



Centre Interuniversitaire sur le Risque,  
les Politiques Économiques et l'Emploi

Cahier de recherche/Working Paper **06-25**

## **Statistical Comparison of Aggregation Rules for Votes**

Michel Truchon  
Stephen Gordon

Avril/April 2007

**Révisé**

---

Truchon: CIRPÉE and Département d'économique, Université Laval

[mtru@ecn.ulaval.ca](mailto:mtru@ecn.ulaval.ca)

Gordon: Corresponding author. CIRPÉE and Département d'économique, Université Laval

[sgor@ecn.ulaval.ca](mailto:sgor@ecn.ulaval.ca)

We thank Mohamed Drissi-Bakhkhat for his detailed comments on a previous version.

**Abstract:**

If individual voters observe the true ranking on a set of alternatives with error, then the social choice problem, that is, the problem of aggregating their observations, is one of statistical inference. This study develops a statistical methodology that can be used to evaluate the properties of a given voting or aggregation rule. These techniques are then applied to some well-know rules.

**Keywords:** vote aggregation, ranking rules, figure skating, maximum likelihood, optimal inference, Monte Carlo, Kemeny, Borda

**JEL Classification:** D71, C15

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	The social choice problem . . . . .	1
1.2	A Monte Carlo approach . . . . .	2
1.3	Summary of the results . . . . .	2
1.4	Contents . . . . .	3
<b>2</b>	<b>Methodology</b>	<b>3</b>
2.1	Vote aggregation: notation and definitions . . . . .	3
2.2	Statistical models of voter choice . . . . .	4
2.3	Criteria for evaluating aggregation rules . . . . .	6
2.4	Summary of the problem . . . . .	6
<b>3</b>	<b>Modelling choices</b>	<b>7</b>
3.1	Statistical model . . . . .	7
3.2	Aggregation rules . . . . .	7
3.2.1	The Borda rule . . . . .	8
3.2.2	The Kemeny Rule . . . . .	9
3.2.3	The Maximum Likelihood rule . . . . .	10
3.2.4	The ISU-94 Rule . . . . .	12
3.2.5	The ISU-98 Rule . . . . .	13
3.3	Loss functions . . . . .	15
3.4	Simulation methodology . . . . .	15
<b>4</b>	<b>The results in brief</b>	<b>18</b>
<b>5</b>	<b>Conclusion</b>	<b>20</b>
	<b>Appendices</b>	<b>24</b>
<b>A</b>	<b>Finding Kemeny orders and choosing among them</b>	<b>24</b>
<b>B</b>	<b>Description of the results</b>	<b>25</b>
B.1	Simulations with three judges and $(\alpha, \beta) = (1.1, 0.9)$ . . . . .	25
B.2	Simulations with nine judges and $(\alpha, \beta) = (1.1, 0.9)$ . . . . .	27
B.3	Simulations with nine judges and $(\alpha, \beta) = (1.0, 0.3)$ . . . . .	29
B.4	Simulations with nine judges and a constant probability . . . . .	31
B.5	Simulations with nine judges and $(\alpha, \beta) = (0.4, 1)$ . . . . .	33
	<b>References</b>	<b>35</b>

## List of Tables

1	Distribution function parameters and choice probabilities . . . . .	7
2	Illustration of the ISU-94 rule . . . . .	13
3	Illustration of the ISU-98 rule . . . . .	14
4	Loss function parameters and implications for costs of errors . . . . .	16
5	Description of the simulations . . . . .	18
6	Relation between different rules in terms of expected loss with 3 judges and $\eta = 1$ . . . . .	26
7	Probabilities and empirical frequencies of cycles with $\alpha = 1.1, \beta = 0.9$ . . . .	27
8	Relation between different rules in terms of expected loss, $\alpha = 1.1, \beta = 0.9, \eta = 1$ . . . . .	28
9	Probabilities and empirical frequencies of cycles with $\alpha = 1, \beta = 0.3$ . . . .	29
10	Relation between different rules in terms of expected loss, $\alpha = 1, \beta = 0.3, \eta \neq 2$ . . . . .	30
11	Relation between different rules in terms of expected loss, $\alpha = 1, \beta = 0.3, \eta = 2$ . . . . .	30

# 1 Introduction

## 1.1 The social choice problem

The question of how best to aggregate individual preferences or rankings is one of the oldest and best-known in the social sciences. Many organization must rank candidates for a position, a scholarship or a prize on the basis of rankings or preferences expressed by a panel of voters, judges or experts. Other organizations seek to rank investment projects on the basis of multiple criteria such as short term and long term profitability, social impact, environmental impact, and so on. In some democratic bodies or societies, the goal is the aggregation of individual preferences, expressed as orders or pre-orders, into a collective preference. In judged sports such as figure skating, diving, synchronized swimming and gymnastics, the task is the aggregation of individual assessments of the performance by a panel of judges into a final ranking.

All the above problems are formally the same. There is a set of candidates or alternatives to be ranked, a set of voters, judges or experts who are asked to provide individual rankings of these candidates. Or there are rankings that have been obtained by applying different criteria to the alternatives.<sup>1</sup> These rankings must then be aggregated according to some rule.

It is well known, at least since Condorcet (1785), that it may be difficult to aggregate individual rankings into a coherent order. Some two centuries later, Arrow (1963) crystallized this difficulty into his famous Impossibility Theorem, which demonstrates that there is no way of transforming individual preferences into a coherent social order that respects apparently innocuous regularity conditions. But as the abundant literature that his theorem has spurred testifies, Arrow's result does not imply that the social choice problem is any less relevant. It merely points to the difficulty of adopting a satisfactory aggregation rule.

In this study, we adopt the point of view of a decision maker, for instance the International Skating Union (ISU), who must adopt an aggregation rule that will be used over a long period. This decision maker has a measure of how bad an erroneous ranking may be, more precisely a loss function. Having a probabilistic model of the individual rankings, she can compute the *ex ante* expected loss, or the risk, of a given aggregation rule. She can choose a rule on this basis. We perform these computations for five aggregation rules: the Borda rule, the Kemeny rule, which can be seen as originating from the Condorcet maximum likelihood

---

<sup>1</sup>See for instance Arrow and Raynaud (1986).

approach to vote aggregation, a more general maximum likelihood (ML) rule and two rules that were used in figure skating prior to the Torino 2006 Olympic Games.

## 1.2 A Monte Carlo approach

Analytical computation of the risk for a given rule requires evaluating the probabilities and the associated losses of all possible outcomes. If the number of candidates is large - as is the case for the figure skating example that motivates this study - computing probabilities and losses for each element in the sample space is prohibitive. To circumvent this difficulty, we use Monte Carlo methods to estimate the risk for the high-dimension problems of interest in this study.

We generate a sample of profiles by having fictitious voters cast their votes according to a distribution function based on the binary approach of Condorcet as extended by Drissi and Truchon (2004). According to this approach there exists a true social order but judges observe it with error; they simply have a probability greater than one half of correctly ordering two alternatives in any pair. Drissi and Truchon (2004) let this probability increase with the distance between the two alternatives in the ‘true’ ranking, and propose a two-parameter logistic probability function to represent to competence of the judges. Since the binary approach does not guarantee that a given profile will be consistent, we discard draws that yield cycles.

In our analysis, we use the family of parametric loss functions proposed by Gordon and Truchon (2007). Given a profile of consistent rankings, we apply the five ranking rules retained for this study and we compute the associated losses for different values of the parameters of the loss function. Taking the sample mean yields an array of *ex ante* expected losses, which runs across the set of rules and the set of parameters of the loss function. This exercise is repeated for different values of the parameters of the probability function. These simulations are done for the case in which there are 3 and 9 judges, and where the number of alternatives ranges from 3 to 10 (sometimes 12 and 20). For computational reasons, this number is kept below 7 (sometimes 8) with the ML rule.

## 1.3 Summary of the results

Our results suggest that it is difficult to conclude that one rule performs better in all circumstances. It is generally the case that if the parameters of the probability function are known with certainty, then the ML rule does best. In practice, of course, this rule is not feasible,

since the form of the probability distribution is not known in advance. On the other hand, it does appear that if the ML rule is applied using parameter combinations that are ‘close’ to the true values, then this ‘quasi-ML’ rule still does well when compared to other rules. As for the Borda rule, it does well with increasing probabilities but poorly with constant probabilities.

## 1.4 Contents

The paper is structured as follows. Section 2 presents the aggregation problem, the statistical model for generating votes, and formalises the criteria for evaluating aggregation rules. Section 3 describes the modelling choices, including those of the parameters of the probability function and those of the loss function. It also summarizes the social choice rules that are evaluated in this study. Section 4 discusses the results and Section 5 concludes. A more detailed description of the results is relegated to an appendix.

# 2 Methodology

## 2.1 Vote aggregation: notation and definitions

Let  $A = \{1, 2, \dots, m\}$  be a set of *alternatives* or *candidates* to be ranked. We denote by  $\mathcal{B}$  the set of *binary relations* on  $A$ , by  $\mathcal{B}^*$  be the set of complete and asymmetric binary relations on  $A$ , by  $\mathcal{R}$  the subset of *complete weak orders* or *rankings* (reflexive and transitive binary relations) on  $A$  and by  $\mathcal{L}$  the subset of *linear orders* (complete, transitive and asymmetric binary relations) on  $A$ . Note that  $\mathcal{L} \subset \mathcal{R} \subset \mathcal{B}$  and  $\mathcal{L} \subset \mathcal{B}^* \subset \mathcal{B}$ .<sup>2</sup> A complete weak order on  $A$  can be represented by a vector  $r = (r_1, r_2, r_3, \dots)$  or  $x = (x_1, x_2, x_3, \dots)$  where  $r_1$  and  $x_1$  are the rank of alternative 1,  $r_2$  and  $x_2$  the rank of 2, and so on.<sup>3</sup>

Let there be a set  $J = \{1, 2, \dots, n\}$  of *voters* or *judges*. For each judge  $j$ , we generate a linear order  $x^j \in \mathcal{L}$ , also called a *vote*, on the set  $A$ . Equivalently, a vote  $x^j$  can be represented by an  $(m \times m)$  binary matrix  $X^j = [x_{st}^j]_{s,t \in A}$  where:

$$x_{st}^j = \begin{cases} 1 & \text{if } x_s^j < x_t^j \\ 0 & \text{otherwise} \end{cases}$$

---

<sup>2</sup>With  $m$  alternatives, the cardinality of  $\mathcal{B}^*$  and  $\mathcal{L}$  are  $2^{\frac{m(m-1)}{2}}$  and  $m!$  respectively. The difference between the two is the number of cyclic binary relations in  $\mathcal{B}^*$ .

<sup>3</sup>Throughout, we shall use  $r$  to represent the true order on  $A$  and  $x$  to represent a vote on  $A$ .

Conversely, given a binary representation  $X^j$  of a linear order, we get the representation  $x^j$  by setting  $x_s^j = m - \sum_{t=1}^m x_{st}^j$ .<sup>4</sup> We shall use the two representations interchangeably.

A *profile of votes* is an array  $X = (x^1, \dots, x^n) \in \mathcal{L}^n$ . A profile may also be written in the binary form  $X = (X^1, \dots, X^n)$ . Once the voters or judges have cast their votes, the problem is to aggregate these votes into a final ranking. We formalize this idea in the following definition.

**Definition 1** An *aggregation or ranking rule* is a function  $\Gamma : \mathcal{L}^n \rightarrow \mathcal{R}$  that assigns to each profile  $X$  a final ranking  $\Gamma(X)$  of the alternatives.  $\Gamma_s(X)$  represents the rank of alternative  $s$  in the final ranking  $\Gamma(X)$ .

## 2.2 Statistical models of voter choice

As in Gordon and Truchon (2007), this study adopts the Condorcet approach to the social choice problem in supposing that the true social order exists, but that judges observe it with error. As in Drissi and Truchon (2004), we let the probability that a judge orders correctly two alternatives increase with the distance between the two alternatives in the allegedly true ranking. This incorporates the possibility that voters or judges will have a better chance to rank correctly two alternatives when one is very good and the other very bad than when facing two similar alternatives.

Formally, there exists a non-decreasing function  $p : \{1, \dots, m-1\} \rightarrow (\frac{1}{2}, 1)$  such that if  $r \in \mathcal{L}$  is the true social order on  $A$ , then the vote of judge  $j$  on a pair of alternatives  $(s, t)$  is a random variable  $\tilde{x}_{st}^j \in \{0, 1\}$  with marginal distribution:

$$\begin{aligned} \Pr(\tilde{x}_{st}^j = 1 \mid r) &= \begin{cases} p(r_t - r_s) & \text{if } r_s < r_t \\ 1 - p(r_s - r_t) & \text{if } r_s > r_t \end{cases} & (1) \\ \Pr(\tilde{x}_{st}^j = 0 \mid r) &= 1 - \Pr(\tilde{x}_{st}^j = 1 \mid r) \end{aligned}$$

Put simply,  $p(r_t - r_s)$  is the probability that a judge orders alternatives  $s$  before  $t$  when  $s$  has a higher rank than  $t$  in the true order, that is, when  $r_s < r_t$ .

---

<sup>4</sup>There is an abuse of notation in using  $x_s$  to represent elements of a vector and  $x_{st}$  to represent elements of a matrix but, given the one-to-one correspondence between the  $x^j$  and the  $X^j$ , this should entail no confusion. Naming both judges and alternatives as 1, 2, 3, ... is also an abuse of notation but this allows for simpler notations in the remaining of the paper.



If the votes  $X^j$  are taken from  $\mathcal{B}^*$ , that is, if cycles are permitted but not ties, and if the elements of the binary matrix  $X^j$  are independently distributed, then the distribution of the whole matrix  $X^j$  is:

$$\widehat{\Pr}(X^j | r, \mathcal{B}^*) = \prod_{\substack{s,t \in A \\ r_s < r_t}} \Pr(\tilde{x}_{st}^j = x_{st}^j | r) \quad (2)$$

If we denote by  $\Pr(\mathcal{L}|r)$  the probability of generating an order from (2), then:

$$\Pr(\mathcal{L}|r) = \sum_{X^j \in \mathcal{L}} \widehat{\Pr}(X^j | r, \mathcal{B}^*)$$

The probability of drawing an order from  $\mathcal{L}$  is therefore:

$$\Pr(X^j | r, \mathcal{L}) = \begin{cases} \frac{\Pr(X^j | r, \mathcal{B}^*)}{\Pr(\mathcal{L}|r)} & \text{if } X^j \in \mathcal{L} \\ 0 & \text{if } X^j \in \mathcal{B}^* \setminus \mathcal{L} \end{cases} \quad (3)$$

Restricting the sample space to  $\mathcal{L}$ , that is, imposing transitivity, implies that the elements of  $X^j$  are no longer independently distributed, even though the joint probabilities are still based on the product of the individual marginal probabilities. Consider the case where  $m = 3$  and where we know that a judge has ranked alternative  $s$  ahead of  $t$  and  $t$  ahead of  $u$ . Even if the *marginal* probability of ranking  $u$  ahead of  $s$  is some  $0 < \hat{p}_{su} < 1$ , transitivity implies that the conditional probability that the judge will rank  $s$  in front of  $u$  is equal to one.

A more satisfactory approach might have consisted in specifying a probabilistic model on  $\mathcal{L}$  directly. Given the number of alternatives that we shall consider, this would have represented a daunting task. The binary approach, which is often used in the literature on social choice, together with the elimination of incoherent relations, is adopted for practical reasons.<sup>5</sup> As we shall see in subsection 3.1, this choice allows us to specify a parsimonious parametric model.

Finally, we suppose that judges produce their orders independently according to (1-3), so that the probability of a given profile  $X$  is:

$$f(X|r) = \prod_{j=1}^n \Pr(X^j | r, \mathcal{L}) \quad (4)$$

---

<sup>5</sup>On this point, see also Remark 2 in Drissi and Truchon (2004).

## 2.3 Criteria for evaluating aggregation rules

An aggregation or ranking rule is usually chosen in advance and is often used over a long period. A question of interest is thus how different ranking rules could fare in repeated applications. In other words, what is the *ex ante* expected loss or risk of a ranking rule? This is the question addressed in this paper. As in Gordon and Truchon (2007), we begin by making explicit the concepts of loss and risk.

**Definition 2** A *loss function* is a mapping  $d : \mathcal{R}^2 \rightarrow \mathbb{R}_+$  such that  $d(r, \hat{r}) > 0 \forall \hat{r} \neq r$  and  $d(r, r) = 0 \forall r$ . The value  $d(r, \hat{r})$  is the loss resulting from the selection of the ranking  $\hat{r}$  when  $r$  is the true order.

**Definition 3** Given a true order  $r$ , a distribution function  $f(\cdot | r)$  on profiles, and a loss function  $d$ , the *risk* or the *ex ante expected loss* of an aggregation rule  $\Gamma$  is given by:

$$\rho(\Gamma; f, d, r) = \sum_{X \in \mathcal{L}^n} f(X | r) d(r, \Gamma(X)) \quad (5)$$

Suppose that a set of aggregation rules  $\mathcal{G} \equiv \{\Gamma^1, \Gamma^2, \dots, \Gamma^G\}$  is available to the decision maker. For a given combination  $(f, d, r, \mathcal{G})$ , an *optimal social choice rule* is a solution  $\Gamma^*$  to the problem:

$$\min_{\Gamma \in \mathcal{G}} \rho(\Gamma; f, d, r) \quad (6)$$

In this paper, we are concerned not only with the optimal rule but with the relative performance of all members of  $\mathcal{G}$  as we vary the parameters of the model.

## 2.4 Summary of the problem

If a decision maker adopts the Condorcet approach of positing the existence of a true social order that is observed with error, and if she is obliged to decide upon a fixed aggregation rule  $\Gamma$  before the votes are observed, then she must specify:

- i) a statistical model for the votes,
- ii) a set of aggregation rules,
- iii) a family of loss functions.

Since there is little reason to believe that a single aggregation rule will be optimal in all settings, care must be taken to ensure that the assumptions are plausible in the setting under study.

### 3 Modelling choices

Although the techniques described here can be applied to any social choice problem, an exhaustive search over all combinations of statistical models and loss functions is beyond the scope of any single study. The context that we have in mind is the ranking of competitors in figure skating; see Truchon (2004), and Gordon and Truchon (2007). It should be noted that assumptions that we believe are plausible in the context of figure skating may not be appropriate in other applications.

#### 3.1 Statistical model

To generate the votes, we adopt the two-parameter logistic probability function used by Drissi and Truchon (2004) and Gordon and Truchon (2007). That is, we let the term  $p(r_t - r_s)$  in (1), which is the probability that a judge  $j$  ranks  $s$  before  $t$  when  $r_s < r_t$ , take the form

$$p_{st}^{\alpha\beta}(r) = \frac{\exp(\alpha + \beta(r_t - r_s - 1))}{1 + \exp(\alpha + \beta(r_t - r_s - 1))} \tag{7}$$

with  $\alpha > 0$  and  $\beta \geq 0$ . If  $\beta > 0$ , this probability increases with the distance between  $s$  and  $t$  in the true order, measured by  $r_t - r_s$ . The larger  $\beta$ , the higher this rate. With  $\beta = 0$ , the probability is constant. And the larger  $\alpha$ , the larger the probability. Thus, the distribution function defined by (4) takes a parametric form, say  $f^{\alpha\beta}(X|r)$ . The values for  $\alpha$  and  $\beta$  used in this study, as well as their implications for the choice probabilities, are listed in Table 1.

Parameters		Distance between alternatives										
$\alpha$	$\beta$	1	2	3	4	5	6	7	8	9	10	11
0.4	1	0.60	0.80	0.92	0.97	0.99	1.00	1.00	1.00	1.00	1.00	1.00
0.4	0	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60
1.1	0	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
2.2	0	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90
1.0	0.3	0.73	0.79	0.83	0.87	0.90	0.92	0.94	0.96	0.97	0.98	0.98
1.1	0.9	0.75	0.88	0.95	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00

Table 1: Distribution function parameters and choice probabilities

#### 3.2 Aggregation rules

We apply four or five (depending on the number of alternatives) aggregation rules to the votes generated according to the above statistical model. The first is the well-known Borda

rule, which was in use in the *Académie des sciences de France* from 1770 until Napoléon’s election to this institution. The second rule is due to Kemeny (1959). It yields orders that have a minimal total number of disagreements with the orders in the profile. Moreover, as shown by Young (1988), the Kemeny rule may be seen as the result of the Condorcet maximum likelihood approach with a constant probability of correctly ordering two alternatives. Drissi and Truchon (2004) extends this approach by having this probability increase with the distance between the two alternatives in the allegedly true ranking. They propose the probability function defined by (7) to represent the competence of the judges. This yields a more general maximum likelihood rule. This is the third rule used in this study when the number of alternatives is not too large.

The final two rules are less well known, and are taken from figure skating. Most fans know that skaters are ranked from the scores that they receive from a panel of judges. However, it is less well known that, prior to the Torino 2006 Olympic Games, the scores of each judge  $j$ , a real number between 0 and 6, merely served to establish a ranking  $r^j$  (ordinals) for this judge. Then, an aggregation or ranking rule, defined by the International Skating Union (ISU), was applied to the profile thus obtained to arrive at a final ranking. There have been at least two such rules. The first one, which we label ISU-94, was used from at least 1982 until 1998. It was replaced in 1998 by a rule that we label ISU-98. It was in use at the Salt Lake City 2002 Olympic Games, but was replaced prior the Torino 2006 Games.<sup>6</sup> The ISU-94 and ISU-98 rules remain nevertheless very interesting, because, *inter alia*, of the controversies that arose from their use. The Borda rule is a component of these two rules. A variant of the Copeland (1951) rule, which is also often discussed in the literature on aggregation of votes, is a component of the ISU-98 rule.<sup>7</sup>

### 3.2.1 The Borda rule

Given a profile  $X$  and alternative  $s$ , let  $b^*(s) = \sum_{j=1}^n (m - x_s^j)$ . This number is the *Borda score* of  $s$ . Given the correspondence between  $x^j$  and  $X^j$ , we also have  $(m - x_s^j) = \sum_{t=1}^m x_{st}^j$  so that  $b^*(s) = \sum_{j=1}^n \sum_{t=1}^m x_{st}^j$ . Let also  $b(s) = \sum_{j=1}^n x_s^j$ . Clearly,  $b^*(s) = mn - b(s)$ . The Borda rule consists in ordering alternatives according to the  $b(s)$  or inversely to the  $b^*(s)$ .<sup>8</sup>

---

<sup>6</sup>The scoring method itself is different and the aggregation rule is now cardinal, more precisely additive, rather than ordinal.

<sup>7</sup>More on the two ISU rules can be found in Truchon (2004). The ISU-94 rule has also been studied extensively by Bassett and Persky (1994).

<sup>8</sup>Formally, the numbers  $b(s)$  and  $b^*(s)$  depend on  $X$ . We drop this argument for the sake of simplicity.

**Definition 4** The *Borda rule* is the function  $B : \mathcal{L}^n \rightarrow \mathcal{R}$  defined by:

$$\forall s, t \in A : B_s(X) \leq B_t(X) \Leftrightarrow b(s) \leq b(t) \Leftrightarrow b^*(s) \geq b^*(t)$$

$B(X)$  is the *Borda ranking*.

### 3.2.2 The Kemeny Rule

The definition of the rule involves the Kemeny “distance” between a linear order and a profile. First, let  $\gamma_{st} : \mathcal{L}^2 \rightarrow \mathbb{R}$ , be a function defined for every couple of alternatives  $(s, t)$  and every pair of linear orders  $(r, x)$  by:

$$\gamma_{st}(r, x) = \begin{cases} 1 & \text{if } r_s < r_t \text{ and } x_s > x_t \\ 0 & \text{otherwise} \end{cases}$$

Then, the *Kemeny metric* on  $\mathcal{L}$  is the function  $d^K : \mathcal{L}^2 \rightarrow \mathbb{R}$  defined by:<sup>9</sup>

$$d^K(r, x) = \sum_{s \in A} \sum_{t \in A} \gamma_{st}(r, x)$$

Finally, the Kemeny “distance”  $\delta^K$  between a linear order  $x$  and a profile  $X$  is given by:

$$\delta^K(r, X) = \sum_{j=1}^n d^K(r, x^j)$$

Kemeny (1959) points out that, given a profile  $X$ , there are two methods of finding a “consensus” ranking. They consist in minimizing  $\sum_{j=1}^n d^K(r, x^j)$  and  $\sum_{j=1}^n d^K(r, x^j)^2$  respectively, with respect to  $r$ . Kemeny and Snell (1962) call the solutions to the two minimization problem *median rankings* and *mean rankings* respectively.<sup>10</sup> Nowadays, the name of Kemeny is associated with the first problem. This is the Kemeny rule.

**Definition 5** The *Kemeny rule* is the correspondence<sup>11</sup>  $K : \mathcal{L}^n \rightarrow \mathcal{L}$  that assigns to each profile  $X$ , the subset  $K(X) = \arg \min_{r \in \mathcal{L}} \delta^K(r, X)$ . The elements of  $K(X)$  are *Kemeny orders*.

The value of  $\delta^K(r, X)$  is the total number of disagreements between  $r$  and the rankings in profile  $X$ . Thus, a Kemeny order has a minimal total number of disagreements with  $X$ . See the Appendix for the computation of Kemeny orders and the handling of their multiplicity.

---

<sup>9</sup>We restrict the definition to the space  $\mathcal{L}$  since we work with profiles of linear orders and since the search with the Kemeny rule is done within this space.

<sup>10</sup>To be more precise, given a profile  $X = (x^1, \dots, x^n)$ , they define them as the median and the mean of the set  $\{x^1, \dots, x^n\}$ .

<sup>11</sup>Contrary to the rules previously defined, we now have a correspondence instead of a single value function.

### 3.2.3 The Maximum Likelihood rule

The Maximum Likelihood approach to vote aggregation dates back to Condorcet (1785), who vehemently opposed Borda’s rule, indeed all scoring rules. He pointed out their major flaw: they can top-rank a candidate that can be defeated by each of the other candidates in pairwise votes or comparisons. As an alternative, Condorcet proposed that candidates be ranked on the basis of these pairwise votes. His objective was to justify the majority principle. On this, he was certainly inspired by Rousseau (1762) in his *Social Contract*, for whom the opinion of the majority is legitimate because it expresses the “general will.”

When in the popular assembly a law is proposed, what the people is asked is not exactly whether it approves or rejects the proposal, but whether it is in conformity with the general will, which is their will. Each man, in giving his vote, states his opinion on that point; and the general will is found by counting votes. When therefore the opinion that is contrary to my own prevails, this proves neither more nor less that I was mistaken and that what I thought to be the general will was not so. [Rousseau (1913), p.93]

Condorcet’s objective was to formulate this proposition rigorously, using the calculus of probability, which was new at the time of his writing. His approach may be cast as follows. There is a true or objective order on the set of candidates. Judges may have different opinions because they are imperfect observers. However, if they are right more often than they are wrong, then the opinion of the majority yields a most probable order. As the number of judges increases, the opinion of the majority converges to the true order.<sup>12</sup>

Condorcet was perfectly aware that the binary relation resulting from his procedure may contain cycles, a phenomenon sometimes referred to as the *Condorcet paradox*. He proposed a method for breaking these cycles but unfortunately this method gives consistent results only for the case of three alternatives. Young (1988) shows that a correct application of the maximum likelihood principle leads to the selection of the *Kemeny orders*. When it exists, that is, in the absence of cycles in the majority relation, the Condorcet order is the unique Kemeny order. As Young (1995) points out, the Kemeny rule can also be seen as a way of compromising between conflicting opinions.

---

<sup>12</sup>In figure skating, which inspired this paper, this is a legitimate point of view since the intent of the ISU regulations is clearly to have judges furnish an evaluation of the relative merits of the competitors in terms of scores. They are even instructed on how to subtract points for different types of mistakes.

Given the distribution function  $f^{\alpha\beta}(\cdot|r)$  for profiles, we may define the *likelihood function*  $L(\cdot; X)$  by:

$$L(r; X, \alpha, \beta) \equiv f^{\alpha\beta}(X|r)$$

This is simply the distribution function re-interpreted as a function of  $r$ , given  $X$ . A *Maximum Likelihood* estimator for the social order is a solution to the problem:

$$\max_{r \in \mathcal{L}} L(r; X, \alpha, \beta) \tag{8}$$

**Definition 6** The correspondence  $ML(\cdot; \alpha, \beta) : \mathcal{L}^n \rightarrow \mathcal{R}$  that assigns to each profile  $X$ , the subset  $ML(X, \alpha, \beta) = \arg \max_{r \in \mathcal{L}} L(r; X, \alpha, \beta)$  is the *Maximum Likelihood (ML) rule* corresponding to the distribution function  $f^{\alpha\beta}(\cdot|r)$ .

In order to take into account the fact that, in real applications, the true parameters of the distribution function are not known with certainty, we also solve the problem (8) using parameters  $(\hat{\alpha}, \hat{\beta})$  that are different from the values  $(\alpha, \beta)$  used to simulate the profiles.<sup>13</sup>

Since  $ML(X; \hat{\alpha}, \hat{\beta})$  is conditioned on the choice of  $(\hat{\alpha}, \hat{\beta})$ , it is not, strictly speaking, a set of true ML estimators. In principle, both  $r$  and  $(\alpha, \beta)$  should be estimated jointly, but since maximizing the likelihood over both  $r$  and  $(\alpha, \beta)$  would require a round of numerical optimization at each of the  $m!$  elements in  $\mathcal{L}$ , this technique could only be used in applications where  $m$  is quite small. In this study, we are obliged to solve (8) for each draw of the Monte Carlo study, a task that requires  $m!$  evaluations of the likelihood function. Note also that even if the true  $(\alpha, \beta)$  were known, the properties of  $ML(\cdot; \alpha, \beta)$  could only be analyzed for small values of  $m$ .

For  $\hat{\beta} = 0$ , the Kemeny and the ML rules coincide, but not when  $\hat{\beta} > 0$ . Moreover, the Condorcet order, when it exists, is not necessarily a most likely order in this more general model.

Another likelihood-based rule is the *minimum expected loss* (MEL) estimator developed in Gordon and Truchon (2007). However, since each application of the MEL rule involves running an application of a Markov chain Monte Carlo algorithm, it is not yet feasible to evaluate the *ex ante* risk of the MEL rule.

---

<sup>13</sup>The ML rule in Drissi and Truchon (2004) sets  $(\hat{\alpha}, \hat{\beta}) = (\alpha, \beta)$ .

### 3.2.4 The ISU-94 Rule

This is essentially rule 371 of the ISU Regulations of 1982 and 1994 - hence the name ISU-94. Its definition involves four principles or criteria. When a principle has been applied, the next one is used only if there remain ties between some competitors.

Formally, the first criterion is the median rank  $\rho(s)$  of each skater  $s$ , that is, the smallest rank such that a majority of judges have placed skater  $s$  at one of the ranks  $1, \dots, \rho(s)$  (par. 1, 2, 6)<sup>14</sup>: the smaller the median rank the better. If two skaters, say  $s$  and  $t$ , obtain the same median rank, one tries to break the tie by reverting to the numbers of judges  $n_\rho(s)$  and  $n_\rho(t)$  who gave the median or a better rank to  $s$  and  $t$  respectively (par. 3): the more the better. If this is not sufficient to break all ties, one then tries to do so by using the sum of ranks of tied skaters according to the judges who gave the median rank or a better rank to these competitors (par. 4). For skater  $s$ , this number is written as  $b_\rho(s)$ : the smaller the better. If there still remains some ties after applying this third criterion, one uses the sum of ranks given to  $s$  by all judges (par. 5), or equivalently the Borda scores  $b(s)$ , as a breaking criterion. Finally, if all ties are not resolved after these four principles have been applied, competitors who tie for a rank obtain the same rank (par. 10).

These criteria are defined as follows. First, given a profile  $X$ , let:

$$J_i(s) = \{j \in J : x_s^j \leq i\} \text{ and } n_i(s) = |J_i(s)|$$

$$\text{Then, } \rho(s) = \min_{i \in \{1, \dots, m\}} i \text{ such that } n_i(s) > \frac{|J|}{2} = \frac{n}{2},$$

Next, for the sake of simplicity, we write  $J_\rho(s) = J_{\rho(s)}(s)$  and  $n_\rho(s) = n_{\rho(s)}(s)$

$$\text{Finally, let } b_\rho(s) = \sum_{j \in J_\rho(s)} x_s^j \text{ and recall that } b(s) = \sum_{j \in J} x_s^j.$$

Note that  $\frac{b(s)}{n}$  is the mean rank of skater  $s$  while  $\frac{b_\rho(s)}{n_\rho(s)}$  is the mean rank of skater  $s$  according to the judges of  $J_\rho(s)$ . Now, for each candidate  $s \in A$ , define the vector  $u(s) \in \mathbb{R}^4$  by

$$u(s) = (\rho(s), n - n_\rho(s), b_\rho(s), b(s))$$

and let  $\leq_\ell$  be the lexicographic weak order<sup>15</sup> on  $\mathbb{R}^4$ . We can now give the definition of the ISU-94 rule.

<sup>14</sup>These are the paragraphs of rule 371 of the regulations. They are reproduced in Truchon (2004).

<sup>15</sup>Let us first define  $<_\ell$ . For  $a, b \in \mathbb{R}^L$ ,  $a <_\ell b$  if  $\exists k \in \{1, \dots, L\} : a_k < b_k$  and  $a_l = b_l \forall l < k$ . Then,  $a \leq_\ell b$  if either  $a = b$  or  $a <_\ell b$ . The reader can check that  $\leq_\ell$  is transitive.



**Definition 7** The *ISU-94 rule* is the function  $ISU^{94} : \mathcal{L}^n \rightarrow \mathcal{R}$  defined by:

$$\forall s, t \in A : ISU_s^{94}(X) \leq ISU_t^{94}(X) \Leftrightarrow u(s) \leq_\ell u(t)$$

$ISU^{94}(X)$  is the *ISU-94 ranking*.

Note that if  $u(s) = u(t)$  for some  $s$  and  $t$ , then  $s$  and  $t$  obtain the same rank. Table 2 illustrates the computation of the four criteria of this rule.

$s$	Profile					Criteria-94				$s$	Final ranking
	1	2	3	4	5	$\rho$	$n_\rho$	$b_\rho$	$b$		
A	1	2	1	2	2	2	5	8	8	A	1
B	2	1	3	2	1	2	4	6	9	B	2
C	5	4	2	4	3	4	4	13	18	C	3
D	3	3	4	5	4	4	4	14	19	D	4
E	4	5	5	1	5	5	5	20	20	E	5

Table 2: Illustration of the ISU-94 rule

### 3.2.5 The ISU-98 Rule

This rule<sup>16</sup> represents an important simplification with respect to the ISU-94 rule. Now, skaters are first ranked according to the numbers of other skaters that they defeat according to a majority of judges, or with whom they tie. The Borda rule is once again used to break ties according to the first criterion.<sup>17</sup>

Formally, let  $\mu_{st} : \mathcal{L}^n \rightarrow \mathbb{R}$ , be a function defined for every pair of alternatives  $(s, t)$  and every profile  $X$ , by:

$$\mu_{st}(R) = \begin{cases} 1 & \text{if } \sum_{j=1}^n x_{st}^j \geq \sum_{j=1}^n x_{ts}^j \\ 0 & \text{otherwise} \end{cases}$$

Then, for each alternative  $s$ , let:

$$c^*(s) = \sum_{t \in A \setminus \{s\}} \mu_{st}(X)$$

<sup>16</sup>This rule is described and illustrated in Communication No. 997 of the ISU, available on the web site [www.isu.org](http://www.isu.org).

<sup>17</sup>This combination of the Copeland and the Borda rules is all the more curious that they may rank candidates in reverse orders. See Straffin (1980) and Moulin (1988, Exercice 9.1).

This number is the *Copeland score* of  $s$ . This is the number of alternatives over which  $s$  gets a strict majority of votes or with which it ties.<sup>18</sup> Let  $v(s) = (c^*(s), b^*(s))$ . The formal definition of the ISU-98 rule is the following.

**Definition 8** The *ISU-98 rule* is the function  $ISU^{98} : \mathcal{L}^n \rightarrow \mathcal{R}$  defined by:

$$\forall s, t \in A : ISU_s^{98}(X) \leq ISU_t^{98}(X) \Leftrightarrow v(s) \geq_\ell v(t)$$

$ISU^{98}(X)$  is the *ISU-98 ranking*.

Put simply, skaters are first ranked according to their Copeland scores  $c^*(s)$ . If two skaters, say  $s$  and  $t$ , obtain the same Copeland score, one tries to break the tie by comparing the Borda scores  $b^*(s)$  and  $b^*(t)$ . Finally, if all ties are not resolved after the application of these two criteria, that is, if  $v(s) = v(t)$  for some  $s$  and  $t$ , competitors who tie for a rank obtain the same rank. Table 3 illustrates the application of this rule. Note that we use the binary matrix  $\left[ \sum_{j=1}^n x_{st}^j \right]_{s,t \in A}$  as the starting point since we can easily compute the number of wins under the majority rule from this matrix. Accordingly,  $b^*$  instead of  $b$  is the breaking criterion. Recall that  $b^*(s) + b(s) = mn$ .

$s$	Profile							
	1	2	3	4	5	6	7	8
A	1	5	1	4	3	3	3	3
B	2	2	3	3	5	2	4	1
C	3	4	4	1	1	4	5	2
D	4	3	2	5	2	5	1	4
E	5	1	5	2	4	1	2	5

$s$	Binary matrix					Criteria-98	
	A	B	C	D	E	$c^*$	$b^*$
A	0	4	4	5	4	4	17
B	4	0	6	5	3	3	18
C	4	2	0	5	5	3	16
D	3	3	3	0	5	1	14
E	4	5	3	3	0	2	15

$s$	Final ranking
A	1
B	2
C	3
D	5
E	4

Table 3: Illustration of the ISU-98 rule

---

<sup>18</sup>Strictly speaking, the Copeland rule is defined with:

$$\mu_{st}(X) = \begin{cases} 1 & \text{if } \sum_{j=1}^n x_{st}^j > \sum_{j=1}^n x_{ts}^j \\ \frac{1}{2} & \text{if } \sum_{j=1}^n x_{st}^j = \sum_{j=1}^n x_{ts}^j \\ 0 & \text{otherwise} \end{cases}$$

That is, a tie is given half the weight of a win. For the difference that it makes, see Truchon (2004).

### 3.3 Loss functions

We use the family of parametric loss functions presented in Gordon and Truchon (2007):

$$d^{m\theta}(r, \hat{r}) = \sum_{t \in A} \left( \sum_{t \in A} \gamma_{st}(r, \hat{r}) \right)^\eta (m - r_s + 1)^\theta, \quad \eta \geq 0, \theta \geq 0 \quad (9)$$

Note that  $d^{1,0}$  is the Kemeny metric. With the convention  $0^0 = 0$ ,  $d^{00}$  is the naive or degenerate distance  $d^N$  defined by

$$d^N(r, \hat{r}) = \begin{cases} \kappa & \text{if } \hat{r} \neq r \\ 0 & \text{if } \hat{r} = r \end{cases}$$

for some  $\kappa > 0$ .

In order to get some intuition for plausible choices of  $\eta$  and  $\theta$ , we consider the case where there are  $m = 4$  alternatives. The errors associated with incorrectly reversing the first- and second-ranked, the second- and third-ranked and the third- and fourth-ranked alternatives are presented in Table 4. In order to keep a common scale, the loss associated with reversing the true order (that is, getting each of the 28 possible pairwise rankings wrong) is normalized to 1. In our judgement, the combinations in bold reflect the most plausible range of variation in the loss function parameters in (9), except maybe those for which  $\eta = 2$ ; see Gordon and Truchon (2007) for a more detailed discussion. However, we retain the combinations with  $\eta = 2$ , in order to test for the robustness of the results to the value of  $\eta$ . We also include the naive distance, represented by  $d^{00}$ .

### 3.4 Simulation methodology

Without loss of generality, the simulations are done under the assumption that the true order is  $r = (1, 2, \dots, m)$ . Thus, we may drop  $r$  from  $f^{\alpha\beta}(\cdot | r)$  and  $d^{m\theta}(r, \cdot)$ , and write the risk defined in (5), that is, the ex ante expected loss of an aggregation rule  $\Gamma$ , given  $f^{\alpha\beta}$  and  $d^{m\theta}$ , as:

$$\rho(\Gamma; f^{\alpha\beta}, d^{m\theta}) = \sum_{X \in \mathcal{L}^n} f^{\alpha\beta}(X) d^{m\theta}(\Gamma(X)) \quad (10)$$

Computing the risk in (10) associated with a given aggregation rule using purely analytical techniques requires evaluating  $f^{\alpha\beta}(X)$  at every profile in  $\mathcal{L}^n$ , and is only feasible for low-dimension problems, that is, with a small number of alternatives and few judges. In this study, we use Monte Carlo sampling techniques to estimate the population expectation in (10). In each experiment, we choose:

Parameters		Ranking error		
$\eta$	$\theta$	$1^{st} \leftrightarrow 2^{nd}$	$2^{nd} \leftrightarrow 3^{rd}$	$3^{rd} \leftrightarrow 4^{th}$
0.5	0.0	0.241	0.241	0.241
0.5	0.5	0.273	0.236	0.193
<b>0.5</b>	<b>1.0</b>	<b>0.304</b>	<b>0.228</b>	<b>0.152</b>
0.5	2.0	0.360	0.203	0.090
<b>1.0</b>	<b>0.0</b>	<b>0.167</b>	<b>0.167</b>	<b>0.167</b>
<b>1.0</b>	<b>0.5</b>	<b>0.184</b>	<b>0.159</b>	<b>0.130</b>
<b>1.0</b>	<b>1.0</b>	<b>0.200</b>	<b>0.150</b>	<b>0.100</b>
<b>1.0</b>	<b>2.0</b>	<b>0.229</b>	<b>0.129</b>	<b>0.057</b>
2.0	0.0	0.071	0.071	0.071
<b>2.0</b>	<b>0.5</b>	<b>0.076</b>	<b>0.066</b>	<b>0.054</b>
<b>2.0</b>	<b>1.0</b>	<b>0.080</b>	<b>0.060</b>	<b>0.040</b>
<b>2.0</b>	<b>2.0</b>	<b>0.087</b>	<b>0.049</b>	<b>0.022</b>

Table 4: Loss function parameters and implications for costs of errors

- i) a number of alternatives  $m \in \{3, 4, 5, 6, 7, 8, 9, 10, 12, 20\}$ ,
- ii) a number of judges  $n \in \{3, 9\}$ ,
- iii) a combination of parameters for the probability function from Table 1,
- iv) a combination of parameters for the loss function from those in bold in Table 4.

We then simulate  $K$  profiles from the chosen statistical model. More precisely, for each voter  $j$  and each pair of alternatives  $(s, t)$  such that  $s < t$ , we draw  $u \in [0, 1]$  from a uniform distribution. If  $u < p_{st}^{\alpha\beta}$ , we set  $\nu_{st}^j = 1$  and  $\nu_{ts}^j = 0$ . Otherwise, we set  $\nu_{st}^j = 0$  and  $\nu_{ts}^j = 1$ . Draws that do not correspond to an order are discarded. We repeat the process for each judge, to obtain a profile  $X$ . Repeating  $K$  times yields a sample  $\Omega^{\alpha\beta mn}$ .

In a second step, we apply the aggregation rules described in subsection 3.2, that is, the ISU-94, the ISU-98, the Borda, and the Kemeny rules<sup>19</sup> to each profile. With 3 to 6 (sometimes 7) alternatives, we also apply the maximum likelihood rule  $ML(\cdot, \hat{\alpha}, \hat{\beta})$ .<sup>20</sup> We

<sup>19</sup>The ISU-94, the ISU-98 and the Borda rules yields a unique ranking. However, there may be more than one Kemeny order. In this case, the mean Kemeny ranking, defined in the Appendix, has been retained when it existed. Otherwise, a ranking close to the Copeland ranking has been chosen.

<sup>20</sup>As is the case for the Kemeny rule, there may be more than one maximum likelihood order. The mean maximum likelihood ranking has been retained when it existed. Otherwise, a ranking close to the chosen Kemeny ranking has been retained. See the Appendix for more on this point.

always start with  $(\hat{\alpha}, \hat{\beta}) = (\alpha, \beta)$  but we also try other values of  $\hat{\beta}$  to check for the robustness of the rule.<sup>21</sup> These four of five rules make up the set  $\mathcal{G}$ .

Finally, for each rule  $\Gamma^g \in \mathcal{G}$  and each loss function  $d^{\eta\theta}$ , we compute the statistic:

$$\hat{\rho}(\Gamma^g; \Omega^{\alpha\beta mn}, d^{\eta\theta}) = \frac{1}{K} \sum_{X \in \Omega^{\alpha\beta mn}} d^{\eta\theta}(\Gamma^g(X)) \quad (11)$$

This statistic is a simulation-consistent estimator for the risk associated with a given aggregation rule. The value of  $K$  is set so that the standard errors are below 1% (2% in one subset of simulations) of their corresponding estimates.

Five sets of simulations were performed. In the first, the number of judges is 3 and the number of alternatives ranges from 3 to 10 and then jumps to 12 and 20. The values of the parameters for simulating the votes are  $\alpha = 1.1$  and  $\beta = 0.9$ . In the second, the number of judges is raised to 9, as in figure skating, and the the number of alternatives ranges from 3 to 10 and 12. The third set is similar to the second, except that  $\alpha = 1$  and  $\beta = 0.3$ . In the fourth set,  $\beta = 0$  and three different values are used for  $\alpha$ , namely 1.1, 2.2 and 0.4. The number of alternatives ranges from 3 to 10 when  $\alpha = 1.1$  and from 3 to 6 otherwise. In the last set,  $\alpha = 0.4$  and  $\beta = 1$ , that is, we go back to variable probabilities, but with a lower competency for the judges. The respective characteristics of the five sets of simulations are given in Table 5. Note that the fourth set is broken into three subsets.

---

<sup>21</sup>The set of maximizers of the likelihood function is invariant to a multiplication of  $(\hat{\alpha}, \hat{\beta})$  by a positive scalar  $\gamma$ . Thus, we vary only  $\hat{\beta}$ . However, the choice of the absolute values of the parameters  $\alpha$  and  $\beta$  in the vote generating function is not innocuous. The larger these values, the larger the probabilities of correctly ordering two alternatives, the larger the probability that a judge selects the true ranking or at least a ranking close to the true one, and the smaller the probability of a cycle in her vote.

Set or subset	1	2	3	4.1	4.2	4.3	5
number of judges	3	9	9	9	9	9	9
number of alternatives	3-10, 12, 20	3-10, 12	3-10, 12	3-10	3-6	3-6	3-6
parameters of the probability function	$\alpha = 1.1$ $\beta = 0.9$	$\alpha = 1.1$ $\beta = 0.9$	$\alpha = 1.0$ $\beta = 0.3$	$\alpha = 1.1$ $\beta = 0$	$\alpha = 2.2$ $\beta = 0$	$\alpha = 0.4$ $\beta = 0$	$\alpha = 0.4$ $\beta = 1$
parameters of the maximum likelihood rule	$\hat{\alpha} = 1.1$ $\hat{\beta} = 0.9$	$\hat{\alpha} = 1.1$ $\hat{\beta} = 0.9, \dots, 3$	$\hat{\alpha} = 1.0$ $\hat{\beta} = 0.3, 0.5$	$\hat{\alpha} = 1.1$ $\hat{\beta} = 1$	$\hat{\alpha} = 2.2$ $\hat{\beta} = 1$	$\hat{\alpha} = 1.1$ $\hat{\beta} = 1$	$\hat{\alpha} = 0.4$ $\hat{\beta} = 1, 3$
number of alt. with maximum likelihood rule	3-7	3-6	3-6	3-6	3-6	3-6	3-6
precision	1%	1%	1%	1%	2%	1%	1%

Table 5: Description of the simulations

## 4 The results in brief

In assessing the behavior of the different rules, four facts should be taken into account:

- $ML(0)$  and the Kemeny rule coincide,
- $ML(\hat{\beta})$  and the Borda rule coincide when  $\hat{\beta} = \hat{\alpha}$ , which, in our case, means  $\hat{\beta} = \alpha$ ,<sup>22</sup>
- the Kemeny and the ISU-98 rules are both Condorcet consistent, that is, they select the Condorcet order when it exists;<sup>23</sup> thus, we can expect them to give similar results,
- $ML(\hat{\beta})$  should do best with  $\hat{\beta} = \beta$  and thus,  $ML$  be decreasing from 0 to  $\beta$  and increasing afterward.

As it turned out,  $ML(\hat{\beta})$  did best, not only in the  $ML$  family, but against all other rules, with a minor exception<sup>24</sup>. Because of the four facts mentioned above, the relative position of the Borda, the Kemeny and the ISU-98 rules, in terms of expected loss, depends on the relative values of  $\alpha$  and  $\beta$ . So do the relative position of the Borda and the ISU-94 rules and the discrepancy between the Borda and the  $ML(\beta)$  rules. *A priori*, we cannot say much about the ISU-94 rule.

<sup>22</sup>This assertion is proved in Truchon (2006) under the hypothesis that a subset of most likely orders that differ only by the permutation of adjacent alternatives, are replaced by a single ranking in which these alternatives tie.

<sup>23</sup>It is well known that the Borda rule is not Condorcet consistent. The ISU-94 rule is not either. See Remark 5 in Truchon (2004) for a proof.

<sup>24</sup>The exception is for the case  $(\alpha, \beta) = (1.0, 0.3)$  and  $\eta = 2$ . See Table 11 in Appendix B. The explanation is given there.

Figures 1 to 4 illustrate four typical behaviors. Expected losses are those obtained for six alternatives, nine judge and  $\eta = \theta = 1$ . The figures have been drawn from a few values of  $\hat{\beta}$ , hence the lack of smoothness. From these figures, we note the following:

- When  $\beta$  (the rate of increase in the probability) is large,  $ML(0)$  is large compared to  $ML(\beta)$ . Thus, the Kemeny and the ISU-98 rules give large and similar expected losses since  $ML(0)$  and the Kemeny rule coincide and since the Kemeny and the ISU-98 rules are both Condorcet consistent. This is the case in Figures 1 and 2. The ISU-94 rule fares better than the Kemeny and the ISU-98 rules in one case and slightly worst in the other.
- When  $\beta$  is low, as in Figures 3 and 4,  $ML(0)$  is close to  $ML(\beta)$ . Thus, the Kemeny rule yields an expected loss close to what  $ML(\beta)$  gives. As a consequence, it does almost<sup>25</sup> always better than the maximum likelihood rule. The ISU-98 rule is the only one that does better but by a very small margin. Then, the chance is high that the ISU-94 rule will give a larger expected loss, as is the case in these two figures. The Borda rule does worst due to the facts that  $\beta < \alpha$  and that  $ML$  is an increasing function past  $\beta$ .
- When  $\alpha$  and  $\beta$  are close to each other, as in Figure 2, the Borda rule fares well since it has an expected loss close to  $ML(\beta)$ , which is the smallest of all. This leaves little room for the other rules. Thus, the chances are that the two ISU and the Kemeny rules give a much larger expected loss than all others, except  $ML(\hat{\beta})$  for values of  $\hat{\beta}$  far from  $\beta$ . This is what we observe in Figure 2.
- The Borda rule may also fare relatively well when  $\beta$  is much larger than  $\alpha$ , as in Figure 1, provided that  $\alpha$  is large enough, which it must be for the probability of correctly ordering two alternatives to be larger than one half. The performance of the maximum likelihood rule also appears to be robust to the value of  $\hat{\beta}$  above  $\beta$ .
- When  $\alpha$  is much larger than  $\beta$ , as in Figures 3 and 4, the Borda rule yields a large expected loss, compared to  $ML(\beta)$ . The reason is that the Borda rule coincides with  $ML(\alpha)$  but, when  $\hat{\beta}$  reaches  $\alpha$ , the expected loss given by  $ML$  has increased by a large amount. Typically, all other rules will do better than Borda, except  $ML(\hat{\beta})$  for  $\hat{\beta} > \alpha$ .

---

<sup>25</sup>The rare exceptions seem to be due to the different handling of multiple Kemeny and most likely orders.

Appendix B presents a full description of the simulations. Here are a few additional points that are worth mentioning here.

- The few simulations that were done with three judges generated results that are sometimes quite different from those obtained with nine judges. For example, with  $(\alpha, \beta) = (1.1, 0.9)$  and less than 6 alternatives, the Borda rule produces the largest expected loss of all rules. Its performance improves as the number of alternatives increases. With 8 alternatives and more, it becomes the best rule.
- Setting  $\eta = 2$  improves the performance of the Borda rule. The reason is that most errors under the Borda rule consist in declaring *ex aequo* two adjacent alternatives in the true order and that kind of error is worth a penalty of  $\frac{1}{2}$  compared to 1 for an inversion. These penalties entering the function  $(\cdot)^\eta$ , the result is a lower expected loss with  $\eta = 2$  than with  $\eta = 1$ .
- Except for the Borda rule, the value of  $\theta$  has an important impact on the expected losses but not on the rankings of the rules in terms of these losses. This may be an indication that, under all rules, most errors involve essentially the same alternatives or close alternatives in the true order. Increasing  $\theta$  from 1 to 2 is often beneficial to the Borda rule. This is an other indication that this rule often produces ties between alternatives.

## 5 Conclusion

If it is supposed that a true social order exists, but that individual voters observe it with error, then the social choice problem is one of optimal inference. This study - along with the companion paper of Gordon and Truchon (2007) - formalises and extends Condorcet's framework in two directions. Firstly, we make the choice-theoretic nature of the problem more explicit by specifying a loss function that describes the consequences of choosing an incorrect social ranking. Secondly, we extend the usual constant-probability case to one in which voters find it easier to distinguish alternatives that are far apart, and more difficult to correctly order similar alternatives.<sup>26</sup>

This study develops a statistical procedure for the *ex ante* evaluation of vote aggregation rules. In our Monte Carlo application, we specify both the distribution function that

---

<sup>26</sup>In Gordon and Truchon (2007), there is actually a third extension that consist in taking into account the possibility that judges may have a bias for alternatives with certain features, such as a shared nationality.



generates individual votes and a loss function that can be used to evaluate the expected loss associated with a given decision rule.

We find that the maximum likelihood (ML) rule, with the correct form and the true parameters of the vote distribution function, generally outperforms its competitors, that is, the Borda, the Kemeny and two rules that have been used for many years in figure skating. Moreover, ML seems to be somewhat robust to the value of  $\hat{\beta}$  when  $\alpha$  is small. However, its feasibility in real-world applications, where the form and the parameters of the distribution function are not known, is limited. In addition, the computational burden forbids the use of ML with a large number of alternatives, such as in figure skating.

The performance of the other rules varies considerably with the value of the parameters of the probability function used to generate the votes. For example, the Borda rule does well with increasing probabilities but poorly with constant probabilities. It appears that no single vote aggregation rule can be expected to perform best in all circumstances.

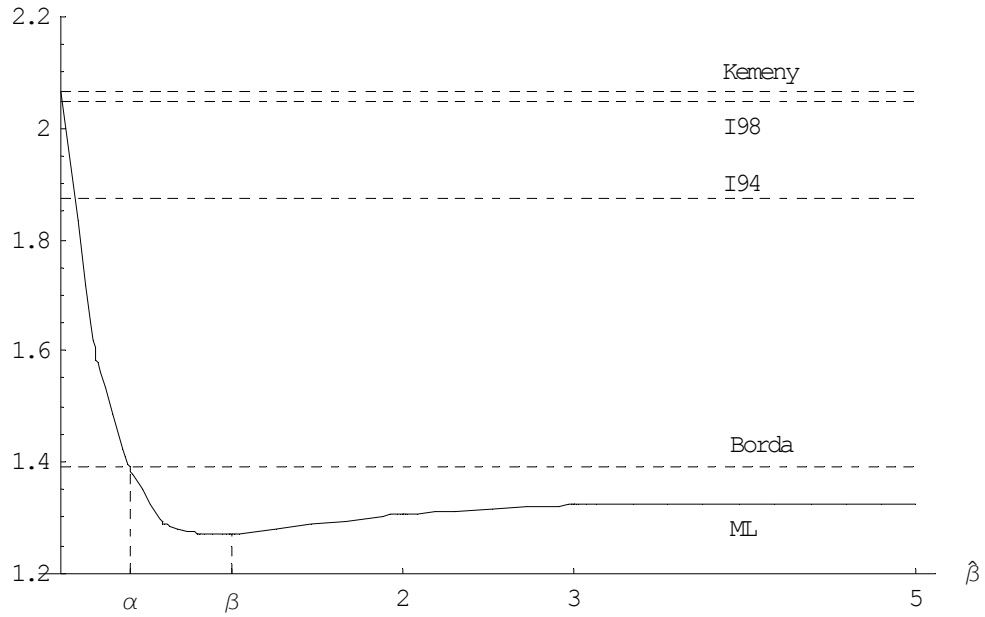


Figure 1: Typical expected losses as a function of  $\hat{\beta}$  when  $\alpha = 0.4$  and  $\beta = 1$

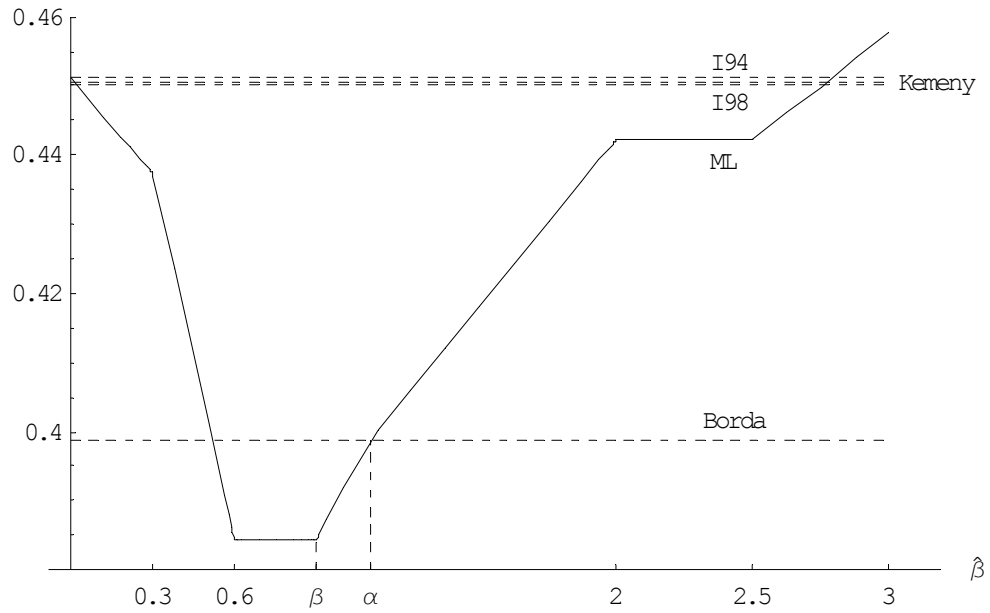


Figure 2: Typical expected losses as a function of  $\hat{\beta}$  when  $\alpha = 1.1$  and  $\beta = 0.9$

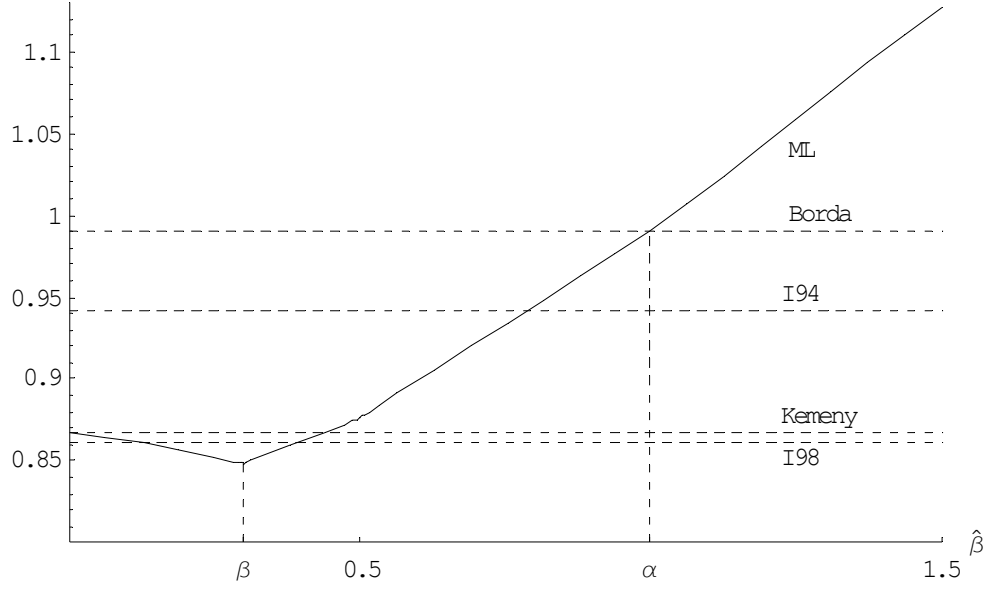


Figure 3: Typical expected losses as a function of  $\hat{\beta}$  when  $\alpha = 1$  and  $\beta = 0.3$

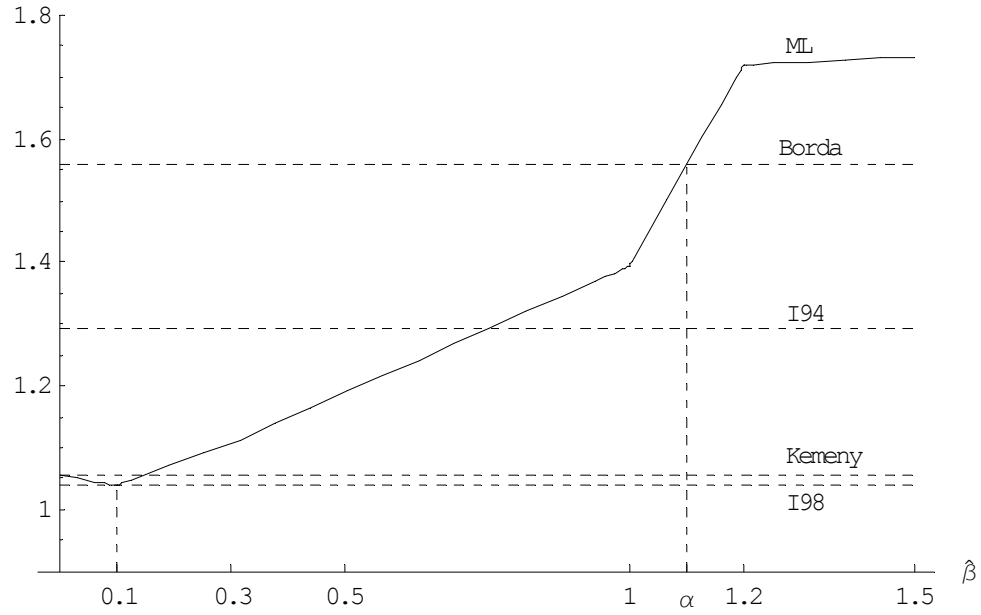


Figure 4: Typical expected losses as a function of  $\hat{\beta}$  when  $\alpha = 1.1$  and  $\beta = 0$

# Appendices

## A Finding Kemeny orders and choosing among them

Finding a Kemeny order by considering all possible permutations of the skaters, as the definition suggests, may prove prohibitive when the number of competitors is large as in figure skating. There are many contributions in the literature that address this specific problem.<sup>27</sup> Whatever the method used, Truchon (2004) shows that one way to ease the task is to break the problem into many subproblems, using an extension of Condorcet Consistency that he calls the *Extended Condorcet Criterion* (XCC). In essence, this criterion says that if the set of alternatives can be partitioned in such a way that all members of a subset of this partition defeat all alternatives belonging to subsets with a higher index, then the former should obtain a better rank than the latter. Such a partition is called a *Condorcet partition*. The finest Condorcet partition can be found by first ordering alternatives according to another rule that satisfies XCC, such as the Copeland rule. For most practical applications, a Kemeny order on each subset of the finest Condorcet partition can be found by simple enumeration. Then, a complete Kemeny order is obtained by juxtaposing the Kemeny orders on the subsets of the partition. This is the approach adopted in our study. However, this approach does not extend to the *ML* rule since this rule does not satisfy XCC. For this reason, we have to limit the number of alternatives to 6 or 7 when applying this rule.

The Kemeny rule may yield more than one order. This occurrence can sometimes be resolved by reverting to the concept of the *mean Kemeny ranking* also taken from Truchon (2004). Given a set of Kemeny orders  $\{\hat{r}^1, \dots, \hat{r}^k\}$ , consider the weak order  $\tilde{r}$  defined by:

$$\forall s, t \in A : \tilde{r}_s \leq \tilde{r}_t \Leftrightarrow \sum_{q=1}^k \hat{r}_s^q \leq \sum_{q=1}^k \hat{r}_t^q$$

This weak order is a ranking according to the mean ranks of alternatives over all Kemeny orders. It will be called the *mean Kemeny ranking* if and only if it weakly agrees with at least one order in  $\{\hat{r}^1, \dots, \hat{r}^k\}$ , that is, if there exists an order  $\hat{r}^q \in \{\hat{r}^1, \dots, \hat{r}^k\}$  such that:

$$\forall s, t \in A : \hat{r}_s^q < \hat{r}_t^q \Rightarrow \tilde{r}_s \leq \tilde{r}_t$$

We may define a *mean ML ranking* in a similar manner.

---

<sup>27</sup>Several of these contributions are based on mathematical programming. See Truchon (2004) for some references.

## B Description of the results

The complete results of the simulations, that is, the expected losses, are available upon request from the authors. Here, we only present the main highlights of these results. There are five subsections corresponding to the five sets of simulations. In the tables of this appendix, the notation  $Y \prec Z$  stands for rule  $Y$  yields a lower expected loss than rule  $Z$ ,  $Y \sim Z$  means that the expected losses are the same under  $Y$  and  $Z$ , and  $B, K, I_{94}, I_{98}$  represent the Borda, the Kemeny, the ISU-94 and the ISU-98 rules respectively.  $ML$  represents the maximum likelihood rule with  $(\hat{\alpha}, \hat{\beta}) = (\alpha, \beta)$  while  $ML(\hat{\beta})$  represents the same rule for the specified value of  $\hat{\beta}$  and for  $\hat{\alpha} = \alpha$ .

### B.1 Simulations with three judges and $(\alpha, \beta) = (1.1, 0.9)$

The main highlights of these simulations are:

- Save the Borda rule, increasing  $\eta$  has a relatively small effect on the expected loss. This is an indication that most errors consist in inverting adjacent alternatives in the true order.
- In the case of Borda, the impact of increasing  $\eta$  is also small but it goes in the reverse direction. It causes the expected error to decrease. This is a confirmation that most errors consist in declaring ex aequo two adjacent alternatives in the true order. Recall that tying two adjacent alternatives is worth a penalty of  $\frac{1}{2}$  compared to 1 for an inversion, hence the decrease in the size of the error as the parameter  $\eta$  increases.
- The Borda rule always produces the largest proportion of erroneous rankings. This translates into the worst expected loss for this rule when  $(\eta, \theta) = (0, 0)$  or  $(\frac{1}{2}, 1)$ . But, for  $\eta = 2$ , the expected loss is always smaller under this rule than under any other rule, for the reason given above.
- Table 6 shows the relation between the different rules for the other values of the parameters, that is,  $(\eta, \theta) = (1, 0), (1, \frac{1}{2}), (1, 1), (1, 2)$ .
- From this table, one can note that, for 3 alternatives, all rules other than Borda give the same expected losses.
- With the number of alternatives ranging from 4 to 7, the expected loss is always smaller under the maximum likelihood rule than under any other rule. This was to be expected since this rule incorporates the hypothesis used in generating the profiles.

Alternatives	Relations between rules
3	$ML \sim I_{98} \sim I_{94} \sim K \prec B$
4-5	$ML \prec I_{98} \prec I_{94} \prec K \prec B$
6	$ML \prec I_{98} \prec B \prec I_{94} \prec K$
7	$ML \prec B \prec I_{98} \prec I_{94} \prec K$
8-20	$B \prec I_{98} \prec I_{94} \prec K$

Table 6: Relation between different rules in terms of expected loss with 3 judges and  $\eta = 1$

- As we increase the number of alternatives, the Borda rule gets a better position with respect to the other rules. The relation between the other rules does not change. Given that the ISU-98 and the Kemeny rules are both Condorcet consistent while the ISU-94 is not, the fact that the ISU-94 rule gives expected losses between those obtained with the two others is somewhat puzzling. However, it must be said that the two ISU and the Kemeny rules produce results that are not markedly different from one another.
- In terms of the naive distance, the order between the two ISU and the Kemeny rules is not always the one described above. This is also often the case with  $\eta = \frac{1}{2}$ . This illustrates the usefulness of a more subtle distance between rankings.
- For  $(\eta, \theta) = (2, \frac{1}{2}), (2, 1), (2, 2)$ , the relations of Table 6 are changed by simply moving the Borda rule at the head.
- Increasing  $\theta$  has an important impact on the expected losses but not on the rankings of the rules in terms of these losses.
- The expected loss under all loss functions increases with the number of alternatives as there is more scope for errors from the judges. The standard error also increases with the number of alternatives but not in proportion to the expected loss, especially for the naive distance. As a consequence, the size of the sample to meet 1% precision decreases with the number of alternatives.
- Table 7 gives the theoretical probabilities of cycles (rounded to 3 digits) for 3 to 10 alternatives and the observed fractions of cycles in the simulations.<sup>28</sup> As expected, they are close to each other. For 12 alternatives, we give only the observed fractions of cycles since computing the theoretical probability is prohibitive in this case.

---

<sup>28</sup>The theoretical probability of a cycle is 1 minus the sum of the probabilities over the set of orders.

Number of alternatives	3	4	5	6	7	8	9	10	12
Probability of cycles	0.122	0.273	0.411	0.527	0.622	0.698	0.759	0.8075	
Frequency of cycles	0.121	0.274	0.409	0.527	0.622	0.698	0.759	0.8081	0.878

Table 7: Probabilities and empirical frequencies of cycles with  $\alpha = 1.1$ ,  $\beta = 0.9$

## B.2 Simulations with nine judges and $(\alpha, \beta) = (1.1, 0.9)$

The main highlights of these simulations and the main differences with those of three judges are:

- The expected losses are markedly smaller than the corresponding losses with three judges. Condorcet (1785) showed that the larger the number of judges, the higher the probability that the majority rule under the binary approach yields the true order. In as much as the ISU-98, the Kemeny and the maximum likelihood rules respect the view of the majority on each pair of alternatives when these views are consistent, these results are not surprising at least for these rules.
- The standard errors are also smaller than the corresponding ones with three judges but the diminution is not as marked as it is for the means. Hence, larger samples have been required to meet 1% precision.
- The behavior of the Borda rule with respect to the parameter  $\eta$  is similar to what we observed with three judges. It is also similar for the other rules, except that, this time, the ISU-94 rule behaves more like the Borda rule, producing a lower expected loss as  $\eta$  increases. This is an indication that, with these two rules, most errors consist in declaring ex aequo two adjacent alternatives in the true order.
- For less than seven alternatives, the maximum likelihood rule has been applied with more than one pair of values for the parameters. This allows us to test for the robustness of this rule with respect to its parameters.
- Table 8 shows the relation between the rules for  $(\eta, \theta) = (1, 0), (1, \frac{1}{2}), (1, 1), (1, 2)$ .
- For  $(\eta, \theta) = (2, \frac{1}{2}), (2, 1), (2, 2)$ , the relations of Table 8 are changed by simply moving the Borda rule at the head in the cases of 3 to 6 alternatives.
- It can be observed that, when applied, the maximum likelihood rule with  $(\hat{\alpha}, \hat{\beta}) = (\alpha, \beta)$  gives expected losses smaller than any other rule. However, its performance deteriorates as  $\hat{\beta}$  is increased above  $\beta$ .

Alternatives	Relations between rules
3	$ML(0.9) \prec B \prec ML(1.5) \prec I_{94} \prec I_{98} \prec K \prec ML(1.75) \sim ML(2)$
4	$ML(0.9) \prec B \prec I_{94} \prec ML(2) \sim ML(2.5) \prec I_{98} \prec K \prec ML(3)$
5	$ML(0.9) \prec B \prec ML(1.5) \prec I_{94} \prec I_{98} \prec K \prec ML(1.75)$
6	$ML(0.9) \prec B \prec ML(2) \sim ML(2.5) \prec I_{94} \prec I_{98} \prec K \prec ML(3)$
7-10, 12	$B \prec I_{94} \prec I_{98} \prec K$

Table 8: Relation between different rules in terms of expected loss,  $\alpha = 1.1$ ,  $\beta = 0.9$ ,  $\eta = 1$

- With 3 and 5 alternatives and  $\hat{\beta} = 1.5$ , the maximum likelihood rule gives a better result than all other rules, save the Borda rule. With  $\hat{\beta} = 1.75$ , the expected loss under the maximum likelihood rule becomes worst than under all other rules. With 4 and 6 alternatives, we have the same kind of pattern but with  $\hat{\beta} = 2.5$  and 3 instead of 1.5 and 1.75 respectively. Thus, the deterioration of the maximum likelihood rule arises for a larger value of  $\hat{\beta}$ . It would be interesting to know if this difference is typical of the odd and even numbers of alternatives but we did not apply the maximum likelihood rule for larger numbers of alternatives.
- Turning to larger numbers of alternatives ( $\geq 7$ ), the ISU-94 rule does slightly better than the ISU-98 and the latter does slightly better than the Kemeny rule. This order is different from the one observed with three judges. However, the differences are not marked, specially between the ISU-98 and the Kemeny rules.<sup>29</sup> Again, these two rules being Condorcet consistent, it is not surprising that they give similar results. What is more surprising is that the ISU-94 does better. Note that the relation between the rules also holds for  $(\eta, \theta) = (\frac{1}{2}, 1)$  when there are more than six alternatives. This is not the case with the naive distance.
- Increasing  $\theta$  has again an important impact on the expected losses but not on the rankings of the rules in terms of these losses.
- For 3 alternatives, different rules give different expected losses, contrary to what we had with 3 judges.
- What was said about expected losses, standard errors and the size of the samples with respect to the number of alternatives, with three judges, remains true with nine judges.

---

<sup>29</sup>However, the difference in the expected losses is, most of the time, significant with a high degree of confidence.



- The theoretical probabilities of cycles are of course the same and the empirical frequencies sensibly the same.<sup>30</sup>

### B.3 Simulations with nine judges and $(\alpha, \beta) = (1.0, 0.3)$

The main highlights of these simulations and the main differences with the preceding are:

- The competency of the judges being lower, their votes are more erratic, resulting in larger expected losses, roughly twice as large as those obtained with  $(\alpha, \beta) = (1.1, 0.9)$ .
- The standard errors are also larger but the increase is not as marked as it is for the means. Hence, smaller samples have been required to meet 1% precision.
- As a result of the lower competency of the judges, the theoretical probabilities of cycles and the empirical frequencies are considerably larger.<sup>31</sup> See Table 9.

Number of alternatives	3	4	5	6	7	8	9	10	12
Probability of cycles	0.171	0.428	0.657	0.815	0.908	0.957	0.980	0.992	
Frequency of cycles	0.171	0.427	0.656	0.815	0.908	0.956	0.981	0.992	0.999

Table 9: Probabilities and empirical frequencies of cycles with  $\alpha = 1, \beta = 0.3$

- Table 10 shows the relation between the rules for  $(\eta, \theta) = (0, 0), (\frac{1}{2}, 1), (1, 0), (1, \frac{1}{2}), (1, 1), (1, 2)$ . Here,  $ML(\hat{\beta})$  represents the maximum likelihood rule for the specified value of  $\hat{\beta}$  and for  $\hat{\alpha} = 1$ .
- The \* indicate exceptions:
  - For 4 to 6 alternatives and the pair  $(\eta, \theta) = (0, 0)$ , the relation between  $ML(0.5)$  and  $I_{94}$  is inverted.
  - For 9 alternatives and the pairs  $(1, 0), (1, \frac{1}{2})$ , the relation between  $I_{94}$  and  $B$  is inverted.
  - For 10 and 12 alternatives and the pairs  $(1, 0), (1, \frac{1}{2}), (1, 1)$ , the relation between  $I_{94}$  and  $B$  is inverted.

---

<sup>30</sup>To give an example of the impact of the possibility of cycles in individual rankings, with 9 judges and 10 alternatives, close to 5 millions of binary matrices have been drawn in order to obtain 105 000 polls of consistent rankings (orders).

<sup>31</sup>As a result, with 12 alternatives, more than 209 millions of binary matrices have been drawn in order to obtain 33 000 polls of consistent rankings (orders).

Alternatives	Relations between rules
3	$ML(0.3) \prec I_{98} \prec K \prec ML(0.5) \prec I_{94} \prec B \prec ML(1.5)$
4-6	$ML(0.3) \prec I_{98} \prec K \prec ML(0.5) \prec^* I_{94} \prec B \prec ML(1.5)$
7-8	$I_{98} \prec K \prec I_{94} \prec B$
9-10, 12	$I_{98} \prec K \prec I_{94} \prec^* B$

Table 10: Relation between different rules in terms of expected loss,  $\alpha = 1$ ,  $\beta = 0.3$ ,  $\eta \neq 2$

- There is a striking difference between the above relations and those of Table 8 obtained with  $(\alpha, \beta) = (1.1, 0.9)$ . The Borda rule, which was the second best performer, after the maximum likelihood rule with  $(\hat{\alpha}, \hat{\beta}) = (\alpha, \beta)$ , is now the worst. The ISU-94 rule also gives worst results and becomes the second worst.
- Again, when applied, the maximum likelihood rule with  $(\hat{\alpha}, \hat{\beta}) = (\alpha, \beta)$  gives expected losses smaller than any other rule. However, its performance deteriorates as  $\hat{\beta}$  is increased above  $\beta$ . It becomes worst than Kemeny with  $\hat{\beta} = 0.5$  and worst of all with  $\hat{\beta} = 1.5$ .
- Table 11 shows the relation between the rules for  $(\eta, \theta) = (2, \frac{1}{2}), (2, 1), (2, 2)$ .

Alternatives	Relations between rules
3	$ML(0.5) \prec B \prec ML(0.3) \prec I_{98} \prec K \prec I_{94} \prec ML(1.5)$
4-6	$ML(0.5) \prec ML(0.3) \prec I_{98} \prec K \prec B \prec I_{94} \prec ML(1.5)$
7	$I_{98} \prec K \prec B \prec I_{94}$
8-10, 12	$B \prec^* I_{98} \prec K \prec I_{94}$

Table 11: Relation between different rules in terms of expected loss,  $\alpha = 1$ ,  $\beta = 0.3$ ,  $\eta = 2$

- There are two exceptions as indicated by the  $*$ : For 9 and 12 alternatives and  $(\eta, \theta) = (2, 2)$ , the relation between  $B$  and  $I_{98}$  is inverted.
- With  $\eta = 2$  instead of 1, the performance of the Borda rule improves. This rule jumps to the first place for 8 alternatives and more. The improvement is marginal for 4 to 7 alternatives. The performance of  $ML(0.5)$  also improves to the point of becoming the best of all, even better than  $ML(0.3)$ .

- The reason for the better performance of  $ML(0.5)$  and  $B$  when  $\eta = 2$  is due to the fact that these two rules produce more ties than the other rules. For example, in the case of 3 alternatives, the first 200 profiles produced more errors in the  $ML(0.5)$  ranking than in  $ML(0.3)$  but 3 of the errors under  $ML(0.5)$  were ties while there were no ties under  $ML(0.3)$ . Since the penalty entering the function  $(\cdot)^\eta$  is  $\frac{1}{2}$  for a tie instead of 1, the result is a lower expected loss with  $\eta = 2$  than with  $\eta = 1$ .

## B.4 Simulations with nine judges and a constant probability

In the preceding subsections, we saw that the maximum likelihood rule with  $(\hat{\alpha}, \hat{\beta}) = (\alpha, \beta)$  does better than the Kemeny rule, with the minor exception mentioned in the last point above. Recall that the Kemeny rule is the maximum likelihood rule with  $\hat{\beta} = 0$ . However, increasing  $\hat{\beta}$  above  $\beta$  causes the performance of the maximum likelihood rule to deteriorate. At some point, it becomes worst than that of the Kemeny rule. We now examine three subsets of simulations where votes are generated with constant probabilities, that is,  $\beta = 0$ .

### Simulations with a constant probability of 0.75

In the first subset of simulations, the votes have been cast with  $\alpha = 1.1$  and  $\beta = 0$ . This means that the probability of correctly ordering two alternatives is 0.75, the same as the probability of correctly ordering two adjacent alternatives with  $\alpha = 1.1$  and  $\beta = 0.9$ . The maximum likelihood rule has been applied with  $\hat{\alpha} = 1.1$  and  $\hat{\beta} = 0.1, 0.3, 0.5, 1.0$ . The main highlights of these simulations and the main differences with the preceding are:

- Keeping the probability constant means that judges vote in a more erratic way, which translates into larger expected losses compared to those with  $\alpha = 1.1$  and  $\beta = 0.9$ . With 3 alternatives, the expected losses are almost twice as large as those with increasing probabilities. The deterioration is slightly worst for the Borda rule. With 6 alternatives, they are more than twice as large, and almost three times for the Borda rule.
- Except for the case of three alternatives, expected losses are also larger than those obtained with  $\alpha = 1$  and  $\beta = 0.3$  and the difference increases with the number of alternatives. The increase is more marked for the Borda rule. For three alternatives, expected losses under the Borda rule are also larger than those obtained with  $\alpha = 1$  and  $\beta = 0.3$  but they are smaller for all other rules.

- For 4 to 6 alternatives, the relation between the rules, in terms of expected loss, is the following for all values of  $\eta$  and  $\theta$ , with two minor exceptions:

$$I_{98} \prec K \prec ML(0.3) \prec ML(0.5) \prec I_{94} \prec ML(1) \prec B \prec ML(1.2)$$

The two minor exceptions concern the relation  $I_{98} \prec K$  in the case of 4 alternatives. It becomes  $I_{98} \sim K$  when  $\eta = 0$  and  $K \prec I_{98}$  when  $\eta = \frac{1}{2}$ .

- For 3 alternatives, this relation is changed by moving  $ML(0.3)$  ahead of  $K$  when  $\eta = 2$  and by having both rules tie when  $\eta = 1$ .
- Introducing  $ML(0.1)$  into the picture has the following impact. In the case of 3 alternatives,  $ML(0.1)$  gives exactly the same expected loss as  $K$ . For 4 to 6 alternatives,  $ML(0.1)$  comes between  $I_{98}$  and  $K$  and sometimes before  $I_{98}$  depending on the values of  $\eta$  and  $\theta$ .
- Thus, contrary to what was to be expected, the maximum likelihood rule with a low value of  $\hat{\beta}$  (0.1 or 0.3) gives a slightly better result than the Kemeny rule. As far as we can see, this is due to the handling of multiple rankings produced by the two rules. However, the performance of the ML rule gets worst as  $\hat{\beta}$  increases. With  $\hat{\beta} = 0.5$ , it is still better than that of the ISU-94 rule but it gets worst than the latter, yet with  $\hat{\beta} = 1$ , it remains better than the Borda rule.
- For 7 to 10 alternatives, the relation is

$$I_{98} \prec K \prec I_{94} \prec B$$

except that  $I_{98} \succ K$  when there are 7 and 8 alternatives.<sup>32</sup>

- The above relations are similar to those obtained with  $(\alpha, \beta) = (1.0, 0.3)$  and  $\eta = 1$ . The striking fact is that they remain the same with  $\eta = 2$ , due to the bad behavior of the Borda rule.

### Simulations with a constant probability of 0.9

In the second subset, the votes have been cast with  $\alpha = 2.2$  and  $\beta = 0$ , that is, with a constant probability of 0.9. This high probability results in more votes that are identical to

---

<sup>32</sup>In the case of 9 alternatives, half a billion of binary matrices have been drawn in order to obtain 36 000 polls of consistent rankings (orders). For 10 alternatives, only 2 278 polls have been generated after four days of computing, due to the very high frequency of cycles. Hence, the precision is only 3.4%.

the true order. As a result, all rules select the true order most of the time. The fraction of erroneous rankings is less than 1%, with one exception. The Borda rule is again and by far the worst performer. The maximum likelihood rule has been applied with  $\hat{\alpha} = 2.2$  and  $\hat{\beta} = 1$  or, equivalently,  $\hat{\alpha} = 1.1$  and  $\hat{\beta} = 0.5$ . Thus, the rate of increase in the probabilities is weak. Yet, in all cases, the Kemeny rule gives a lower expected loss than the maximum likelihood rule. The relation between the rules in terms of expected losses is the following for the four numbers of alternatives considered and all values of  $\eta$  and  $\theta$  :

$$I_{98} \sim K \prec ML(0.5) \prec I_{94} \prec B$$

This relation indicates that the Kemeny and the ISU-98 rules give expected losses that are equal or not significantly different. The explanation is straightforward. Most votes being identical to the true order, there is a Condorcet order with most profiles and both rules being Condorcet consistent, select this order.

This subset of simulations has been done with a precision of 2% instead of 1%. As remarked before, the lower the expected loss, the larger the number of profiles required to achieve a given degree of precision. The reason is that the variance does not decrease proportionally to the expected loss. In this case, more than one million of draws have been necessary to achieve a precision of 2%. These numbers would have had to be multiplied by 4 to achieve a precision of 1%.

### Simulations with a constant probability of 0.6

In the third subset, the votes have been cast with  $\alpha = 0.4$  and  $\beta = 0$ , that is, with a constant probability of 0.6. This means that the competency of the judges in assessing the true order is very low. As a result, all rules do badly in terms of the expected losses. One cannot ask rules to aggregate bad votes into something close to the true order. With three alternatives, all rules miss the true order more than 50% of the time. With six alternatives, this percentage is higher. With the naive distance, the relation between the rules is:

$$I_{98} \prec K \prec ML(1) \prec I_{94} \prec B$$

However, this relation varies when a more general distance is used and it also changes with the values of the parameters. In short, everything is more erratic in this case.

## B.5 Simulations with nine judges and $(\alpha, \beta) = (0.4, 1)$

In the preceding subsection, we saw that lowering  $\beta$  could have a non-negligible impact on the relative performances of the rules. The impact was most important for the Borda rule,

whose performance passed from one of the best to the worst when  $\beta$  was set to 0. We now examine what happens when  $\alpha$  is decreased while  $\beta$  retains the same value. We report the results of simulations with  $\alpha = 0.4$  and  $\beta = 1$ . This amounts to a lower competency for the judges. Again, the number of judges is 9 and the number of alternatives ranges from 3 to 6.

The maximum likelihood rule has been applied with  $(\hat{\alpha}, \hat{\beta}) = (\alpha, \beta) = (0.4, 1)$  but also with  $(\hat{\alpha}, \hat{\beta}) = (0.4, 3), (0.4, 0.2), (0.4, 0.6)$ . The most striking fact about the results is that, not only did the maximum likelihood rule perform better than all other rules, except when  $\eta = 2$ , but this performance appears to be robust to the choice of the parameters, as long as  $\hat{\beta} > \alpha$ . Setting  $\hat{\beta} = 3$  increases the expected loss under the maximum likelihood rule only slightly without changing the relative position of this rule with respect to others. Setting  $\hat{\beta} = 7$  produces the same expected losses as  $\hat{\beta} = 3$ . For  $\hat{\beta} < \alpha$ ,  $ML$  does worst than  $B$ . This was to be expected since  $ML$  and  $B$  are identical when  $\hat{\beta} = \alpha$ . The relation between the rules, in terms of expected losses when  $\eta \neq 2$ , is now:

$$ML(1) \prec ML(3) \prec ML(0.6) \prec B \prec ML(0.2) \prec I_{94} \prec I_{98} \prec K$$

When,  $\eta = 2$ , Borda takes the head. The reason is again that ties get a penalty of  $\frac{1}{2}$ . This is not very different from what we had with  $\alpha = 1.1, \beta = 0.9$ . The main difference is again the robustness of the maximum likelihood rule with respect to the choice of the parameters  $\hat{\alpha}$  and  $\hat{\beta}$ .

## References

- Arrow, K.J. (1963): *Social Choice and Individual Values*, second edition, New York: Wiley.
- Arrow, K.J. and H. Raynaud (1986): *Social choice and multicriterion decision-making*, Cambridge, Mass.: MIT Press.
- Bassett, W. and J. Persky (1994): "Rating Skating," *Journal of the American Statistical Association*, **89**, 1075-1079.
- Borda, J.C. de (1784): *Mémoire sur les élections au scrutin*, Histoire de L'Académie Royale des Sciences.
- Bühlmann, H. and P. Huber (1963): "Pairwise Comparison and Ranking in Tournaments", *Annals of Mathematical Statistics*, **34**, 501-510.
- 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.
- Copeland, A.H. (1951): "A Reasonable Social Welfare Function," University of Michigan Seminar on the Applications of Mathematics to Social Sciences, mimeo.
- De Groot, M.H. (1970): *Optimal Statistical Decision*, New York: McGraw-Hill.
- Drissi-Bakhkhat, M. (2002): *A Statistical Approach to the Aggregation of Votes*, Ph. D. Thesis, Université Laval, Québec.
- Drissi-Bakhkhat, M. and M. Truchon (2004), "Maximum Likelihood Approach to Vote Aggregation with Variable Probabilities," *Social Choice and Welfare*, **23**, 161-185.
- Gordon, S. and M. Truchon (2007): "Social choice, optimal inference and figure skating", *Social Choice and Welfare*, forthcoming. See also Cahier de Recherche 06-24, Centre interuniversitaire sur le risque, les politiques économiques et l'emploi (CIRPÉE).  
<http://132.203.59.36/CIRPEE/cahierscirpee/2006/2006.htm>
- 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.
- Rousseau, J.J. (1913): *The Social Contract*, London: J.M. Dent & Sons.
- Truchon, M. (2004): "Aggregation of Rankings in Figure Skating," Cahier de Recherche 0402, Département d'Économie, Université Laval.  
[www.ecn.ulaval.ca/pages/Recherche/cahiers.html](http://www.ecn.ulaval.ca/pages/Recherche/cahiers.html)
- Truchon, M. (2006): "Borda and the Maximum Likelihood Approach to Vote Aggregation," Cahier de Recherche 06-23, Centre interuniversitaire sur le risque, les politiques économiques et l'emploi (CIRPÉE).  
<http://132.203.59.36/CIRPEE/cahierscirpee/2006/2006.htm>
- Young, H.P. (1988): "Condorcet's Theory of Voting," *American Political Science Review*, **82**: 1231-1244.