

Ranking-Based Voting Revisited: Maximum Entropy Approach Leads to Borda Count (and Its Versions)

Olga Kosheleva¹, Vladik Kreinovich¹, and Guo Wei²
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
olgak@utep.edu, vladik@utep.edu

²Department of Mathematics and Computer Science
University of North Carolina at Pembroke
PO Box 1510
Pembroke, NC 28372 USA
guo.wei@uncp.edu

Abstract

In many practical situations, we need to make a group decision that takes into account preferences of all the participants. Ideally, we should elicit, from each participant, a full information about his/her preferences, but such elicitation is usually too time-consuming to be practical. Instead, we only elicit, from each participant, his/her ranking of different alternatives. One of the semi-heuristic methods for decision making under such information is Borda count, when for each alternative and each participant, we count how many alternatives are worse, and then select the alternatives for which the sum of these numbers is the largest. In this paper, we explain the empirical success of the Borda count technique by showing that this method naturally follows from the maximum entropy approach – a natural approach to decision making under uncertainty.

1 Formulation of the Problem

Need for voting and group decision making. In many real-life situations, we need to make a decision that affects many people. Ideally, when making this decision, we should take into account the preferences of all the affected people. This group decision making situation is also known as *voting*.

What information can be used for voting: from the simplest majority voting to the most comprehensive situations. The simplest – and most

widely used – type of voting is when each person selects one of the possible alternatives. After this selection, all we know is how many people voted for each alternative.

Clearly, the more people vote for a certain alternative, the better is this alternative for the community as a whole. Thus, if this is all the information we have, and we do not plan to extract any additional information from the participants, then a natural idea is to select the alternative that gathered the largest number of votes. (Another idea is to keep only the alternatives with the largest number of votes and vote again.)

In this scheme, for each person, we only take into account one piece of information: which alternative is preferable to this person. To make more adequate decision, it is desirable to use more information about people’s preferences. An ideal case is when we use full information about people’s preferences; we will discuss this case in the following text. This is ideal but this requires too much elicitation and is, thus, not used in practice.

An intermediate stage – when we use more information than in the simple majority voting – is when we ask the participants to rank all the alternatives, and use these rankings to make a decision.

Ranking-based voting: a brief reminder. The famous result by a Nobelist Kenneth Arrow shows that it is not possible to have a ranking-based voting scheme that would satisfy all reasonable fairness-related properties [14, 16, 17]. So what can we do?

One of the schemes used in such voting is the *Borda count* (see, e.g., [16, 17], when for each participant i and for each alternative A_j , we count the number b_{ij} of alternatives that the i -th participant ranked lower than A_j . Then, for each alternative A_j , we add up the numbers corresponding to different participants, and we select the alternatives with the largest value of the corresponding sum $\sum_{i=1}^n b_{ij}$.

Why Borda count? Borda count is often successfully used in practice. However, the fact that there are several other alternative schemes prompts a natural question: why namely Borda count and why not one of these other schemes?

In this paper, we provide an explanation for the success of Borda count: namely, we show that the Borda count (and its versions) naturally follow from the maximum entropy approach – a known way for making decisions under uncertainty.

2 What If We Have Complete Information About the Preferences: Reminder

How to describe individual preferences. In order to describe what should we do when only know the rankings, let us first recall what decision we should make when we have full information about the preferences. To describe this, we need to recall how to describe these preferences.

In decision theory (see, e.g., [4, 8, 13, 14, 15]), a user's preferences are described by using the notion of *utility*. To define this notion, we need to select two extreme alternatives:

- a very bad alternative A_- which is worse than anything that we will actually encounter, and
- a very good alternative A_+ which is better than anything that we will actually encounter.

For each number p from the interval $[0, 1]$, we can then form a lottery $L(p)$ in which:

- we get A_+ with probability p and
- we get A_- with the remaining probability $1 - p$.

Then:

- For $p = 0$, the lottery $L(p)$ coincides with the very bad alternative A_- and is, thus, worse than any of the alternatives A that we encounter:

$$L(0) = A_- < A.$$

- For $p = 1$, the lottery $L(p)$ coincides with the very good alternative A_+ and is, thus, better than any of the alternatives A that we encounter:

$$A < L(1) = A_+.$$

Clearly, the larger p , the better the lottery. Thus, there exists a threshold p_0 such that:

- for $p < p_0$, we have $A(p) < A$, and
- for $p > p_0$, we have $A < A(p)$.

This threshold is known as the *utility* of the alternative A ; it is usually denoted by $u(A)$.

In particular, according to this definition:

- the very bad alternative A_- has utility 0, while
- the very good alternative A_+ has utility 1.

To fully describe people's preferences, we need to elicit, from each person i , this person's utility $u_i(A_j)$ of all possible alternatives A_j .

Utility is defined modulo linear transformations. The numerical value of utility depends on the selection of values A_- and A_+ .

One can show that if we use a different pair of alternatives (A'_-, A'_+) , then the resulting new utility values $u'(A)$ are related to the original values $u(A)$ by a linear dependence: $u'(A) = k + \ell \cdot u(A)$ for some k and $\ell > 0$.

Utility-based decision making under probabilistic uncertainty. In many practical situations, we do not know the exact consequences of different actions. For each action, we may have different consequences c_1, \dots, c_m , with different utilities $u(c_1), \dots, u(c_m)$. We can also usually estimate the probabilities p_1, \dots, p_m of different consequences. What is the utility of this action?

This action is equivalent to selecting c_i with probability p_i . By definition of utility, each consequence c_i is, its turn, equivalent to a lottery in which we get A_+ with probability $u(c_i)$ and A_- with the remaining probability $1 - u(c_i)$. Thus, the original action is equivalent to the corresponding two-stage lottery as a result of which we get either A_+ or A_- .

One can easily conclude that the probability of getting A_+ in this 2-stage lottery is equal to the sum $p_1 \cdot u(c_1) + \dots + p_m \cdot u(c_m)$. Thus, by definition of utility, this sum – which happens to be the expected value of utility – is the utility of the corresponding action.

How to make a group decision: simplest choice situation. Once we know the utility $u_i(A_j)$ of each alternative A_j for each participant i , we need to decide which alternative to select. Each alternative is thus characterized by the tuple of the corresponding utility values $(u_1(A_j), \dots, u_n(A_j))$. Based on the tuples corresponding to different alternatives, we need to select the best one. In other words, we need to be able, given two tuples $(u_1(A_j), \dots, u_n(A_j))$ and $(u_1(A_k), \dots, u_n(A_k))$, to decide which of the two alternatives is better, i.e., whether

$$(u_1(A_j), \dots, u_n(A_j)) < (u_1(A_k), \dots, u_n(A_k))$$

or

$$(u_1(A_k), \dots, u_n(A_k)) < (u_1(A_j), \dots, u_n(A_j)).$$

In the voting situation, there is usually a status quo state – a state that exists right now and that will remain if we do not make any decision. For example, if we are voting on different plans to decrease the traffic congestion in a city, the status quo situation is not to do anything and to continue suffering traffic delays. The status quo situation is worse than any of the alternatives. Thus, we can take this status quo situation as the value A_- . In this case, for all participants, the utility of the status quo situation is 0. The only remaining freedom is selecting A_+ . If we replace the original very good alternative A_+ with a new alternative A'_+ , then the corresponding linear transformation should transform 0 into 0 and thus, should have the form $u'_i(A) = \ell_i \cdot u_i(A)$.

In principle, each participant can select his/her own scale. It is reasonable to require that the resulting group choice should not change if one of the participants selects a different option A_+ . Thus, the corresponding order of the set of all the tuples must satisfy the condition that if $(u_1, \dots, u_n) < (u'_1, \dots, u'_n)$ then $(\ell_1 \cdot u_1, \dots, \ell_n \cdot u_n) < (\ell_1 \cdot u'_1, \dots, \ell_n \cdot u'_n)$. Other requirements include monotonicity (if an alternative is better for everyone it should be preferred) and fairness (the order should not change if we simply rename the participants.)

It turns out that the only order with this property is the comparison of the

products:

$$(u_1, \dots, u_n) < (u'_1, \dots, u'_n) \Leftrightarrow \prod_{i=1}^n u_i < \prod_{i=1}^n u'_i.$$

This comparison is known as *Nash's bargaining solution* after the Nobelist John Nash [5, 13, 14].

How to make a group decision: case of transferable utility. The above analysis refers to the case when we make a simple decision: e.g., when we simply elect an official. In many other group decision situations, however, the situation is more complicated. For example, some people may be opposed a road construction plan, since during this construction, their access to their homes and businesses will be limited. In such situations, if this particular alternative seems to be overall the best, a reasonable idea is to use some of its benefits to compensate those who will experience temporary inconveniences. The possibility of such a compensation is known as *transferable utility*: in contrast to the above simple choice situation, we can transfer utility from one participant to another.

The fact that we can move utility from one person to another means that now we have a common unit for such a utility; so, when some utility is transferred, the sum of all utilities remains constant. Suppose that without the transfers, the utilities corresponding to some alternative are u_1, \dots, u_n . The possibility of transfers means that we can have different values u'_1, \dots, u'_n – as long as the sum of all the utilities remains the same: $\sum_{i=1}^n u_i = \sum_{i=1}^n u'_i$.

The optimal transfer corresponds, as before, to the case when the product of the individual utilities attains the largest possible value. To find the resulting utility values, we need, given the values u_1, \dots, u_n , to find the values u'_1, \dots, u'_n for which the product $\prod_{i=1}^n u'_i$ attains the largest possible value among all the tuples for which $\sum_{i=1}^n u_i = \sum_{i=1}^n u'_i$. By applying the Lagrange multiplier method, we can reduce this constraint optimization problem to the unconstrained problem of optimizing the following objective function:

$$\prod_{i=1}^n u'_i + \lambda \cdot \left(\sum_{i=1}^n u_i - \sum_{i=1}^n u'_i \right).$$

Differentiating this expression with respect to each unknown u'_i and equating the derivative to 0, we conclude that $\prod_{i' \neq i} u'_{i'} - \lambda = 0$, i.e., $\prod_{i' \neq i} u'_{i'} = \lambda$. Thus, for each i , we have

$$u'_i = \frac{\prod_{i'=1}^n u'_{i'}}{\prod_{i' \neq i} u'_{i'}} = \frac{\prod_{i'=1}^n u'_{i'}}{\lambda}.$$

The right-hand side of this formula does not depend on i , thus we have

$$u'_1 = \dots = u'_n.$$

From the condition that $\sum_{i=1}^n u_i = \sum_{i=1}^n u'_i$, we conclude that

$$u'_1 = \dots = u'_n = \frac{1}{n} \cdot \sum_{i=1}^n u_i$$

and thus, that

$$\prod_{i=1}^n u'_i = \left(\frac{1}{n} \cdot \sum_{i=1}^n u_i \right)^n .$$

Among several alternatives, we should select the one for which this product is the largest – which is equivalent to selecting the alternative for which the sum $\sum_{i=1}^n u_i$ attains its largest possible value.

3 Let Us Apply These Formulas to Ranking-Based Voting

The problem of ranking-base voting: reminder. In the situation of ranking-based voting, we do not know the utilities. All we know, for each participant, is the ranking given by this participants to possible alternatives.

How can we apply the above formulas to this situation. Ranking

$$A_{i_1} < A_{i_2} < \dots$$

means that we can have different utility values $u(A_i) \in [0, 1]$ as long as these utility values are consistent with this ranking, i.e., as long as

$$u(A_{i_1}) < u(A_{i_2}) < \dots$$

In line with the above description of decision making under uncertainty, to find an actual utility of each alternative for this participant, we must find the expected value of the corresponding utility $u(A_j)$. To find this expected value, we need to select some probability distribution on the set of all possible tuples.

Maximum entropy approach: idea. There may be many different probability distributions on the set of all the property ordered tuples, we need to select one of them. Some of these distributions may have more uncertainty, some less. To select one of these distributions, a reasonable idea is to keep the original uncertainty and not to add artificial certainty – i.e., to select, among all possible distributions, a distribution with the largest possible value of uncertainty. A natural measure of this uncertainty is the entropy, so we select the distribution with the largest possible value of the entropy; see, e.g., [6].

What happens when we apply the maximum entropy approach. In our case, the largest possible entropy is attained for a uniform distribution on

the set of all the property sorted tuples. Thus, the utility of each alternative is equal to the expected value of the corresponding utility under such uniform distribution.

It is known (see, e.g., [1, 2, 3, 7, 9, 10, 11, 12]) that for k alternatives $A_{i_1} < A_{i_2} < \dots < A_{i_k}$, the resulting expected utility values $\bar{u}(A_j)$ take the form

$$\bar{u}_i(A_{i_q}) = \frac{q}{k+1}.$$

For each alternative $A_j = A_{i_q}$, its Borda count b_{ij} for this participant i – i.e., its number of worse-then- A_j alternatives – is equal to $b_{ij} = q - 1$. Thus, $q = b_{ij} + 1$, and in terms of this Borda count, the expected utility of each alternative A_j is equal to

$$\bar{u}_i(A_j) = \frac{b_{ij} + 1}{k + 1}.$$

In the case of transferable utility, this explains the Borda count. For transferable utility, as we have discussed in the previous section, we must select the alternative A_j for which the sum $\sum_{i=1}^n u_i(A_j)$ of the utilities is the largest possible. In our case, this means that we compare the values

$$\sum_{i=1}^n \bar{u}_i(A_j) = \sum_{i=1}^n \frac{b_{ij} + 1}{k + 1}.$$

This sum is, in its turn, equal to

$$\sum_{i=1}^n \frac{b_{ij} + 1}{k + 1} = \frac{1}{k + 1} \cdot \sum_{i=1}^n b_{ij} + \frac{n}{k + 1}.$$

Thus, the largest value of this sum corresponds to the largest value of the Borda sum $\sum_{i=1}^n b_{ij}$ – so we arrive exactly at the Borda count approach to voting.

Comment: in the simplest selection case, we get a version of the Borda count. It is worth mentioning that in the case of simple selection, we should use the following version of Borda count – select the alternative A_j for which the *product*

$$\prod_{i=1}^n \bar{u}_i(A_j) = \prod_{i=1}^n \frac{b_{ij} + 1}{k + 1}$$

attains the largest possible value. This, in its turn, is equivalent to maximizing the product $\prod_{i=1}^n (b_{ij} + 1)$, or, alternatively, to maximizing the sum $\sum_{i=1}^n \ln(b_{ij} + 1)$.

Acknowledgments

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science) and HRD-1242122 (Cyber-ShARE Center of Excellence).

References

- [1] M. Ahsanullah, V. B. Nevzorov, and M. Shakil, *An Introduction to Order Statistics*, Atlantis Press, Paris, 2013.
- [2] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A First Course in Order Statistics*, Society of Industrial and Applied Mathematics (SIAM), Philadelphia, Pennsylvania, 2008.
- [3] H. A. David and H. N. Nagaraja, *Order Statistics*, Wiley, New York, 2003.
- [4] P. C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.
- [5] A. Jaimes, C. Tweedie, T. Magoc, V. Kreinovich, and M. Ceberio, “Selecting the best location for a meteorological tower: a case study of multi-objective constraint optimization”, In: M. Ceberio and V. Kreinovich (eds.), *Constraint Programming and Decision Making*, Springer Verlag, Berlin, Heidelberg, 2014, pp. 61–66.
- [6] E. T. Jaynes and G. L. Bretthorst, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, UK, 2003.
- [7] O. Kosheleva, V. Kreinovich, J. Lorkowski, and M. Osegueda, “How to transform partial order between degrees into numerical values”, *Proceedings of the 2016 IEEE International Conferences on Systems, Man, and Cybernetics SMC’2016*, Budapest, Hungary, October 9–12, 2016.
- [8] V. Kreinovich, “Decision making under interval uncertainty (and beyond)”, In: P. Guo and W. Pedrycz (eds.), *Human-Centric Decision-Making Models for Social Sciences*, Springer Verlag, 2014, pp. 163–193.
- [9] J. Lorkowski and V. Kreinovich, “Interval and symmetry approaches to uncertainty – pioneered by Wiener – help explain seemingly irrational human behavior: a case study”, *Proceedings of the 2014 Annual Conference of the North American Fuzzy Information Processing Society NAFIPS’2014*, Boston, Massachusetts, June 24–26, 2014.
- [10] J. Lorkowski and V. Kreinovich, “Likert-type fuzzy uncertainty from a traditional decision making viewpoint: how symmetry helps explain human decision making (including seemingly irrational behavior)”, *Applied and Computational Mathematics*, 2014, Vol. 13, No. 3, pp. 275–298.

- [11] J. Lorkowski and V. Kreinovich, “Granularity helps explain seemingly irrational features of human decision making”, In: W. Pedrycz and S.-M. Chen (eds.), *Granular Computing and Decision-Making: Interactive and Iterative Approaches*, Springer Verlag, Cham, Switzerland, 2015, pp. 1–31.
- [12] J. Lorkowski and V. Kreinovich, “Fuzzy logic ideas can help in explaining Kahneman and Tversky’s empirical decision weights”, In: L. Zadeh et al. (Eds.), *Recent Developments and New Direction in Soft-Computing Foundations and Applications*, Springer Verlag, 2016, pp. 89–98.
- [13] R. D. Luce and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.
- [14] H. T. Nguyen, O. Kosheleva, and V. Kreinovich, “Decision making beyond Arrow’s ‘impossibility theorem’, with the analysis of effects of collusion and mutual attraction”, *International Journal of Intelligent Systems*, 2009, Vol. 24, No. 1, pp. 27–47.
- [15] H. Raiffa, *Decision Analysis*, McGraw-Hill, Columbus, Ohio, 1997.
- [16] D. G. Saari, *Chaotic Elections!*, American Mathematical Society, Providence, Rhode Island, 2001.
- [17] D. G. Saari, *Disposing Dictators, Demystifying Voting Paradoxes: Social Choice Analysis*, Cambridge University Press, New York, 2008.