Soc Choice Welfare (2004) 22: 539–566
DOI: 10.1007/s00355-003-0232-z

Social Choice and Welfare

Approval voting and positional voting methods: Inference, relationship, examples

Michel Regenwetter[1], Ilia Tsetlin[2]

[1] Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA (e-mail: regenwet@uiuc.edu)
[2] INSEAD, 1 Ayer Rajah Ave, Singapore, 138676 (e-mail: ilia.tsetlin@insead.edu)

Received: 22 May 2002/Accepted: 4 March 2003

Abstract. *Approval voting* is the voting method recently adopted by the *Society for Social Choice and Welfare*. *Positional voting methods* include the famous plurality, antiplurality, and Borda methods. We extend the inference framework of Tsetlin and Regenwetter (2003) from majority rule to approval voting and all positional voting methods. We also establish a link between approval voting and positional voting methods whenever Falmagne et al.'s (1996) *size-independent model of approval voting* holds: In all such cases, approval voting mimics some positional voting method.

We illustrate our inference framework by analyzing approval voting and ranking data, with and without the assumption of the size-independent model. For many of the existing data, including the Society for Social Choice and Welfare election analyzed by Brams and Fishburn (2001) and Saari (2001), low turnout implies that inferences drawn from such elections carry low (statistical) confidence. Whenever solid inferences are possible, we find that, under certain statistical assumptions, approval voting tends to agree with positional voting methods, and with Borda, in particular.

Michel Regenwetter thanks the National Science Foundation for funding this research through NSF grant SBR 97-30076. Both authors thank the Fuqua School of Business for financially supporting their collaboration. Most of this research was done while Regenwetter was a faculty member at Fuqua. We thank Prof. Steven Brams for his valuable comments as a discussant of a previous version of this paper, given at the 2002 Public Choice meeting, and Prof. Donald Saari for his helpful comments in conversations and on another draft. We also thank the editor in charge and a referee for their valuable comments.

Tsetlin acknowledges the support of the Centre for Decision Making and Risk Analysis at INSEAD.

In 1999 the Society for Social Choice and Welfare and, in 2000, the Society for Judgment and Decision Making joined the family of academic and professional organizations that use *approval voting* for electing their president and board members. These societies include the American Statistical Association, the Mathematical Association of America, the National Academy of Sciences, and the Institute for Operations Research and the Management Sciences. Under approval voting, each voter indicates those candidates that s/he approves of. Each approved candidate receives a point, and the winner(s) is (are) the candidate(s) with the highest point total, summed over all voters. We abbreviate *'approval voting'* by 'AV'.

One of the advantages of AV is its obvious ease of use: Each voter's task amounts to simply choosing a subset of acceptable candidates from all listed candidates. The voter can consider each candidate in turn and either approve of him/her or not. Despite the great simplicity of the vote casting process, the ballots contain more information than just the voter's single top choice, the information used in plurality elections. When there are m many candidates, there are $2^m - 1$ degrees of freedom in AV, but only $m - 1$ degrees of freedom in plurality ballots.

Some researchers consider positional voting methods (also commonly referred to as *scoring rules*), in particular the Borda method, as a superior alternative to AV. Under positional voting methods, each voter provides a complete linear order of the candidates, and to each rank order position corresponds a certain score. For instance, the Borda scoring method assigns $m - 1$ points to the first ranked among m candidates, $m - 2$ points to the second ranked, etc. Under the plurality scoring rule, the candidate ranked first in a given ballot receives a point, and the remaining candidates score zero points. The winner(s) is (are) the candidate(s) with the highest score total, summed over all voters. We abbreviate *'positional voting method'* by 'PVM'.

The 2000 United States presidential election has taught us that the complexity of filling out even relatively simple ballots is not to be taken lightly. More generally, we have argued [30] and continue to make the point that the process, that leads from the voter's decision to vote all the way to the final certified vote counts, should be studied from a statistical point of view. Social choice theorists and analysts ought to take into account the multi-faceted uncertainties attached to such a process, especially for large electorates. The simplest nontrivial approach to treating this process statistically is to view ballots as outcomes of a sampling process, and make inferences from those ballot counts about the likely underlying distribution of intended ballot responses in the electorate.

In this paper, we lay out such an inference framework to quantify the likelihoods of all possible latent social welfare orders by approval voting given actual empirical approval voting (subset choice) ballot frequency counts. Similarly, we provide an inference framework to quantify the likelihoods of all possible latent social welfare orders by any positional voting method given actual empirical frequency counts on rankings. We also provide a statistical inference framework to quantify the confidence we can have in discrepancies or overlaps between approval voting and positional voting methods, for 1) experimental elections where both approval voting ballots were cast and preference rankings assessed, as well as 2) approval voting elections where only approval voting ballots were cast (i.e., no rankings were assessed).

Following our earlier work, we argue that to evaluate the adequacy of a voting method, we need to evaluate how it performs not just in principle, but also in actuality, namely in those situations where it is or has been actually used. Our statistical methodology allows to evaluate and compare AV and PVMs in a fashion that is intimately connected to empirical data.

Little effort has been spent on empirically evaluating and comparing the performance of AV and related methods (for some rare examples see [2, 4, 5, 13, 14, 23, 28]). For example, Brams and Fishburn [2] and Saari [28] compare several voting methods using the first election of the *Society for Social Choice and Welfare* conducted under AV. (This was an experimental election in the sense that the voters were asked to add to their official AV ballot an additional hypothetical ranking they would have used, if the election had been carried out under the Borda method.) Brams and Fishburn conclude that "the election was essentially a toss up."

In order to compare AV with PVMs in the absence of ranking data one needs to make some (ideally statistically testable) assumptions about the process by which the subsets are generated under AV. Various models have been proposed or used to illuminate this process [12, 17, 19–21, 26]. Only few of these papers, however, provide any statistical estimations and none provide thorough and rigorous quantitative assessments of the statistical confidence one can have in their findings.

This paper is a first theoretical study in that direction. Our statistical assumptions have multiple effects that trade off with each other: As correctly pointed out by a referee, forcing attention to specialized lower dimensional settings may seriously influence and even bias the conclusions that our results suggest. On the other hand, it is routine in most areas of science to develop simple, and thus restrictive, models: 1) We gain a great deal of mathematical tractability, and thus are able to derive analytical results and offer relevant illustrations of our statistical methodology. 2) The constraints imposed by our models make them statistically testable. This is a prerequisite for comparing theoretical models against empirical data in a meaningful fashion. 3) To the extent that our models provide a parsimonious, yet statistically adequate, description of empirical data we believe that their implications are informative. On the other hand, we readily acknowledge that other parsimonious explanations for the same data might exist and they, in turn, might suggest interpretations that would negate our findings. (For a whole paper on model dependence of social choice analyses see [25].)

The paper is organized as follows. In Sect. 2 we introduce a statistical framework to make inferences about population social welfare orders under approval voting or positional voting methods based on subset choice or ranking data. In Sect. 3 we provide some additional background information on the analysis of AV and PVMs (Borda in particular). We also re-introduce the *size-independent model of approval voting* of Falmagne and Regenwetter [12], which Regenwetter and Grofman [23] used in order to compare AV data with the Borda score and majority rule. We then state and discuss new theoretical results on the correspondence between AV and PVMs in situations where the size-independent model is an appropriate description of the probabilistic process that generates the AV ballot data. We show that in such situations, approval voting precisely mimics some positional voting method. In Sect. 4, we extend our inference framework to subset choice data that have originated via the size-independent model. In Sect. 5 we analyze several election data sets, including older experimental data of the *Institute of Management Science* and new data from the 2000 elections of the *Society for Social Choice and Welfare*, as well as the *Society for Judgment and Decision Making*.[1] The final section summarizes our findings. The Appendix describes a *method of bounds*, which uses probabilities on pairwise comparisons to place upper and lower bounds on our statistical confidence in the inferred social welfare orders over any number of candidates.

2 An inference framework

During the 2000 United States presidential election, the world was transfixed with the question who was the election winner and whether the majority of the electoral college votes, based on the certified ballot counts, was reflecting the 'correct' winner. From the uncertainty and eventual decision of each voter to vote, through the various steps involved in resolving his or her own uncertainties as to which candidate(s) they prefer and which candidate(s) they ought to vote for, to correctly or incorrectly casting that vote, having it correctly or incorrectly counted and the final counts certified, there are multiple facets and stages where noise or other probabilistic components enter the process. As we have argued before [30], the probability that a given voter's preference is correctly reflected in the certified counts of an election is strictly less than one. To our knowledge, social choice and voting theory has made little effort to integrate such uncertainties into its formal treatment of collective choice. Both this and the predecessor paper are attempting to redress that gap and to make the case that social choice theorists should more systematically incorporate statistical considerations into theories, models, and empirical analyses of social choice procedures.[2]

---

[1] We thank the Society for Judgment and Decision Making and the Society for Social Choice and Welfare for providing their ballot counts. We also requested recent ballot counts from the Institute for Operations Research and the Management Sciences, but our request was turned down by the executive committee.

[2] While the social choice and voting literature provides a wealth of knowledge about strategic aspects of voting, we will initially develop our inference framework with the implicit assumption of sincere voting. We leave it for future work to reconcile and synthesize statistical inference with game theoretic considerations.

We propose a natural and simple way to view the vote casting and ballot counting process as a process with an uncertain outcome: We treat actual ballot counts as outcomes of an experiment in the statistical sense. We ask how much statistical confidence we can have in the actual/empirical outcome of a given election with a given number of ballot counts and a given preference aggregation method if the process generating the ballot counts includes probabilistic components. We start with the simplest nontrivial scenario: A first approximation is to view the actual ballot counts as a simple random sample from an underlying population with some unknown probability distribution over ballot responses. In other words, we treat the ballot counts as originating from a multinomial distribution. This is one of the simplifying assumptions that may affect our final findings.[3]

More specifically, in this paper, when we use AV (or PVM), we will think of the certified AV ballots (or rankings) as a random sample of the true population of AV ballots (or rankings). In particular, the AV (or PVM) winner based on the certified ballot counts may or may not be the same as the true AV (or PVM) winner in the population from which the ballots were sampled. Besides integrating uncertainty into the analysis of voting ballots, this view point also naturally applies to traditional sample survey data.

We now introduce an inference framework to analyze AV and all PVMs from such a statistical point of view. We state mathematical formulae for the statistical confidence in social welfare relations derived from ballots or other sample data. These results are used and illustrated in the data analyses provided in Sect. 5.

Following a similar logic as in Tsetlin and Regenwetter [30] we tackle inference from a standard Bayesian view point. We assume a prior probability distribution over each variable of interest, and we update this distribution based on available data, to obtain a posterior distribution.[4] Following standard usage in theoretical Bayesian studies, and for mathematical convenience, we use the conjugacy classes of the multinomial and Dirichlet distributions. Throughout the analysis, we assume the prior distribution to be noninformative, which yields results that are also consistent with maximum likelihood estimation. We assume that the data come from random sampling with replacement (or without replacement from an extremely large population). As a consequence, the sampling distribution of the ballot counts is multinomial and conjugated to it is the Dirichlet distribution. We use the latter to derive aggregate pairwise comparison probabilities, i.e., probabilities that a given candidate beats a given other candidate in the population under a given social welfare function, for a given set of ballot counts. A particular case of a multinomial distribution is the binomial, and naturally conjugated to it is the beta distribution. The Appendix provides a method of bounds

---

[3] Note also, that much more constrained assumptions, such as the impartial culture assumption, are common practice in social choice theory.

[4] For more details on Bayesian methodology and terminology see, e.g., [6].

which allows to estimate the statistical confidence we can have in any social welfare order, based on these aggregate paired comparison probabilities. This method was originally proposed by Tsetlin and Regenwetter [30] for the parallel problem of inferring majority rule outcomes from probabilistic ballots. In what follows, we use the word *population* in the statistical sense of a theoretical population from which empirical data are generated. We use the word *sample* to refer to empirical ballot data that were generated via a probabilistic process. We use boldface letters to denote random variables and regular font to denote numbers.

2.1 Approval voting

We now provide the probabilities of all possible AV outcomes in the population given subset data. Writing p_S for the probability that a randomly drawn member of the population approves of set S, the AV score for candidate c, in a sample of size N, has a binomial distribution with number of trials N and probability of success $\sum_{S:c\in S} p_S$. Its natural conjugate family is the beta distribution. To make statistical inferences based on empirical data, we need to derive the probability that candidate a is preferred to candidate b, by AV, in the population, given the data. For two candidates a, b we denote by N_{ab} the number of observed subset choices in which candidate a is approved and candidate b is not approved, and by $\boldsymbol{p}_{ab}$ the probability that a randomly drawn member of the population approves a and does not approve b. Following the standard convention in Bayesian statistics, we write D_{av} for the observed (AV ballot) data. We write $a \succ_{\boldsymbol{p},av} b$ for the event that a beats b by AV in the population (the boldface $\boldsymbol{p}$ indicates that this relationship is uncertain, i.e., a probabilistic event), $f_\beta(x, \alpha, \beta)$ for the beta distribution with parameters α and β and $F_\beta(x, \alpha, \beta)$ for its cumulative distribution function.

Observation 1. *Given N_{ab} and N_{ba} in the sample, and given the parameters α_{ab} and α_{ba} of the prior beta distribution, the posterior probability that a is preferred to b by AV in the population,* $P((a \succ_{\boldsymbol{p},av} b)\ |D_{av})$, is given by

$$P\Big((a \succ_{\boldsymbol{p},av} b)\ |D_{av}\Big) = F_\beta\left(\frac{1}{2}, N_{ba} + \alpha_{ba}, N_{ab} + \alpha_{ab}\right). \tag{1}$$

Let $N = N_{ab} + N_{ba}$, $\alpha = \alpha_{ab} + \alpha_{ba}$, $t_{ab} = \frac{N_{ab}+\alpha_{ab}}{N+\alpha}$. It follows from the properties of the beta distribution that, for large N, the posterior distribution of $(\boldsymbol{p}_{ab} - \boldsymbol{p}_{ba})$ can be approximated by a normal distribution with mean $\mu = t_{ab} - t_{ba}$ and variance $\sigma^2 = \frac{1}{N+\alpha}(t_{ab} + t_{ba} - (t_{ab} - t_{ba})^2)$.

2.2 Positional voting methods

This subsection provides the probabilities of all possible PVM outcomes in the population given ranking data. Every PVM for m candidates can be specified by a voting vector $\bar{w} = (w_1, w_2, \ldots, w_m)$ as follows (Saari [27], p. 43).

Given the ranking $c_1 c_2 \ldots c_m$ of m many candidates, candidate c_1, who is ranked first, receives w_1 points, candidate c_2 receives w_2 points, ..., while the last candidate, c_m, receives w_m points. In the standard definition the sequence $w_1 \ldots w_m$ is nonincreasing and there is at least one i with $w_i > w_{i+1}$. The score of each candidate is the sum of all points for this candidate over all voters. Usually $\bar{w}$ is normalized so that $w_1 = 1$, $w_m = 0$, or such that $w_m = 0$, $\sum_{i=1}^{m} w_i = 1$. PVMs are also often referred to as *scoring rules*. Among the best known voting procedures that can be written as a PVM are the Borda method ($w_i = 1 - \frac{i-1}{m-1}$), plurality rule ($w_1 = 1, w_i = 0, i > 1$), and antiplurality ($w_i = 1$, $i < m$, $w_m = 0$).

To make statistical inferences based on empirical data, we need to derive the probability that candidate a is preferred to candidate b in the population by the PVM with voting vector $\bar{w}$, given the sample data. We denote by $N_{r_a=i,r_b=j}$ the number of voters who rank candidate a at the ith place from the top, and rank candidate b at the jth place from the top. Similarly, the corresponding probability that a is ranked at position i and that b is ranked at position j in the population is written as $\boldsymbol{p}_{r_a=i,r_b=j}$. We denote the collection $N_{r_a=i,r_b=j}$, $i = 1, \ldots, m$; $j = 1, \ldots, m$ by $\bar{N}$, and the collection $\boldsymbol{p}_{r_a=i,r_b=j}$, $i = 1, \ldots, m$; $j = 1, \ldots, m$ by $\bar{\boldsymbol{p}}$.

Observation 2. *If the parameters of the prior Dirichlet distribution $P(\bar{p})$ are $\alpha_{r_a=i,r_b=j}$, $\sum_{i,j} \alpha_{r_a=i,r_b=j} = \alpha$, then the posterior Dirichlet distribution $P(\bar{p}\,|\bar{N})$ is given by*

$$P(\bar{p}\,|\bar{N}) = \frac{\Gamma(\alpha + N)}{\prod_{i,j} \Gamma(\alpha_{r_a=i,r_b=j} + N_{r_a=i,r_b=j})} \prod_{i,j} p_{r_a=i,r_b=j}^{N_{r_a=i,r_b=j} + \alpha_{r_a=i,r_b=j} - 1}. \tag{2}$$

Let $t_{r_a=i,r_b=j} = \frac{\alpha_{r_a=i,r_b=j} + N_{r_a=i,r_b=j}}{\alpha + N}$. From the properties of the Dirichlet distribution (e.g., [6]) it follows that the mean of the posterior $\boldsymbol{p}_{r_a=i,r_b=j}$ is given by

$$E(\boldsymbol{p}_{r_a=i,r_b=j}) = t_{r_a=i,r_b=j}, \tag{3}$$

and the variance-covariance matrix is given by

$$Cov(\boldsymbol{p}_{r_a=i,r_b=j}, \boldsymbol{p}_{r_a=k,r_b=l}) = \frac{1}{\alpha + N} \left(t_{r_a=i,r_b=j} \delta_{ik} \delta_{jl} - t_{r_a=i,r_b=j} t_{r_a=k,r_b=l} \right), \tag{4}$$

where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

For $\alpha_{r_a=i,r_b=j} \ll N_{r_a=i,r_b=j}$, the quantity $t_{r_a=i,r_b=j}$ is the relative frequency with which a is ranked at the ith place and b is ranked at the jth place in the sample. The probability of the event $a \succ_{\boldsymbol{p},\bar{w}} b$ that candidate a is preferred to candidate b in the population by the PVM with voting vector $\bar{w}$, given the observed ranking data D_R, is then given by

$$P\Big((a \succ_{\boldsymbol{p},\bar{w}} b)|D_R\Big) = P\left(\sum_{i,j}(w_i - w_j)p_{r_a=i,r_b=j} > 0\right).$$

From (3) and (4) it follows that

$$E\left(\sum_{i,j}(w_i - w_j)\boldsymbol{p}_{r_a=i,r_b=j}\right) = \left(\sum_{ij}(w_i - w_j)t_{r_a=i,r_b=j}\right),$$

$$Var\left(\sum_{i,j}(w_i - w_j)\boldsymbol{p}_{r_a=i,r_b=j}\right) = \frac{1}{\alpha + N}$$

$$\times\left(\sum_{i,j}(w_i - w_j)^2 t_{r_a=i,r_b=j} - \left(\sum_{i,j}(w_i - w_j)t_{r_a=i,r_b=j}\right)^2\right). \tag{5}$$

As we state next, for sufficiently large N, we can approximate $(\sum_{i,j}(w_i - w_j)\boldsymbol{p}_{r_a=i,r_b=j})$ by a normal distribution with mean and variance given by (5). Denote by $F_N(x)$ the standard normal cumulative distribution function evaluated at x.

Observation 3. *Let the sample size be* N, *consider the voting vector* $\bar{w} = (w_1, w_2, \ldots, w_m)$, *and let the number of voters that rank candidate a at the ith place and candidate b at the jth place be* $N_{r_a=i,r_b=j}$. *For sufficiently high N the posterior probability* $P((a \succ_{\boldsymbol{p},\bar{w}} b)|D_R)$ *that candidate a is preferred to candidate b in the population by the PVM with voting vector* $\bar{w}$, *given the observed ranking data* D_R, *can be approximated as follows:*

$$P\Big((a \succ_{\boldsymbol{p},\bar{w}} b)|D_R\Big) \overset{approx}{=} F_N\left(\frac{\mu}{\sigma\sqrt{\alpha + N}}\right), \text{ where}$$

$$\mu = \sum_{i,j}(w_i - w_j)t_{r_a=i,r_b=j},$$

$$\sigma = \sqrt{\sum_{i,j}(w_i - w_j)^2 t_{r_a=i,r_b=j} - \left(\sum_{i,j}(w_i - w_j)t_{r_a=i,r_b=j}\right)^2},$$

$$t_{r_a=i,r_b=j} = \frac{\alpha_{r_a=i,r_b=j} + N_{r_a=i,r_b=j}}{\alpha + N}.$$

This completes our discussion of inferences for AV and PVMs when the only assumption about the data is that they were generated via random sampling from some unknown theoretical distribution. As a consequence, the ballots were assumed to have an unknown multinomial distribution. We now move on to a situation with more restrictive modeling assumptions, namely situations where we specify an explicit model that relates the ranking probabilities and subset choice probabilities in these sampling processes.

3 On the relationship between AV and PVMs

3.1 Some additional background on AV and PVMs

AV allows the voters to approve of any subset of candidates. A common assumption in social choice theory is to assume that each voter's preferences can be adequately described by a complete linear ordering of the choice alternatives. Contrary to AV, a PVM requires the voter to elicit and report such an ordering. The ballot generating process under AV has been modeled in several ways. We briefly describe four of the models that have been used in the literature because they have certain communalities that are of critical importance for some of the work that follows. Merrill argues that a voter should maximize his/her expected utility and approve of all candidates whose utilities are above the average utility (see also Brams and Fishburn's [1] discussion of optimal voting strategies). If a preference ranking is translated into utilities where neighboring candidates are equidistant from each other, and if each candidate is assumed to have an equal chance of winning, then the top half candidates are above average. This leads to the prescriptive rule that one *should* approve of approximately the top half of all candidates (irrespective of one's particular preferences). We will refer to this special case as *Merrill's prescriptive rule*.[5]

A second model for approval voting data was proposed by Marley [19], which was later labeled the *latent scale model* for approval voting [26]. According to this model, the voter considers each candidate in turn and, with some fixed probability that depends on the candidate, includes or excludes that candidate in the approval voting ballot.

A third model, used by Gehrlein and Lepelley [17], assumes that individual preferences are drawn from a uniform distribution over linear orders (impartial culture) and that set sizes are randomly drawn from a uniform distribution with the two random processes being independent. We refer to this as the *Gehrlein and Lepelley model*.

These three models are all special cases of the size-independent model of Falmagne and Regenwetter [12].[6] The latter is, however, considerably less restrictive. It assumes that 1) a voter approves of a top subset of candidates consistent with a linear ordering of the candidates and 2) that for any set of size k, the event that this set is a top set in the ranking and the event that the voter approves of k many candidates, are independent.

---

[5] Gehrlein [16] established that the prescriptive rule maximizes the probability among so-called 'constant scoring rules' (i.e., PVMs with weights equal to 0 or 1) that a Borda winner is elected, under assumption of the impartial culture and for large electorates.

[6] Merrill's prescriptive rule and Gehrlein and Lepelley's model are obvious special cases. Marley's latent scale model is a special case in a less obvious and somewhat indirect sense: It has been shown that its subset choice probabilities can always also be generated from a size-independent model [9, 22].

Formally, as before, denote by p_S the probability that a randomly drawn member of the population approves of set S. We use the abbreviation $p_{|S|} = \sum_{S':|S'|=|S|} p_{S'}$, and denote by p_{Π_S} the total probability of all those rank orders in which the elements of S are ranked higher than all other candidates. According to the size-independent model, a randomly chosen voter approves of $|S|$ many candidates with probability $p_{|S|}$, and

$$p_S = p_{|S|} p_{\Pi_S}. \tag{6}$$

In words: The probability that a randomly drawn voter chooses set S is the product of the probability that s/he votes for as many alternatives as S contains times the probability that s/he likes all candidates in S better than all others. There are also natural equivalent formulations as distribution-free random utility models [12, 26]. The geometry and combinatoric structure of the size-independent model via its associated approval voting (convex) polytope, including its dimensionality, are now well understood [7, 8, 10, 11].[7]

An interesting special case of the size-independent model arises by requiring that the voter independently elicits a linear ordering and a subset size k, and then approves of the top k elements in the linear ordering. This includes, e.g., Merrill's prescriptive rule and the Gehrlein and Lepelley model. For instance, the prescriptive rule corresponds to the size-independent model with the following probability distribution over subset sizes, regardless of one's preference order. When the number of candidates m is even,

$$p_{|S|} = \begin{cases} 0 & \text{if } |S| \neq \frac{m}{2}, \\ 1 & \text{if } |S| = \frac{m}{2}. \end{cases}$$

A similar correspondence obtains for odd m. Since a voter should be indifferent between approving a subset of size $\frac{m-1}{2}$ and a subset of size $\frac{m+1}{2}$, the corresponding probability distribution over subsets ought to be as follows:

$$p_{|S|} = \begin{cases} 0 & \text{if } |S| \neq \frac{m-1}{2} \text{ and } |S| \neq \frac{m+1}{2}, \\ \frac{1}{2} & \text{if } |S| = \frac{m-1}{2} \text{ or } |S| = \frac{m+1}{2}. \end{cases}$$

However, it is important to notice that the general size-independent model does not require the rankings and the set sizes to be independent. For instance, each linear order over 3 candidates, a, b, c might be held by $\frac{1}{6}$ of the voters. Suppose those voters with orders abc, bca, cab all vote for the single best candidate, whereas all others vote for their top two candidates. Then each subset of size one or two has probability $\frac{1}{6}$, but given a person's preference ordering, we know the chosen subset with probability one. Thus, the

[7] Various colleagues have expressed concern about the model's intuitive plausibility and about the strong constraints it places on empirical data. We view this as an asset of the model, given that it has held up quite well in empirical tests [12, 23]. We know of no other parsimonious probabilistic choice model that has been so successfully fit to approval voting data.

preference orderings and the subset sizes are maximally dependent. Yet, the size-independent model holds.

Using the size-independent model, Regenwetter and Grofman [24] compare AV with majority rule and the Borda score and, e.g., provide a general (model-dependent) formula to calculate the Borda outcome from approval voting ballots. They also show that the orders by AV scores and by Borda scores empirically tend to, but theoretically need not, be identical (even when the size-independent model holds). However, they do not investigate the relationship between AV and other PVMs, nor do they evaluate the statistical confidence one can have in any of their empirical findings.

We now provide some additional background information about PVMs. For three candidates a, b, c any voting vector $\bar{w}$ may be normalized as

$$\bar{w} = (1 - s, s, 0), \quad 0 \leq s \leq \frac{1}{2}. \tag{7}$$

As Saari ([27], p. 56) shows, the scores of three candidates under the PVM with this voting vector are linear (convex) combinations of the scores of these candidates with voting vectors corresponding to plurality and antiplurality: For a given preference profile let a_s, b_s, c_s be the scores of candidates a, b, c under voting vector $(1 - s, s, 0)$, a_p, b_p, c_p be the scores of these candidates under plurality voting vector $(1, 0, 0)$ and a_a, b_a, c_a be the scores of these candidates under antiplurality voting vector $(1, 1, 0)$. Then the following holds:

$$(a_s, b_s, c_s) = (1 - 2s)(a_p, b_p, c_p) + 2s(a_a, b_a, c_a). \tag{8}$$

Equation (8) defines the so-called "procedure line" (Saari [27], Eq. 2.4.3). Two consequences follow directly from (8): 1) If the plurality and antiplurality outcomes coincide, then *every* scoring rule has that same outcome, too. 2) If the plurality and antiplurality outcomes share the same winner (loser) then *every* PVM has that same winner (loser), too.

There has been some theoretical work on the relationship between AV and PVMs, as well as the relationship among PVMs (e.g., [16–18, 27, 29]). For instance, Gehrlein and Lepelley [17] compare the Condorcet efficiencies of PVMs and AV using a special case of the size-independent model with uniform distributions over rankings (impartial culture) and set sizes. Their results apply only to the impartial culture, i.e., a population where all candidates are tied (at the level of the population distribution) regardless of the voting method.[8] The next section establishes the much more general correspondence between AV and PVMs. This correspondence has its own restrictions, in that it assumes the size-independent model. On the other hand, it eliminates standard restrictions in the social choice literature by allowing for arbitrary populations (or cultures).

[8] In view of the nature of the impartial culture assumption this might also be interpreted as being a 'worst case scenario' comparison of AV and PVMs [31].

3.2 The relationship between AV and PVM under the size-independent model

We write $r_c = i$ for the event that candidate c is ranked at the ith place from the top by a given voter and $p_{r_c=i}$ for the corresponding probability. The probability that a randomly chosen voter approves of candidate c is given by $\sum_{S:c\in S} p_S$. Under the assumption of the size-independent model this equals

$$\sum_{S:c\in S} p_{|S|} p_{\Pi_S} = \sum_{|S|=1}^{m} \left(p_{|S|} \sum_{i=1}^{|S|} p_{r_c=i} \right)$$

$$= \sum_{i=1}^{m} \left(p_{r_c=i} \sum_{|S|=i}^{m} p_{|S|} \right) = \sum_{i=1}^{m} w_i p_{r_c=i}, \tag{9}$$

where $w_i = \sum_{|S|=i}^{m} p_{|S|=i}$.

By definition $p_{|S|} \geq 0$, since it is the total probability of all subsets with size $|S|$. One can see that if at least one $p_{|S|}$ is positive, $0 < |S| < m$, then we can normalize the weights w_i so that $w_1 = 1$ and $w_m = 0$, by setting

$$w_i = \frac{\sum_{|S|=i}^{m-1} p_{|S|}}{1 - p_0 - p_m}. \tag{10}$$

The following observation summarizes these results.

Observation 4. *Under the assumption of the size-independent model with subset size probabilities* $p_{|S|}$, *AV yields scores that are identical to the scores obtained from a PVM with voting vector* $w_i = \frac{\sum_{|S|=i}^{m-1} p_{|S|}}{1-p_0-p_m}$.

Corollary 1. *Merrill's prescriptive rule, as a special case of the size-independent model, corresponds to a PVM with voting vector (10) as follows, for even m,*

$$w_i = \begin{cases} 1 & \text{if } i \leq \frac{m}{2}, \\ 0 & \text{if } i > \frac{m}{2}. \end{cases}$$

A similar correspondence obtains for odd m. Here, the corresponding voting vector (10) is

$$w_i = \begin{cases} 1 & \text{if } i < \frac{m+1}{2}, \\ \frac{1}{2} & \text{if } i = \frac{m+1}{2}, \\ 0 & \text{if } i > \frac{m+1}{2}. \end{cases}$$

For $m = 3$ *candidates, this prescriptive rule yields approval voting ballot counts equal (in expectation) to the Borda score.*

In general, for large electorates and given the size-independent model, if all subset sizes are equally frequent, then AV and Borda mimic each other. (See [16] for a similar result using the impartial culture assumption.) Furthermore, if all set sizes are not equally frequent, but the size-independent model holds, then, by Observation 1, the AV outcomes for a large electorate lie on the procedure line of the PVM. There is thus a PVM, with voting vector specified by the subset size probabilities, that mimics AV. We will calculate these voting vectors in our data analysis.

While our earlier two developments about statistical inference, provided in Sect. 2, were atheoretical (other than presuming random sampling, they made no assumptions about the nature of the probabilistic process) we now provide an inference methodology to analyze subset choices when the probabilistic process takes the form spelled out by the size-independent model. Consequently, any conclusions drawn from the results in this section, stand and fall with the size-independent model.

4 Inference framework under the assumption of the size-independent model

The size-independent model allows to reconstruct from subset data the probability that a given candidate is ranked at a particular rank order position [10, 12]. This subsection provides the inference calculus to obtain rank position probabilities and the probabilities of all possible outcomes under PVMs in the population given subset data, assuming the size-independent model. We again state mathematical formulae for the statistical confidence in social welfare relations derived from ballots or other sample data, but now under the assumption that the size-independent model holds. These results are also used and illustrated in the data analyses in Sect. 5.

In an inference framework the posterior subset probabilities are given by a Dirichlet distribution. The probabilities of rank positions, therefore, can be inferred using the properties of the Dirichlet distribution. For three candidates a, b, c, the rank position probability that candidate a is ranked first, is given by $\boldsymbol{p}_{r_a=1} = \frac{\boldsymbol{p}_{\{a\}}}{\boldsymbol{p}_{\{a\}}+\boldsymbol{p}_{\{b\}}+\boldsymbol{p}_{\{c\}}}$, where each $\boldsymbol{p}_S$ has a Dirichlet distribution. As before, $\boldsymbol{p}_S$ denotes the probability that a randomly drawn member of the population approves of set S. Let N_S be the number of times the subset $S \subseteq \mathscr{C}$ occurs in a sample of subset choice data. Then the posterior probability of $\boldsymbol{p}_{r_a=1}$ is given by the following beta distribution

$$\boldsymbol{p}_{r_a=1} \sim f_\beta\left(N_{\{a\}} + \alpha_{\{a\}}, N_{\{b\}} + \alpha_{\{b\}} + N_{\{c\}} + \alpha_{\{c\}}\right). \tag{11}$$

Similarly, the posterior rank position probability that candidate a is ranked third, $\boldsymbol{p}_{r_a=3}$, has the following distribution:

$$\boldsymbol{p}_{r_a=3} \sim f_\beta\left(N_{\{b,c\}} + \alpha_{\{b,c\}}, N_{\{a,b\}} + \alpha_{\{a,b\}} + N_{\{b,c\}} + \alpha_{\{b,c\}} + N_{\{a,c\}} + \alpha_{\{a,c\}}\right), \tag{12}$$

and the posterior rank position probability that candidate a is ranked second is

$$\boldsymbol{p}_{r_a=2} = 1 - \boldsymbol{p}_{r_a=1} - \boldsymbol{p}_{r_a=3}. \tag{13}$$

Conditional on the observed subset choices, the random variables $\boldsymbol{p}_{r_a=1}$ and $\boldsymbol{p}_{r_a=3}$ are independent, and the variance of $\boldsymbol{p}_{r_a=2}$ equals the sum of the variances of $\boldsymbol{p}_{r_a=1}$ and $\boldsymbol{p}_{r_a=3}$.

Since any PVM score for candidate a is a linear combination of probabilities of the form $\boldsymbol{p}_{r_a=i},\ i=1,\ldots,m$, for any given voting vector and for any given observed subset data one can calculate the probability that candidate a is preferred to candidate b by a PVM with voting vector $(1-s,s,0),\ 0\le s\le\frac{1}{2}$ given by (7). We first state the result and then prove it.

Observation 5. *Consider a voting vector $\bar{w}=(1-s,s,0)$ for three candidates. For sufficiently large sample size N the posterior probability $P((a\succ_{\boldsymbol{p},\bar{w},SIM} b)|D_{av})$ that candidate a is preferred to candidate b in the population by the PVM with voting vector $\bar{w}$ under the assumption of the size-independent model, and given the observed AV data D_{av}, equals*

$$P\Big((a\succ_{\boldsymbol{p},\bar{w},SIM} b)\Big|D_{av}\Big)=F_N\left(\frac{\mu}{\sigma}\right),\ \text{where}$$

$$\mu=(1-2s)(t_{r_a=1}-t_{r_b=1})+s(t_{r_b=3}-t_{r_a=3}),$$

$$\sigma=\sqrt{\begin{array}{c}(1-2s)^2\ \frac{t_{r_a=1}+t_{r_b=1}-\left(t_{r_a=1}-t_{r_b=1}\right)^2}{N_1+\alpha_1}\\ +s^2\ \frac{t_{r_a=3}+t_{r_b=3}-\left(t_{r_a=3}-t_{r_b=3}\right)^2}{N_2+\alpha_2}\end{array}},$$

and $N_1=N_{\{a\}}+N_{\{b\}}+N_{\{c\}},\quad \alpha_1=\alpha_{\{a\}}+\alpha_{\{b\}}+\alpha_{\{c\}},\quad N_2=N_{\{a,b\}}+N_{\{b,c\}}+N_{\{a,c\}},\ \alpha_2=\alpha_{\{a,b\}}+\alpha_{\{b,c\}}+\alpha_{\{a,c\}},\ t_{r_a=1}=\frac{\alpha_{\{a\}}+N_{\{a\}}}{\alpha_1+N_1},\ t_{r_a=3}=\frac{\alpha_{\{b,c\}}+N_{\{b,c\}}}{\alpha_2+N_2}.$

Proof.

$$\begin{aligned}&P((a\succ_{\boldsymbol{p},\bar{w},SIM} b)|D_{av})\\ &\quad=P((1-s)\boldsymbol{p}_{r_a=1}+s\boldsymbol{p}_{r_a=2}-(1-s)\boldsymbol{p}_{r_b=1}-s\boldsymbol{p}_{r_b=2})>0)\\ &\quad=P((1-2s)\big(\boldsymbol{p}_{r_a=1}-\boldsymbol{p}_{r_b=1}\big)+s(\boldsymbol{p}_{r_b=3}-\boldsymbol{p}_{r_a=3})>0).\end{aligned}$$

Under the assumption of the size-independent model the random variables $\boldsymbol{p}_{r_a=1}$ and $\boldsymbol{p}_{r_b=1}$ have a Dirichlet distribution, so

$$E(\boldsymbol{p}_{r_a=1}-\boldsymbol{p}_{r_b=1})=t_{r_a=1}-t_{r_b=1}$$

and

$$Var(\boldsymbol{p}_{r_a=1}-\boldsymbol{p}_{r_b=1})=\frac{t_{r_a=1}+t_{r_b=1}-\left(t_{r_a=1}-t_{r_b=1}\right)^2}{N_1+\alpha_1}.$$

Similar formulae hold for $\boldsymbol{p}_{r_a=3}-\boldsymbol{p}_{r_b=3}$. Also, given the observed data and the size-independent model, the two random variables $(\boldsymbol{p}_{r_b=1}-\boldsymbol{p}_{r_a=1})$ and $(\boldsymbol{p}_{r_b=3}-\boldsymbol{p}_{r_a=3})$ are independent, so

$$Var\Big((1-2s)\big(\boldsymbol{p}_{r_a=1}-\boldsymbol{p}_{r_b=1}\big)+s\big(\boldsymbol{p}_{r_b=3}-\boldsymbol{p}_{r_a=3}\big)\Big)$$
$$=(1-2s)^2 Var\big(\boldsymbol{p}_{r_a=1}-\boldsymbol{p}_{r_b=1}\big)+s^2 Var\big(\boldsymbol{p}_{r_b=3}-\boldsymbol{p}_{r_a=3}\big). \quad \blacksquare$$

5 Data analysis

We analyze three AV elections in which the voters were also asked to provide a full ranking of the candidates and two AV elections where only subset choices were recorded. Table 1 presents the original ballot counts for all five elections (SSCW, TIMS President, TIMS Council, SJDM, MAA). For the first three of these elections, the reported (experimental) ranking data are also included in the table. Because we use these data to compare AV to scoring rules based on ranking data, we only consider those ballots that include both a subset choice and a full ranking.[9] Regenwetter and Grofman [23] have analyzed the full subset choice ballot counts of TIMS P (their TIMS E1) and TIMS C (their TIMS E2) without a comparison with rankings. Our reduced data set has the same AV outcomes as the full set.

The table can be read as follows. The first row means that, in the SSCW election, among 11 respondents who reported *abc* as their ranking, one person approved of no candidates, 10 approved of *a* alone, nobody approved of $\{a, b\}$, and nobody approved of all three candidates. The 7th row provides the ballot totals by subset size. For instance, 49 out of 52 voters approved of a single candidate only. For the SJDM and MAA elections, we do not have any ranking data. Here we only report the observed subset frequencies. Brams and Fishburn [2] and Saari [28], in this journal, analyze the same SSCW election.

Table 2 collects a wealth of information about the inferences we can draw from the AV ballot data via the size-independent model and directly from the ranking data: 1) Using the size-independent model, we infer from the AV ballot counts the rank position probability of each candidate for each rank, using equations (11)–(13). For instance, in the SSCW election, our mean estimate of the probability that candidate *a* is ranked first is 0.423 with a standard deviation of 0.068. 2) AV data for three candidates are consistent with the size-independent model if and only if each inferred rank position

[9] We also drop those ballots whose approved subset is ‘inconsistent’ with the reported ranking. A subset and a ranking are ‘inconsistent’ when the approval subset contains a candidate *a* without containing all candidates that are ranked better than *a* in the ranking. For SSCW, no rankings are inconsistent with the subset choices. However, out of 71 ballots, 19 are excluded from this analysis because they do not provide a full ranking. For TIMS P, of 1567 ballots, two are left out because they provide a ranking that is inconsistent with the approval subset and 156 are left out because they fail to provide a full ranking. For TIMS C, out of 1377 ballots, 6 are left out because the subset and the ranking are inconsistent and 218 ballots are left out because they do not have a full ranking of the candidates.

Table 1. Original data

Rankings\subset size	0	1	2	3	Total
	SSCW				
abc	1	10	0	0	11
acb	1	11	1	0	13
bac	0	0	0	0	0
bca	0	9	0	0	9
cab	0	11	0	0	11
cba	0	8	0	0	8
Total	2	49	1	0	52
	TIMS P				
abc	1	26	42	1	70
acb	3	31	24	6	64
bac	8	143	127	10	288
bca	8	141	169	13	331
cab	9	69	75	9	162
cba	28	174	262	30	494
Total	57	584	699	69	1,409
	TIMS C				
abc	6	85	43	26	160
acb	3	60	44	20	127
bac	7	102	43	23	175
bca	13	82	119	29	243
cab	6	69	73	21	169
cba	15	131	98	35	279
Total	50	529	420	154	1,153
	SJDM Total	MAA Total			
Subsets empty set	0	510			
$\{a\}$	22	413			
$\{b\}$	24	1,798			
$\{c\}$	9	1,019			
$\{a, b\}$	3	29			
$\{a, c\}$	4	52			
$\{b, c\}$	4	199			
$\{a, b, c\}$	1	20			
Total	67	4,040			

probability is well-defined, i.e., between zero and one [10]. The table shows that we can be reasonably confident that the rank position probabilities are indeed between zero and one. (For TIMS P, TIMS C and MAA a similar analysis, but without a Bayesian assessment of confidence levels, was already carried out by Regenwetter and Grofman, [23].) 3) We can also make inferences about the rank position probabilities directly from the ranking data in those three elections where the latter are available. Because this analysis is straightforward, we do not provide a formula for it. We find, for instance for SSCW, that the ranking based mean estimate that candidate a is ranked first is 0.455 with a standard deviation of 0.067. (Note that this result does not use

Table 2. Inferred rank position probabilities derived from approval voting data using the size-independent model and from ranking data (where available)

Rank position	Mean	Stdev	Mean	Stdev	p-value
	SSCW, SIM		SSCW Ranking		SSCW
a1	0.423	0.068	0.455	0.067	0.741
a2	0.327	0.205	0.218	0.055	0.609
a3	0.250	0.194	0.327	0.063	0.704
b1	0.192	0.054	0.182	0.052	0.888
b2	0.308	0.230	0.364	0.064	0.815
b3	0.500	0.224	0.455	0.067	0.846
c1	0.385	0.067	0.364	0.064	0.821
c2	0.365	0.205	0.418	0.066	0.806
c3	0.250	0.194	0.218	0.055	0.874
w	0.038	0.027			
	TIMS P, SIM		TIMS P, Ranking		TIMS P
a1	0.099	0.012	0.096	0.008	0.826
a2	0.286	0.022	0.319	0.012	0.185
a3	0.615	0.018	0.585	0.013	0.178
b1	0.486	0.021	0.439	0.013	0.058
b2	0.372	0.024	0.400	0.013	0.311
b3	0.142	0.013	0.161	0.010	0.264
c1	0.416	0.020	0.465	0.013	0.041
c2	0.342	0.026	0.280	0.012	0.031
c3	0.242	0.016	0.254	0.012	0.543
w	0.545	0.014			
	TIMS C, SIM		TIMS C, Ranking		TIMS C
a1	0.274	0.019	0.249	0.013	0.274
a2	0.210	0.031	0.298	0.013	0.009
a3	0.515	0.024	0.452	0.015	0.026
b1	0.348	0.021	0.362	0.014	0.556
b2	0.373	0.030	0.381	0.014	0.825
b3	0.279	0.022	0.257	0.013	0.383
c1	0.378	0.021	0.388	0.014	0.677
c2	0.417	0.029	0.321	0.014	0.003
c3	0.206	0.020	0.291	0.013	0.000
w	0.443	0.016			
	SJDM, SIM				
a1	0.397	0.064			
a2	0.246	0.139			
a3	0.357	0.124			
b1	0.431	0.064			
b2	0.212	0.140			
b3	0.357	0.124			
c1	0.172	0.049			
c2	0.542	0.127			
c3	0.286	0.117			
w	0.176	0.046			
	MAA, SIM				
a1	0.128	0.006			
a2	0.165	0.028			
a3	0.707	0.027			

Table 2. (continued)

Rank position	Mean	Stdev	Mean	Stdev	p-value
b1	0.556	0.009			
b2	0.256	0.025			
b3	0.187	0.023			
c1	0.315	0.008			
c2	0.578	0.020			
c3	0.106	0.018			
w	0.080	0.005			

the size-independent model.) 4) We also compare how different the rank position probability estimates are for the two methods. The table provides two-sided p-values for the difference between the rank position probability estimates. For instance, the p-value of 0.741 for rank position 1 of candidate a in SSCW means that the probability of finding a larger difference between these estimates, when in fact they are equal, is 0.741. In other words, these estimates are not statistically distinguishable. In contrast, at the bottom of the table, for election TIMS C and candidate c, the third rank position probability estimates are significantly different ($p < 0.001$). In general, some rank position probability estimates are significantly different, but many of them have high p-values. For the latter, there is high agreement between the rank position probability estimates derived from subset data (via the size-independent model) and from ranking data (without the size-independent model), respectively. 5) Using Observation 1, we derive w, the weight of rank position two under that PVM with voting vector $(1, w, 0)$ for which AV and PVM would mimic each other in a given data set under the size-independent model. For SSCW we find $w = 0.038$, i.e., AV essentially mimics plurality here. For SJDM the estimated $w = 0.176$ means that AV again roughly mimics plurality. For TIMS P and TIMS C we find w very close to 0.5, i.e., AV mimics Borda.

Table 3 provides a systematic comparison of outcomes by different voting rules in our inference framework. (None of the results in this table use the size-independent model.) For each pair of candidates x, y and every voting rule, we provide the inferred preference relation in the population according to that voting rule, and our confidence level. For instance, for a, b in SSCW we are 99.00% confident, based on the AV ballots, that a beats b by AV in the population from which the ballots were sampled. We are 99.55% (86%, 99.37%, 97.7%) confident, based on the ranking data, that a beats b by plurality (antiplurality, majority, Borda). The last column provides the inferred overall social welfare order in the population for each aggregation method, as well as a lower bound on our confidence for this inference. The formula for this bound is provided in the Appendix. Our main conclusions from this table are the following: 1) There are no dramatic differences between different aggregation methods (e.g., with high confidence they all well agree on the losers, even if they disagree on the winners). 2) In those cases

Table 3. Preference relations and confidence according to different voting procedures

Voting method	$\{a, b\}$		$\{b, c\}$		$\{a, c\}$		$\{a, b, c\}$	
	Preference	Confidence (%)	Preference	Confidence (%)	Preference	Confidence (%)	Preference	Confidence (%)
SSCW								
AV	$a > b$	99.00	$c > b$	97.86	$a > c$	62.24	$a > c > b$	59.10
Plurality	$a > b$	99.55	$c > b$	96.93	$a > c$	77.43	$a > c > b$	73.91
AntiPlurality	$a > b$	86.00	$c > b$	98.56	$c > a$	86.75	$c > a > b$	71.31
Majority	$a > b$	99.37	$c > b$	15.08	$c > a$	70.84	$c > a > b$	65.29
Borda	$a > b$	97.70	$c > b$	99.29	$c > a$	53.62	$c > a > b$	50.61
TIMS P								
AV	$b > a$	100.00	$b > c$	100.00	$c > a$	100.00	$b > c > a$	100.00
Plurality	$b > a$	100.00	$c > b$	84.99	$c > a$	100.00	$c > b > a$	84.99
Antiplurality	$b > a$	100.00	$b > c$	100.00	$c > a$	100.00	$b > c > a$	100.00
Majority	$b > a$	100.00	$c > b$	79.55	$c > a$	100.00	$c > b > a$	79.55
Borda	$b > a$	100.00	$b > c$	96.46	$c > a$	100.00	$b > c > a$	96.46
TIMS C								
AV	$b > a$	100.00	$c > b$	97.37	$c > a$	100.00	$c > b > a$	97.37
Plurality	$b > a$	100.00	$c > b$	84.59	$c > a$	100.00	$c > b > a$	84.59
Antiplurality	$b > a$	100.00	$b > c$	93.96	$c > a$	100.00	$b > c > a$	93.96
Majority	$b > a$	100.00	$b > c$	53.52	$c > a$	100.00	$b > c > a$	53.52
Borda	$b > a$	100.00	$b > c$	57.61	$c > a$	100.00	$b > c > a$	57.61

where the different methods disagree about the winners, the confidence in those inferences is not high. This means that there is hardly any statistical evidence that the various methods disagree for any of these data sets. 3) In those cases where the methods disagree, they furthermore support their own winner by only a small margin: An 'accidental' reversal among the top two candidates (a, c for SSCW, b, c for TIMS P and TIMS C) would represent a comparably less dramatic loss than would, say, the 'accidental' election of someone to whom a very large part of the electorate is strongly opposed.

Table 4 provides a similar comparison based only on the subset choice data assuming the size-independent model (but without majority rule, for lack of identifiability).[10] Here, the confidence levels are calculated based on Observation 5. For TIMS P and TIMS C, AV, plurality, and antiplurality generate the same social welfare order with high confidence (the confidence is low only for TIMS C plurality). As a consequence of Eq. (8), AV matches all PVMs in these two elections. For SSCW, AV and all PVMs agree on the winner and all but antiplurality agree on the entire social welfare order (because b and c are tied according to antiplurality). However, the confidence for any of these conclusions is low, consistent with Brams and Fishburn's [2] conclusion that this election was a toss-up. For SJDM, AV, plurality, and Borda fully agree, but the winner under these three is the same as the loser under antiplurality. However, these findings are supported with extremely low confidence. The most illustrative case is MAA, where AV and all positional voting methods agree on the loser with very high confidence, but plurality and antiplurality have different winners. From Eq. (8) we find that AV coincides with all PVMs specified by voting vector $(1, w, 0)$ with $w \leq 0.748$, and thus, in particular, with Borda ($w = 0.5$) and plurality ($w = 0$). The confidence depends on the value of w. It is extremely high for the Borda outcome.

A comparison of Tables 3 and 4 and the p-values in Table 2 reveals that the ranking based and the subset based analyses do not agree with each other. This could have several different causes. The format of the available data, and possibly the instructions to the respondents, may impact the inferences and the confidence assessments. For instance, in SSCW the respondents elected a society president by AV and were asked to indicate a hypothetical ranking that they would have provided, had the election been carried out under the Borda count. It is conceivable that they would have provided different rankings if they had been asked to provide the ranking underlying their approval vote, or the ranking they would have given if majority (or some other) rule had been used. A more thorough empirical comparison of different voting methods needs to disentangle such response biases.

Table 5 provides the same analysis as Table 3, but for a data set with 5 candidates where 2 candidates were to be elected to the TIMS council.

[10] Doignon and Regenwetter [10] have shown that, while the rank position probabilities are uniquely determined from the subset choice probabilities, the ranking probabilities are not uniquely determined.

Table 4. Preference relations and confidence according to different voting procedures reconstructed from approval voting data under the assumption of the size-independent model

Voting method	$\{a,b\}$		$\{b,c\}$		$\{a,c\}$		$\{a,b,c\}$	
	Preference	Confidence (%)	Preferences	Confidence (%)	Preferences	Confidence (%)	Preferences	Confidence (%)
SSCW								
AV	$a > b$	99.00	$c > b$	97.86	$a > c$	62.24	$a > c > b$	59.10
Plurality	$a > b$	97.91	$c > b$	96.16	$a > c$	62.11	$a > c > b$	56.18
Antiplurality	$a > b$	71.05	$c > b$	71.05	$a \sim c$	50.00	$a \sim c > b$	0.00
Borda	$a > b$	84.95	$c > b$	83.00	$a > c$	54.09	$a > c > b$	22.04
TIMS P								
AV	$b > a$	100.00	$b > c$	100.00	$c > a$	100.00	$b > c > a$	100.00
Plurality	$b > a$	100.00	$b > c$	96.23	$c > a$	100.00	$b > c > a$	96.23
Antiplurality	$b > a$	100.00	$b > c$	100.00	$c > a$	100.00	$b > c > a$	100.00
Borda	$b > a$	100.00	$b > c$	99.99	$c > a$	100.00	$b > c > a$	99.99
TIMS C								
AV	$b > a$	100.00	$c > b$	97.37	$c > a$	100.00	$c > b > a$	97.37
Plurality	$b > a$	98.36	$c > b$	79.21	$c > a$	99.83	$c > b > a$	77.40
Antiplurality	$b > a$	100.00	$c > b$	98.43	$c > a$	100.00	$c > b > a$	98.43
Borda	$b > a$	100.00	$c > b$	98.02	$c > a$	100.00	$c > b > a$	98.02
SJDM								
AV	$b > a$	60.61	$b > c$	98.62	$a > c$	97.34	$b > a > c$	56.58
Plurality	$b > a$	61.35	$b > c$	99.19	$a > c$	98.50	$b > a > c$	59.04
Antiplurality	$a \sim b$	50.00	$c > b$	63.01	$c > a$	63.01	$c > a \sim b$	0.00
Borda	$b > a$	55.37	$b > c$	78.18	$a > c$	73.89	$b > a > c$	7.44
MAA								
AV	$b > a$	100.00	$b > c$	100.00	$c > a$	100.00	$b > c > a$	100.00
Plurality	$b > a$	100.00	$b > c$	100.00	$c > a$	100.00	$b > c > a$	100.00
Antiplurality	$b > a$	100.00	$c > b$	99.37	$c > a$	100.00	$c > b > a$	99.37
Borda	$b > a$	100.00	$b > c$	100.00	$c > a$	100.00	$b > c > a$	100.00

Table 5. Preference relations and confidence according to different voting procedures for TIMS council data, two candidates (out of five) to be elected

Voting method	Preference	Confidence (%)	Preference	Confidence (%)	Preference	Confidence (%)	Preference	Confidence (%)	Preference	Confidence (%)	Preference	Confidence (%)
	$\{a,b\}$		$\{a,c\}$		$\{a,d\}$		$\{a,e\}$		$\{b,c\}$		$\{b,d\}$	
AV	$b > a$	100.00	$c > a$	87.01	$d > a$	73.54	$a > e$	100.00	$b > c$	100.00	$b > d$	100.00
Plurality	$b > a$	99.77	$a > c$	99.99	$a > d$	98.69	$a > e$	100.00	$b > c$	100.00	$b > d$	100.00
Antiplurality	$b > a$	100.00	$c > a$	100.00	$d > a$	100.00	$a > e$	92.30	$b > c$	100.00	$b > d$	100.00
Majority	$b > a$	100.00	$c > a$	68.53	$d > a$	84.07	$a > e$	100.00	$b > c$	100.00	$b > d$	100.00
Borda	$b > a$	100.00	$c > a$	70.77	$d > a$	80.73	$a > e$	100.00	$b > c$	100.00	$b > d$	100.00
	$\{b,e\}$		$\{c,d\}$		$\{c,e\}$		$\{d,e\}$		$\{a,b,c,d,e\}$			
AV	$b > e$	100.00	$c > d$	72.35	$c > e$	100.00	$d > e$	100.00	$b > c > d > a > e$	32.90		
Plurality	$b > e$	100.00	$d > c$	94.49	$c > e$	100.00	$d > e$	100.00	$b > a > d > c > e$	92.95		
Antiplurality	$b > e$	100.00	$c > d$	64.95	$c > e$	100.00	$d > e$	100.00	$b > c > d > a > e$	57.25		
Majority	$b > e$	100.00	$c > d$	75.04	$c > e$	100.00	$d > e$	100.00	$b > c > d > a > e$	27.64		
Borda	$b > e$	100.00	$d > c$	62.57	$c > e$	100.00	$d > e$	100.00	$b > d > c > a > e$	14.07		

(Again, none of the results in this table use the size-independent model.) These data were analyzed in Fishburn and Little [15]. The overall pattern of results is as follows: 1) The five aggregation methods agree with high confidence that e is the loser and that b is the winner. 2) All methods except plurality agree, also with high confidence, that a is the fourth ranked candidate. 3) Comparing AV, majority, and Borda we can see that majority rule and AV coincide, whereas Borda exchanges the second and the third candidates. Since two candidates were to be elected here, it actually matters who is ranked second. However, the disagreement between the methods regarding the second rank is supported only with low confidence.

6 Conclusions

Social choice theorists have spent much effort comparing voting methods and studying the behavior of various social choice procedures. Maybe the majority of that work compares voting procedures with respect to their strategy proneness, their likelihood to make voters vote sincerely, and their ability to satisfy certain normative criteria, such as, e.g., their Condorcet efficiency. Arguments have also been made that some social choice procedures are less expensive to implement than others. For instance, it has been pointed out [3] that existing voting machines, say, for presidential elections, could be used for subset based voting, but not for ranking based voting methods. Similarly, the computations involved in tallying are more costly for some methods than for others (e.g., the tally procedure for the single transferable vote is particularly computationally expensive). We argue that a comparative cost-benefit analysis of social choice procedures needs to include not just the price of using a given method, and the probabilities of all possible overlaps or discrepancies between methods, but also the likelihood of all possible erroneous election outcomes in elections where the process leading up to the certified tally counts is subject to probabilistic components. This has to be compounded with the cost of all possible errors. Furthermore, we argue that the comparison of social choice procedures based on thought experiments, probability calculations derived from theoretical distributions, and the listing of all possible election outcomes, needs to be supplemented by a rigorous empirical investigation of existing election ballots in order to empirically evaluate the extent to which each possible problem appears to occur in practice.

In this paper, we have made progress on several dimensions.

A) We have provided statistical inference tools to evaluate the possible latent outcomes under approval voting and all positional voting methods when the empirically observed subset choice or ranking tallies are assumed to result from a probabilistic process that follows a multinomial distribution. Only in those analyses where we inferred rankings from subsets did we further parametrize (constrain) this multinomial distribution by the size-independent model.

B) We have made new contributions to the understanding of the theoretical and empirical relationship between approval voting and positional voting methods. More specifically, we have shown a close theoretical link between AV and PVMs whenever the size-independent model holds. 1) AV, under the size-independent model, always mimics some positional voting method where the weights are determined by the subset size probabilities. Because the weights can take any values, the full theory of all possible PVMs [27] has practical and theoretical implications to the study of AV. 2) We point out the theoretically interesting result that, whenever each subset size is equally frequent/probable, it follows that AV mimics the Borda method in expectation (see Table 2 for an example). 3) The size-independent model is forced to hold whenever the voters follow Merrill's prescriptive rule to vote for approximately the "top half of the candidates independent of their preferences." As a consequence, for three candidates, when the voters follow the prescriptive rule, AV and the Borda method mimic each other.

C) We have extended the statistical inference framework of Sect. 2 to the analysis of subset and ranking data in the case where the multinomial distribution over subset choices is parameterized via the size-independent model.

D) We have applied our theoretical results to several data sets. These applications show that the calculated bounds are usually close together, thereby allowing us to get quite concise evaluations of the probabilities of all possible social welfare relations. Our statistical analysis fails to find any significant evidence that AV grossly disagrees with any of the standard PVMs (e.g., AV did not elect a Borda loser in any of these data sets). Even the small inconsistencies between voting methods, that we find in our data analysis, are by and large supported only with low statistical confidence.

Notice that a number of these conclusions hold with and without the use of the size-independent model.

We have made five types of restrictive assumptions in this paper. 1) As a first approximation, we have treated tallies as a random sample from a latent population. 2) In the Bayesian analysis, we have relied on conjugacy classes and diffuse priors. 3) In parts of the paper, we have relied on the size-independent model. 4) In all analyses that involved rankings, we implicitly assumed that individual preferences are complete linear orders. This idealization is standard in social choice theory. (The inference of approval voting outcomes from approval voting data did not involve rankings and thus did not assume linear order preferences.) 5) We have assumed throughout that the voters provided sincere ballots and rankings.

On the other hand, however, we have also eliminated one of the most restrictive assumptions commonly used in the social choice literature, namely that ballots are randomly sampled from an impartial (or similar) culture. Our ballots can originate from any distribution (culture).

Of course, as with any statistical analysis, our conclusions are subject to the assumptions of the theoretical statistical models that we use. Restrictive assumptions about how social choice theoretical quantities are related to empirical data are a two-edged sword: On the one hand, as we have demonstrated in another paper [25], social choice theoretical computations based on empirical data can dramatically vary with the implicit or explicit model used in the analysis/inference. On the other hand, as elsewhere in science, we wish to account for empirical data in a parsimonious fashion. We believe that future work, like our work here, should spell out as precisely as possible, what assumptions went into the analysis of empirical data, and should assess the statistical fit of a model to those data.

In addition to these general comments, we now briefly comment more specifically on the use of the size-independent model, the assumption that ballots are the outcome of a probabilistic process and the role of strategies.

Some of our analyses rely on the size-independent model of approval voting [12]. We have shown that some models used or proposed by other researchers are special cases of this model and we have pointed out that the size-independent model holds up quite well in statistical tests on empirical data. We have also included detailed analyses that did not rely on the size-independent model in cases where both subset choice and 'experimental' ranking data were available. One of the size-independent model's strengths lies in its ability to serve as a bridge between rankings and subset choices in situations where no ranking data are available.

One can argue[11] that voters who participate in elections of scientific societies are sophisticated and careful, and that the tally counts for such elections may not be safely viewed as resulting from a probabilistic process. Although we do not share this opinion we readily concede that this may place some limitations on the interpretation of our results for the data sets analyzed here. While opinions may differ on whether or not the voting behavior of scientists has probabilistic components, we believe that the sheer complexity and size of mass electorates alleviate such concerns for national elections. For mass electorates, we doubt that one can reasonably challenge our basic assumption that the certified ballot counts are a probabilistic account of the latent/intended ballot counts. One can however, naturally challenge the simple view that the ballots result from a multinomial (sampling) distribution. In line with a long tradition in mathematical psychology, we also strongly believe that one should model voter preferences (not just the ballot casting and tally process) as a probabilistic process. Future work will have to investigate to what extent the sampling processes assumed here should be replaced by different, possibly more sophisticated, probabilistic processes.

While much social choice literature is preoccupied with strategic components of voting processes, we have ignored such issues here. We leave it to

---

[11] Saari, personal communication.

future work to synthesize statistical inference with game theoretic considerations.

We close with a comment regarding Table 2 and the comparison of Tables 3 and 4. The rank position probabilities inferred from subsets via the size-independent model and those inferred from ranking data do not fully agree with each other. One could expect this phenomenon if the primitives of decision making are, in fact, cardinal utilities, involving a broad range of strengths of preference. In such a case, the strength of preference would be discarded in the empirical ranking data, but reflected by the empirical AV data. As a consequence, the rank position probabilities inferred from the AV data would not need to be consistent with the empirical ranking data.

Appendix
Bounds based on pairwise comparison probabilities

We provide a method of bounds similar to the one we used elsewhere for majority rule [30]. Suppose we have m candidates c_i, $i = 1, \ldots, m$, and for each pair of candidates c_i, c_j, $i \neq j$ we know the probability $P(c_i \succ_p c_j)$ that c_i is preferred to c_j according to a given tally procedure in the population. Denote by $Err(i,j) = 1 - P(c_i \succ_p c_j)$ the probability of making an inference error, i.e., the probability that c_i is in fact not preferred to c_j in the population. For instance, if c_i beats c_j by a tally procedure in the sample (denoted by $c_i \succ_s c_j$) then $P(c_i \succ_p c_j)$ denotes the probability that the preference relation between candidates c_i and c_j in the population is the same as in the sample.

One might be interested in 1) $P(A \succ_p B)$, the probability that all candidates in subset A are preferred to all candidates in subset B, 2) the probability that the full social welfare relation $c_1 \succ_p c_2 \ldots \succ_p c_m$ holds.

Observation 6. *The probability that all candidates in A are preferred to all candidates in B has the following upper and lower bounds:*

$$1 - \sum_{c_i \in A, c_j \in B} Err(i,j) \leq P(A \succ_p B) \leq 1 - max_{c_i \in A, c_j \in B} Err(i,j).$$

The probability $P(\succ_p)$ that the full social welfare order $c_1 \succ_p c_2 \ldots \succ_p c_m$ holds has the following upper and lower bounds:

$$1 - \sum_{1 \leq i < j \leq m} Err(i,j) \leq P(\succ_p) \leq 1 - max_{1 \leq i < j \leq m} Err(i,j).$$

So, if $max_{1 \leq i < j \leq m} Err(i,j)$ is small then both $P(A \succ_p B)$ and $P(\succ_p)$ are close to one, and our confidence that the social welfare order in the sample coincides with the social welfare in the population is high. Otherwise, even if only for one pair (i,j) the error probability $Err(i,j)$ is not small, then our confidence $P(\succ_p)$ is low. In that case, given $c_i \in A$ and $c_j \in B$, our confidence $P(A \succ_p B)$

is also low. Notice that if $A = \{a\}$ then we are looking at the probability to correctly choose the (single) top-ranked candidate.

As we see from Observation 6, the pairwise comparison probabilities play an important role in estimating our confidence in social welfare relations. This is the reason why our analysis in the paper is devoted to formulae for $Err(i, j)$.

References

[1] Brams SJ, Fishburn PC (1983) Approval voting. Birkhäuser, Boston

[2] Brams SJ, Fishburn PC (2001) A nail-biting election. Soc Choice Welfare 18: 409–414

[3] Brams SJ, Herschbach DR (2001) The science of elections. Science 292: 1449

[4] Chamberlin JR, Cohen JL, Coombs CH (1984) Social choice observed: Five presidential elections of the American Psychological Association. J Polit 46: 479–502

[5] Chamberlin JR, Featherston F (1986) Selecting a voting sytem. J Polit 48: 347–370

[6] De Groot MH (1970) Optimal statistical decisions. McGraw-Hill, New York

[7] Doignon J-P, Fiorini S (2003) The approval-voting polytope: Combinatorial interpretation of the facets. Math Inf Sci Hum 161: 29–39

[8] Doignon J-P, Fiorini S (2004) The facets and the symmetries of the approval-voting polytope. J Comb Theory B (forthcoming)

[9] Doignon J-P, Pekeč A, Regenwetter M (2004) The repeated insertion model for rankings: Missing link between two subset choice models. Psychometrika (forthcoming)

[10] Doignon J-P, Regenwetter M (1997) An approval-voting polytope for linear orders. J Math Psychol 41: 171–188

[11] Doignon J-P, Regenwetter M (2002) On the combinatorial structure of the approval-voting polytope. J Math Psychol 46: 554–563

[12] Falmagne J-Cl, Regenwetter M (1996) Random utility models for approval voting. J Math Psychol 40: 152–159

[13] Felsenthal DS, Maoz Z, Rapoport A (1986) Comparing voting systems in genuine elections: Approval voting vs selection-plurality. Soc Behav 1: 41–53

[14] Felsenthal DS, Maoz Z, Rapoport A (1993) An empirical evaluation of 6 voting procedures – do they really make any difference? Brit J Polit Sci 23: 1–27

[15] Fishburn PC, Little JDC (1988) An experiment in approval voting. Manag Sci 34: 555–568

[16] Gehrlein WV (1981) Borda efficiency of constant scoring rules with large electorates. RAIRO-Oper Res 15: 287–296

[17] Gehrlein WV, Lepelley D (1998) The Condorcet efficiency of approval voting and the probability of electing the Condorcet loser. J Math Econ 29: 271–283

[18] Gehrlein WV, Lepelley D (2001) The Condorcet efficiency of Borda rule with anonymous voters. Math Soc Sci 41: 39–50

[19] Marley AAJ (1993) Aggregation theorems and the combination of probabilistic rank orders. In: Critchlow DE, Fligner MA, Verducci JS (eds) Probability models and statistical analyses for ranking data. Springer, Berlin Heidelberg New York, pp 216–240

[20] Merrill S (1981) Strategic decisions under one-stage multi-candidate voting systems. Publ Choice 36: 115–134

[21] Regenwetter M (1997) Probabilistic preferences and topset voting. Math Soc Sci 34: 91–105

[22] Regenwetter M, Doignon J-P (1998) The choice probabilities of the latent-scale model satisfy the size-independent model when *n* is small. J Math Psychol 42: 102–106 (Theoretical Note)

[23] Regenwetter M, Grofman B (1998) Approval voting, Borda winners and Condorcet winners: Evidence from seven elections. Manag Sci 44: 520–533

[24] Regenwetter M, Grofman B (1998) Choosing subsets: A size-independent probabilistic model and the quest for a social welfare ordering. Soc Choice Welfare 15: 423–443

[25] Regenwetter M, Grofman B, Marley AAJ (2002) On the model dependence of majority preferences reconstructed from ballot or survey data. Math Soc Sci (Spec issue on random utility theory and probabilistic measurement theory) 43: 453–468

[26] Regenwetter M, Marley AAJ, Joe H (1998) Random utility threshold models of subset choice. Austr J Psychol (Spec issue on mathematical psychology) 50: 175–185

[27] Saari DG (1994) Geometry of voting. Springer, Berlin Heidelberg New York

[28] Saari DG (2001) Analyzing a nail-biting election. Soc Choice Welfare 18: 415–430

[29] Tabarrok A (2001) President Perot or fundamentals of voting theory illustrated with the 1992 election. Publ Choice 106: 275–297

[30] Tsetlin I, Regenwetter M (2003) On the probability of correct or incorrect majority preference relations. Soc Choice Welfare 20: 283–306

[31] Tsetlin I, Regenwetter M, Grofman B (2003) The impartial culture maximizes the probability of majority cycles. Soc Choice Welfare 21: 387–398