analysis of supreme court. *Mathematics Magazine*, *79*, 340-346.

Niemi, R. G., & Riker, W. H. (1976). The choice of voting systems. *Scientific American*, *234*, 21-27.

Nurmi, H. (1987). *Comparing voting systems*, Dordrecht: Reidel.

Pattanaik, P. K. (2002). Positional rules of collective decision making, In *Handbook of Social Choice and Welfare*, *Vol.1*, K. J. Arrow, A. K. Sen and K. Suzumura (Eds.), Amsterdam: Elsevier Science, 361-394.

Regenwetter, M., & Grofman, B. (1998). Approval voting, Borda winners, and Condorcet winners: Evidence from seven elections. *Management Science*, *44*, 520-533.

Saari, D. G. (1987). The source of some paradoxes from social choice and probability. *Journal of Economic Theory*, *41*, 1-22.

Saari, D. G. (1989). A dictionary for voting paradoxes. *Journal of Economic Theory*, *48*, 443-475.

Saari, D. G. (1990). Susceptibility to manipulation. *Public Choice*, *64*, 21-41.

Saari, D. G. (2001). *Chaotic elections! A mathematician looks at voting*. Providence, RI: American Mathematical Society.

Sertel, M. R. (1987). A non-dictatorial compromise. *Social Choice and Welfare*, *4*, 1-11.

Vandercruyssen, D. (1999). Analysis of voting procedures in one-seat elections: Condorcet efficiency and Borda efficiency, Discussion Paper Series DPS 99.11. Kathlieke Universiteit Leuven, Leuven: Center for Economic Studies.