CIRPÉE
Centre interuniversitaire sur le risque, les politiques économiques et l'emploi

Cahier de recherche/Working Paper **06-23**

Borda and the Maximum Likelihood Approach to Vote Aggregation

Michel Truchon

Juin/June 2006
Révisé octobre 2006/revised October 2006

Truchon: CIRPÉE and Département d'économique, Université Laval
mtru@ecn.ulaval.ca

I thank Mohamed Drissi-Bakhkhat, Stephen Gordon and Michel Le Breton for their comments on a previous version.

1 Introduction

The controversy that opposed Borda (1784) and Condorcet (1785) is well known; indeed, recent articles by Saari (2003, 2006) and Risse (2005)[1] suggest that it is still very much alive. The aggregation rules proposed by these two eighteenth-century scholars are very different in nature. The Borda rule is a scoring method that yields a unique ranking, although not necessarily strict. The Condorcet binary procedure may yield a cyclic binary (majority) relation, but when it does produce an unambiguous ranking - more precisely, an order - the Condorcet rule is a maximum likelihood estimator of the true order, under the assumptions that there is indeed a true order and that all judges or voters are able to order any two alternatives as they are in the true order with the same probability. Young (1988) extends the Condorcet result to the case where the majority relation is cyclic, by showing that the most likely orders are the Kemeny orders.

Condorcet made a distinction between the most likely ranking and the alternative most likely to be the best; so he was apparently aware that the alternative with the largest probability of being the best is not necessarily the top alternative in the most likely ranking. Young (1988) shows that indeed with a constant probability close to one half, the alternative most likely to be the best is the Borda winner, which may be different from the top alternative in the most likely ranking.

Drissi and Truchon (2004) extend the above analyses by relaxing the assumption that the probability of correctly ordering two alternatives is the same for all pairs of alternatives. They let this probability increase with the distance between the two alternatives in the true order, to reflect the intuition that a judge or voter is more prone to errors when confronted to two comparable alternatives than when confronted to a good alternative and a bad one. They make a thorough study of the case of three alternatives, using the class of logistic probability functions defined by

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

where k is the distance between two alternatives in the true order.

With $\{a,b,c\}$ as the set of alternatives, and restricting themselves to the polls for which abc is the unique Kemeny order, Drissi and Truchon show that abc is also the most likely

[1] See also the references within these articles.

order when α is sufficiently large with respect to β. Otherwise, the most likely order may be *acb* or *bac*. They also identify a subset of polls for which the Borda ranking turns out to be the most likely order for all values of α and β. However, they did not realize that this identity also holds for all polls when $\alpha = \beta$. The purpose of this note is to prove this result[2], which is interesting in itself since it puts the Borda rule on the same footing as the maximum likelihood rule, for a particular probability function. A parallel can also be drawn between this result and Young's result on the alternative most likely to be the best. Finally, in as much as the Condorcet procedure belongs to the maximum likelihood tradition (Condorcet approach was in fact one of the first application of the maximum likelihood approach), this result may be seen as a reconciliation of the Borda and Condorcet methods.

2 The social choice problem

Let $A = \{1, 2, \ldots, m\}$ be a set of *alternatives* or *candidates* to be ranked. Let us denote the subset of *complete weak orders* or *rankings* (reflexive and transitive binary relations) on A by $\mathcal{R}$ and the subset of *(linear) orders* (complete, transitive and asymmetric binary relations) on A by $\mathcal{L}$.[3] A complete weak order on A can be represented in a *rank form,* that is, by a vector $r = (r_1, r_2, r_3, \ldots)$, where r_1 is the rank of alternative 1, r_2 the rank of 2, and so on. A complete weak order on A can also be represented in a *sequence form,* that is, by a sequence $s_1 s_2 \ldots$, where s_1 and s_2 are the alternatives with ranks 1 and 2 respectively, etc. In this form, parentheses are used to identify alternatives with the same rank, as in $ab(cde)fg$.

To complete the problem, there is a set $I = \{1, 2, \ldots, n\}$ of *voters* or *judges.* Each is asked to compare the alternatives pair by pair, as in the Condorcet procedure. His or her vote is summarized in a matrix $X^i = [x^i_{st}]_{s,t \in A}$. For any pair of alternatives $(s, t) \in A^2$, $x^i_{st} = 1$ if voter i chooses s over t and $x^i_{st} = 0$ otherwise, and $x^i_{st} = 0$ if $s = t$. Alternatively, we can

[2] Actually, the result is proved for an extension of the maximum likelihood rule that consists in replacing multiple most likely orders by a single weak order in a sort of averaging process.

[3] I depart somewhat from Drissi and Truchon (2004) by introducing weak orders as well as linear orders. I also allow aggregation rules to produce weak orders instead of linear orders. This happens quite often with the Borda rule.

ask each voter i to fill in X^i according to the previous convention. Since only the aggregate information will be needed, we define a *poll* by:

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

Once voters or judges have expressed their opinions in a poll, the problem is to aggregate these opinions in order to select a final ranking. The Borda rule and the maximum likelihood rule are the two methods considered here to accomplish this aggregation.

The Borda rule is a scoring method with the vector of scores $(m-1, m-2, \ldots, 2, 1, 0)$. An alternative s receives $m-1$ points if it is ranked first by a voter, $m-2$ if it is ranked second, ..., and 0 points if it is last. These points are then aggregated across voters to give the *Borda score* $b_s(X)$ of s.[4] Alternatives are ordered according to these scores. Thus, the ranking $s_1 s_2 s_3 \ldots$ such that

$$b_{s_1}(X) \geq b_{s_2}(X) \geq b_{s_3}(X) \geq \cdots$$

is the *Borda ranking.* It is unique but it admits ties. Alternatively, given a poll X, the *Borda ranking* is the weak order $B(X)$ such that:

$$\forall s, t \in A : B_s(X) \leq B_t(X) \Leftrightarrow b_s(X) \geq b_t(X)$$

where $B_s(X)$ it the rank of alternative s. $B(\cdot)$ is the Borda rule.

The maximum likelihood approach to vote aggregation, which was initiated by Condorcet (1785), starts from the point of view that there exists a true order $r \in \mathcal{L}$ on the set of alternatives. The true order, however, is not known. The role of voters or judges is to provide an opinion as to what should be considered the true order. Their opinions are collected in $X^i,\ i = 1, \ldots, n$. The vote of expert i on a pair of alternatives (s, t) is a random variable $x_{st}^i \in \{0, 1\}$, conditional on the true order r. Condorcet assumed that the votes are independent between voters and pairs of alternatives. Moreover, each voter has the same probability, say $\bar{p} \in \left(\frac{1}{2}, 1\right)$, of ordering correctly two alternatives and this probability is the same for all pairs of alternatives. Now, given an order r and a poll X, let:

$$K(r; X) = \sum_{\substack{s,t \in A \\ r_s < r_t}} x_{st}^i$$

[4]Since $b_s(X) = \sum_{t=1}^{m} x_{st}$, it is legitimate to define the score of s as a function of the aggregate matrix X.

Young (1988) shows that a *most likely order* is an order r^* such that:

$$K(r^*; X) = \max_{r \in \mathcal{L}} K(r; X)$$

Such an order is also known in the literature as a *Kemeny order,* after the contribution of Kemeny (1959, 1962). The value of $K(r; X)$, called the *Kemeny score* of r given X, is the total number of agreements between r and the individual rankings making up profile X.

Drissi and Truchon (2004) maintain the assumption that votes are independent and that each voter has the same probability of ordering correctly two alternatives. They also assume that this probability is the same for any two couples of alternatives $(s, t), (u, v) \in A^2$ such that $r_s - r_t = r_u - r_v$. However, this probability is a non-decreasing function of the distance between the two alternatives in the true order. More precisely, they consider the class P^m of non-decreasing functions $p : \{1, \ldots, m-1\} \to \left(\frac{1}{2}, 1\right)$, where the domain is the set of the possible distances between alternatives in the true order. Thus, given a true order r and two alternatives $s, t \in A$ such that $r_s < r_t$, the probability that a judge orders correctly these two alternatives is given by $p(r_t - r_s)$.

Now, given a probability function $p \in P^m$, let

$$L_p(k) = \ln\left(\frac{p(k)}{1 - p(k)}\right) \ \forall k \in \{1, \ldots, m-1\} \tag{2}$$

and for every order r and every poll X, consider the function:

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

As shown by Drissi and Truchon (2004), the *most likely* or *most probable orders* are the elements of the set:

$$r_p(X) = \arg\max_{r \in \mathcal{L}} M_p(r; X)$$

Clearly, $r_p(\cdot)$ is another aggregation rule, the *maximum likelihood rule* with respect to the function p.

Now, if p is defined by (1), we have:

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

Thus, the function L_p defined in (2) takes the form:

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

3 The result

Before stating and proving the result of this paper, the possibility of multiple most likely orders must be addressed. One instance of this is when $r_p(X)$ contains two orders say r^1 and r^2 that are identical except that, in the sequence form, r^1 contains the subsequence $s\,t$ while r^2 contains the subsequence $t\,s$. Since $M_p(r^1;X) = M_p(r^2;X)$, it makes sense to declare s and t ex aequo, that is, to replace both r^1 and r^2 by a single weak order r in which $s\,t$ and $t\,s$ become $(s\,t)$. This means $r_s = r_t$ in the rank form.

More generally, suppose that, for some $\kappa \geq 2$ and $0 \leq k \leq m - \kappa$, the set $r_p(X)$ contains the order

$$s_1\, s_2 \ldots s_k s_{k+1} s_{k+2} \ldots s_{k+\kappa} s_{k+\kappa+1} \ldots s_m$$

as well as the orders obtained by permuting the κ alternatives in the subsequence

$$s_{k+1} s_{k+2} \ldots s_{k+\kappa} s_{k+\kappa+1},$$

for a total of $\kappa!$ orders. Then, across these $\kappa!$ orders, each alternative in $\{s_{k+1}, s_{k+2}, \ldots, s_{k+\kappa}\}$ occupies each rank from $k+1$ to $k+\kappa$ exactly $(\kappa-1)!$ times. Thus, it makes sense to "average" these $\kappa!$ orders, that is, to replace them by the single weak order

$$s_1\, s_2 \ldots s_k\, (s_{k+1} s_{k+2} \ldots s_{k+\kappa})\, s_{k+\kappa+1} \ldots s_m$$

in which the alternatives $s_{k+1}, s_{k+2}, \ldots, s_{k+\kappa}$ tie. In the rank form, one would have $r_{s_{k+1}} = r_{s_{k+2}} = \cdots = r_{s_{k+\kappa}}$. [5]

The result to be proved assumes that the above averaging process is applied in due circumstances. More precisely, let $\mathcal{N}$ be the set of all profiles and $r_p^w : \mathcal{N} \to \mathcal{R}$ be the extension of r_p obtained by replacing every maximal subset of orders in $r_p(X)$ that differ only by the permutation of adjacent alternatives, by a single weak order in which these alternatives tie. This extension is the object of the proposition that follows.

Proposition 1 *The extension r_p^w of the maximum likelihood rule r_p, with p defined by (1) and $\alpha = \beta$, coincides with the Borda rule.*

Proof. Without loss of generality, let $r = (1, 2, \ldots, m)$ be a ranking from $r_p^w(X)$ and let $\alpha = \beta = 1$. Then, $L_p(k) = k$ and

$$M_p(r;X) = \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} (v-u)\, x_{uv}$$

[5] For more on the multiplicity of most likely orders, see Truchon (2004).

Next, consider an alternative $s < m$ and another ranking $\hat{r}$ obtained from r by simply interchanging the position of s and $s+1$.[6] I shall show that $M_p(r;X) - M_p(\hat{r};X) = b_s(X) - b_{s+1}(X)$. Since $M_p(r;X) - M_p(\hat{r};X) \geq 0$, it will follow that $b_s(X) - b_{s+1}(X) \geq 0$ and, by transitivity of $\geq$, that $b_s(X) \geq b_t(X)$ $\forall s,t \in A : s < t$. Thus, r is the Borda ranking. In particular, if $M_p(r;X) - M_p(\hat{r};X) = b_s(X) - b_{s+1}(X) = 0$, then, by definition of r_p^w, we have $\hat{r} = r$, that is, alternatives s and $s+1$ tie in $\hat{r} = r$. This is consistent with the tieing in the Borda ranking.

For the remaining of the proof, let $t = s+1$. To ease reading, I shall use t instead of $s+1$ as a subscript. The only difference between $M_p(r;X)$ and $M_p(\hat{r};X)$ is that the terms x_{us} are replaced by x_{ut} for $u = 1, \ldots, m$. Although tedious, it is easy to check that:

$$\begin{aligned}
M_p(r;X) - M_p(\hat{r};X) &= \sum_{u=1}^{s-1} \left((s-u)(x_{us} - x_{ut}) + (t-u)(x_{ut} - x_{us}) \right) + (x_{st} - x_{ts}) \\
&\quad + \sum_{v=s+2}^{m} (v-s)(x_{sv} - x_{tv}) + \sum_{v=s+2}^{m} (v-s-1)(x_{tv} - x_{sv}) \\
&= \sum_{u=1}^{s-1} (x_{ut} - x_{us}) + (x_{st} - x_{ts}) + \sum_{v=s+2}^{m} (x_{sv} - x_{tv}) \\
&= \sum_{v=1}^{s-1} (x_{sv} - x_{tv}) + (x_{st} - x_{ts}) + \sum_{v=s+2}^{m} (x_{sv} - x_{tv}) \\
&= \sum_{v=1}^{m} x_{sv} - \sum_{v=1}^{m} x_{tv} \\
&= b_s(X) - b_t(X)
\end{aligned}$$

The second equality results from the simplification of the first term. The third equality follows from the fact that $x_{us} + x_{su} = m$. The fourth equality holds under the assumption $x_{ss} = x_{tt} = 0$. The last equality uses the well known, and easily checked, fact that $b_s(X) = \sum_{v=1}^{m} x_{sv}$. ■

A priori, nothing guarantees that $r_p^w(X)$ is a singleton. However, since the Borda ranking is unique, we get the single-valuedness of $r_p^w(X)$ as a corollary to the theorem.

Corollary 2 *Consider the extension r_p^w of the maximum likelihood rule r_p, with p defined by (1) and $\alpha = \beta$. Then, for any profile X, $r_p^w(X)$ is a singleton.*

[6] In this case, r and $\hat{r}$ can be seen as both the rank and the sequence forms of the respective orders.

References

Borda, J.C. de (1784): *Mémoire sur les élections au scrutin,* Histoire de L'Académie Royale des Sciences.

Condorcet, Marquis de (1785): *Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à Probabilité des Voix,* Paris: De l'Imprimerie Royale.

Drissi-Bakhkhat, M. and M. Truchon (2004), "Maximum Likelihood Approach to Vote Aggregation with Variable Probabilities," *Social Choice and Welfare*, **23**, 161-185.

Kemeny, J. (1959): "Mathematics without Numbers," *Daedalus*, **88**, 571-591.

Kemeny, J. and J.L. Snell (1962): *Mathematical Models in the Social Sciences,* Boston: Ginn.

Risse, M. (2005): "Why the count de Borda cannot beat the Marquis de Condorcet," *Social Choice and Welfare*, **25**, 95-113.

Saari, D.G. (2003): "Capturing the "Will of the People"," *Ethics,* **113**: 333–334.

Saari, D.G. (2006): "Which is better: the Condorcet or Borda winner?," *Social Choice and Welfare*, **26**, 107-129.

Truchon, M. (2004): "Aggregation of Rankings in Figure Skating," Cahier de recherche 0402, Département d'Économique, Université Laval.
www.ecn.ulaval.ca/pages/Recherche/cahiers.html

Young, H.P. (1988): "Condorcet's Theory of Voting," *American Political Science Review*, **82**: 1231-1244.