Aggregation of Rankings: a Brief Review of Distance-based Rules and Loss Functions for the Expected Loss Approach

Michel Truchon

December 2005
Revised April 2007

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

Abstract:
Some researchers have addressed the problem of aggregating individual preferences or rankings by seeking a ranking that is closest to the individual rankings. Their methods differ according to the notion of distance that they use. The best known method of this sort is due to Kemeny. The first part of this paper offers a brief survey of some of these methods. Another way of approaching the aggregation of rankings is as a problem of optimal statistical inference, in which an expected loss is minimised. This approach requires a loss function, a concept closely related the notion of distance between rankings. The second part of this paper examines two classes of parametric functions and proposes one class for the optimal statistical inference problem.

Contents

List of Tables

1 Introduction

The question of how best to aggregate individual preferences or rankings is one of the oldest and best-known in the social sciences. Some researchers have addressed the issue by seeking a ranking that is 'closest' to the individual rankings. Their methods differ according to the notion of distance that they use. The best known method of this sort is due to Kemeny (1959). Other methods are due to Bogart (1975), Cook and Seiford (1978), and Cook and Seiford (1982). The first part of this paper offers a brief survey of some of these methods.

Another way of approaching the aggregation of rankings is as a problem of optimal statistical inference, in which an expected loss is minimised. This approach requires a loss function, a concept closely related to the notion of distance between rankings. The second part of this paper examines two classes of parametric functions and proposes one class for the optimal statistical inference problem.

This approach can be traced back to Condorcet (1785), whose objective was to justify the majority principle. It may be cast as follows. There is a true or objective order on the set of candidates. If a panel of judges is asked for their opinion about the true order, they may not come out with the correct answer 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.

Condorcet's rigorous formulation of this proposition is one of the earliest applications of the calculus of probability and of the maximum likelihood approach to inference. He assumed that every voter chooses the best of two alternatives with a probability larger than one half, and that this judgment is independent between pairs and voters. If the binary relation obtained by applying the simple majority rule to each pair of alternatives is consistent, then it is the solution to the problem, that is, the most probable order.

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 ranking is the unique Kemeny ranking.

Drissi and Truchon (2004) relax the assumption that the probability of comparing correctly two alternatives is the same for any pair of alternatives. They let the probability increase with the distance between two alternatives in the allegedly true ranking, thus allowing for the possibility that it may be more difficult to correctly rank two alternatives that are 'close' than two that are far apart in the true ranking. They postulate a two-parameter probability function to represent the competence of the judges and they analyze the behaviour of the resulting maximum likelihood rule as a function of these parameters. A most likely order is not necessarily a Kemeny order.

From the point of view of statistical decision theory, the maximum likelihood approach assumes implicitly that all errors with respect to the true order have the same weight. For example, if the true order on the set $\{a, b, c, d\}$ is $abcd$, this loss function says that choosing $dcba$ and $abdc$ are equally costly. Yet, it is unlikely that a decision maker would view reversing the last two candidates as being as serious an error as reversing the entire ranking. Gordon and Truchon (2007) formalise and make explicit the optimal inference problem facing the decision maker by introducing a loss function in the aggregation problem. The objective is then to find a ranking that minimises the expected loss.

The loss function is also the key concept in Truchon and Gordon (2006), who take the point of view of a decision maker who must adopt an aggregation rule that will be used over a long period. Having a loss function and 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. Truchon and Gordon perform these computations for five aggregation rules.

The distances surveyed in the first part of this paper could be used as loss functions in an expected loss approach. However they have a serious drawback: all errors of the same kind are assigned the same loss whether they concern a top ranked or a bottom ranked alternative in the reference order. Going back to the above example, $bacd$ and $abdc$ would have the same loss with respect to the true order $abcd$.

In many situations, an error on a bottom ranked alternative in the true ranking would be of less consequence than the same error on a top ranked alternative. In the second part of the paper, we define, discuss, illustrate and compare two classes of parametric functions that offer much latitude in this respect. They also allow for some forms of concavity or convexity, as one may wish. The second class, which is an extension of the Kemeny metric on the set of rankings, is the class of loss functions used by Gordon and Truchon (2007) and Truchon and Gordon (2006).

We start with the definition of the aggregation problem. We then present a survey of some distance-based aggregation rule. The two classes of loss functions are discusses in the last section.

2 The aggregation problem

Let $A = \{1, 2, \ldots, m\}$ 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 $\mathcal{L} \subset \mathcal{B}^* \subset \mathcal{B}$.[1] A complete weak order on A can be represented by a vector $r = (r_1, r_2, r_3, \ldots)$ or $x = (x_1, x_2, x_3, \ldots)$ where r_1 and x_1 are the rank of alternative 1, r_2 and x_2 the rank of 2, and so on.[2]

Let there be a set $J = \{1, 2, \ldots, n\}$ of *voters* or *judges.* Each judge j has a weak order $x^j \in \mathcal{R}$, also called a *vote*, on the set A. Equivalently, a vote x^j can be represented by an $(m \times m)$ binary matrix $X^j = \left[x^j_{st} \right]_{s,t \in A}$ where:

$$x^j_{st} = \begin{cases} 1 & \text{if } s \neq t \text{ and } x^j_s \leq x^j_t \\ 0 & \text{otherwise} \end{cases}$$

Conversely, given a binary representation X^j of a weak order, we get the representation x^j by setting $x^j_s = m - \sum_{t=1}^{m} x^j_{st}$. We shall use the two representations interchangeably.[3]

A *profile of votes* is an array $X = (x^1, \ldots, x^n) \in \mathcal{R}^n$. A profile may also be written in the binary form $X = (X^1, \ldots, X^n)$. 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 rule* is a correspondence $\Gamma : \mathcal{R}^n \to \mathcal{R}$ that assigns to each profile R, a final ranking or a subset of final rankings $\Gamma(X)$ of the alternatives. $\Gamma_s(X)$ represents the rank of alternative s in the final ranking $\Gamma(X)$.

---

[1] 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 $\mathcal{B}^*$.

[2] Throughout, we shall use r to represent the true order on A and x to represent a vote on A.

[3] 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 $1, 2, 3, \ldots$ is also an abuse of notation but this allows for simpler notations in the remaining of the paper.

Before embarking on the presentation of genuine aggregation rules, let us consider the correspondence $M : \mathcal{R}^n \rightarrow \mathcal{B}$ defined by $s\, M(X)\, t \Leftrightarrow \sum_{j=1}^{n} x_{st}^j > \sum_{j=1}^{n} x_{ts}^j$. The binary relation $M(X)$ is the *majority relation* issued from profile X. This relation is not necessarily transitive: it may contain cycles. It is not necessarily complete either except when n is odd. Thus, M is not an aggregation rule. Yet, if $M(X)$ is a linear order, it is of special interest.

Definition 2 Given a profile X, if $M(X)$ is a linear order, it is called the *Condorcet ranking*.

Although M is not necessarily an aggregation rule, it is often considered as desirable that an aggregation rule gives a result that agrees with the Condorcet ranking whenever possible. This prompts the following definition.

Definition 3 An aggregation rule $\Gamma : \mathcal{R}^n \rightarrow \mathcal{R}$ has the *Condorcet property* or is *Condorcet consistent* if $\Gamma(X) = M(X)$ for every profile X such that $M(X)$ is a linear order.

3   Distance-based aggregation rules

Suppose that we are given a *metric d* on $\mathcal{R}$, that is, a function $d : \mathcal{R}^2 \rightarrow \mathbb{R}$ satisfying the three following axioms:

Axiom 1 *Given $r, \hat{r} \in \mathcal{R}$, $d(r, \hat{r}) \geq 0$ and $d(r, \hat{r}) = 0$ if $r = \hat{r}$.*

Axiom 2 *Given $r, \hat{r} \in \mathcal{R}$, $d(r, \hat{r}) = d(\hat{r}, r)$.*

Axiom 3 *Given $r, \tilde{r}, \hat{r} \in \mathcal{R}$, $d(r, \hat{r}) \leq d(r, \tilde{r}) + d(\tilde{r}, \hat{r})$.*

Given d, let us define the function $\delta : \mathcal{R}^{n+1} \rightarrow \mathbb{R}$ by:

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

$\delta(r, X)$ is a "distance" between profile X and the weak order r. Given a metric d on $\mathcal{R}$ and its derived "distance" δ between orders and profiles, we can define an aggregation rule $\Gamma^\delta : \mathcal{R}^n \rightarrow \mathcal{R}$ by:

$$\Gamma^\delta(X) = \arg\min_{r \in \mathcal{R}} \delta(r, X)$$

Several rules of this sort have been proposed in the literature. One of them is the well known Kemeny rule.

3.1 The Kemeny Rule

The first metric for weak orders has been proposed by Kemeny and Snell (1962). They derive this metric from the following axioms, in addition to the usual axioms of a metric. Kemeny (1959) gave a summary of the results of Kemeny and Snell before the publication of the book. One of their axioms involves the following notion of betweenness.

Definition 4 (Kemeny-Snell) Given $r, \tilde{r}, \hat{r} \in \mathcal{R}$, one says that $\tilde{r}$ is *between* r and $\hat{r}$ and one writes $[r, \tilde{r}, \hat{r}]$ if $\forall s, t \in A : \tilde{r}_s < \tilde{r}_t \Rightarrow r_s < r_t$ or $\hat{r}_s < \hat{r}_t$ and $\tilde{r}_s = \tilde{r}_t \Rightarrow (r_s - r_t)(\hat{r}_s - \hat{r}_t) \leq 0$.

Axiom 4 *Given $r, \tilde{r}, \hat{r} \in \mathcal{R}$, $d(r, \hat{r}) = d(r, \tilde{r}) + d(\tilde{r}, \hat{r})$ whenever $[r, \tilde{r}, \hat{r}]$.*

Axiom 5 *$d(\sigma(r), \sigma(\hat{r})) = d(r, \hat{r})$ for every permutation σ of the elements of r and $\hat{r}$.*

Axiom 6 *If an alternative is added to A and if this alternative is ranked first or last relative to both r and $\hat{r}$, so that r and $\hat{r}$ become r^* and $\hat{r}^*$ respectively, then,[4]*

$$d(r^*, \hat{r}^*) = d(r, \hat{r}).$$

Axiom 7 $\min\limits_{r, \hat{r} \in \mathcal{R}} d(r, \hat{r}) \in \{0, 1\}$, *that is, the minimum positive distance is 1.*

Kemeny and Snell show that there is a unique metric, satisfying Axioms 1 to 7. This is the Kemeny metric defined as follows. Let $\gamma_{st} : \mathcal{R}^2 \to \mathbb{R}$, be a function defined for every couple of alternatives (s, t) and every pair of weak orders $(r, \hat{r})$ by:

$$\gamma_{st}(r, \hat{r}) = \begin{cases} 2 & \text{if } r_s < r_t \text{ and } \hat{r}_s > \hat{r}_t \\ 1 & \text{if } r_s < r_t \text{ and } \hat{r}_s = \hat{r}_t \\ 1 & \text{if } r_s = r_t \text{ and } \hat{r}_s > \hat{r}_t \\ 0 & \text{otherwise} \end{cases}$$

Then, the *Kemeny metric* on $\mathcal{R}$ is the function $d^K : \mathcal{R}^2 \to \mathbb{R}$ defined by:

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

Using this metric, we get the Kemeny "distance" δ^K between a weak order r and a profile X:

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

---

[4]This formulation is from Cook and Seiford (1978) (See the next subsection). Given the invariance to permutations, it is equivalent to the formulation of Kemeny and Snell.

Kemeny 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 call the solutions to the two minimization problem *median rankings* and *mean rankings* respectively.[5]

Nowadays, the name of Kemeny is associated with the first problem, the one that consists in minimizing the total number of disagreements with the rankings of the voters. This is the Kemeny rule.

Definition 5 The *Kemeny rule* is the correspondence $K : \mathcal{R}^n \to \mathcal{L}$ 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 the *Kemeny orders*.

Remark 1 Bogart (1973 and 1975) extends the characterization of d^K by Kemeny and Snell to partial and non transitive binary relations. Ali, Cook, and Kress (1986) gave a similar characterization eleven years later.

Remark 2 Bogart (1975) points out the well known fact that if the majority relation $M(X)$ is a linear order, then, it is the unique Kemeny order. This suggests that if $M(X)$ contains cycles, then, the solution to[6]

$$\min_{r \in \mathcal{L}} d^K(r, M(X))$$

might be a Kemeny order. Bogart shows that this is not the case.

Example 1 Let $A = \{a, b, c, d\}$ and suppose there are 10 voters with the linear order $bdac$, 10 others with $cdab$ and 1 with $abcd$. Then, $M(X) = \{(a, b), (a, c), (b, c), (b, d), (c, d), (d, a)\}$. $M(X)$ gives the cycle $abcda$. The closest linear order to $M(X)$ according to d^K is $abcd$. Yet, this is not the closest linear order to the profile, that is, it does not minimize $\sum_{j=1}^{n} d^K(r, x^j)$. Indeed, $d^K(abcd, M(X)) = 2$. However, $\sum_{j=1}^{n} d^K(abcd, x^j) = 140$ while $\sum_{j=1}^{n} d^K(r, x^j) = 106$ for $r = dabc$, $bdac$, and $bcda$. These three linear orders are Kemeny orders.

---

[5] To be more precise, given a profile R, they define them as the median and the mean of the set $\{r^1, \ldots, r^n\}$.

[6] Here, r and $M(R)$ must be seen as subsets of X^2 and the definition of γ_{st} changed for:

$$\gamma_{st}(r, \hat{r}) = \begin{cases} 2 & \text{if } (s, t) \in r \text{ and } (t, s) \in \hat{r} \\ 1 & \text{if } (s, t) \in r \text{ and } (s, t), (t, s) \in \hat{r} \\ 1 & \text{if } (s, t), (t, s) \in r \text{ and } (t, s) \in \hat{r} \\ 0 & \text{otherwise} \end{cases}$$

The function d^K is an l^1 metric. With a different set of axioms, Bogart (1975) obtains an l^2 metric on $\mathcal{R}$, defined by:

$$d^{2K}(r, \hat{r}) = \frac{1}{\sqrt{2}} \left(\sum_{s \in A} \sum_{t \in A} (\gamma_{st}(r, \hat{r}))^2 \right)^{\frac{1}{2}}$$

Again, one obtains a corresponding "distance" δ^{2K} from this metric. Bogart calls the elements of $\arg \min_{r \in \mathcal{R}} \delta^{2K}(r, X)$ "mean rankings" but these rankings may be different from those of Kemeny and Snell (1962) since they are obtained with respect to a different distance.[7]

3.2   A Median Ranking Rule

Using essentially the same axioms as those of Kemeny and Snell (1962), Cook and Seiford (1978) derive another l^1 metric on $\mathcal{R}$. The reason for obtaining a different metric for apparently the same set of axioms is their use of a different definition of betweenness. Thus, Axiom 4 has not the same meaning in both papers.

Definition 6 (Cook-Seiford) Given $r, \tilde{r}, \hat{r} \in \mathcal{R}$, on says that $\tilde{r}$ is *between* r and $\hat{r}$ and one writes $[r, \tilde{r}, \hat{r}]$ if $\forall s : r_s \leq \tilde{r}_s \leq \hat{r}_s$ or $r_s \geq \tilde{r}_s \geq \hat{r}_s$.

Moreover, they use the following convention, borrowed from Kendall (1970), to represent weak orders. If two alternatives s and t tie for say rank k, then $r_s = r_t = k + \frac{1}{2}$. More generally, if κ alternatives $s_1, \ldots, s_\kappa$ tie for say rank k, then $r_{s_1} = \cdots = r_{s_\kappa} = k + \frac{\kappa - 1}{2}$ and the rank of the next alternative not involved in this tie is $k + \kappa$. With this convention, we always have $\sum_{s=1}^m r_s = \frac{m(m+1)}{2}$. For example, if there are only 3 alternatives and they all tie, then $r_1 = r_2 = r_3 = 2$.

Cook and Seiford show that the unique metric satisfying Axioms 1-7 is the function $d^{CS} : \mathcal{R}^2 \to \mathbb{R}$ defined by:

$$d^{CS}(r, \hat{r}) = \sum_{s=1}^m |r_s - \hat{r}_s|$$

---

[7]Indeed, without loss of generality, we could as well define

$$d^{2K}(\hat{r}, r) = \sum_{s \in A} \sum_{t \in A} (\gamma_{st}(\hat{r}, r))^2$$

which must not be confounded with $\left(\sum_{s \in A} \sum_{t \in A} \gamma_{st}(\hat{r}, r) \right)^2$ as in Kemeny and Snell (1962).

Using d^{CS}, they then get the "distance" δ^{CS} between weak orders and profiles:

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

Then, given a profile X, *median rankings,* with respect to d^{CS}, are the elements of:

$$MED(X) = \arg\min_{r \in \mathcal{R}} \delta^{CS}(r, X)$$

Thus, a median ranking, with respect to d^{CS}, minimizes the sum of the absolute differences with the ranks in the x^j.

Remark 3 The metric d^{CS} is quite different in spirit from d^K. The basic ingredient in d^{CS} is the absolute difference in ranks while in d^K it is the number of reversals in the position of s with respect to all other alternatives in $\hat{r}$ as compared to its position in r ($|r_s - \hat{r}_s|$ versus $\sum_{s \in A} \gamma_{st}(r, \hat{r})$). Examples 2 and 3 below illustrate this difference. Their first objective is however to show that the definitions of betweenness of Kemeny-Snell and Cook-Seiford are completely independent from one another. Example 4 further illustrates the difference between d^K and d^{CS}.[8]

Example 2 Let:

$$
\begin{aligned}
r &= (3, 2, 1, 4, 5, 6) \\
\tilde{r} &= (2, 3, 1, 4, 5, 6) \\
\hat{r} &= (3, 4, 5, 6, 2, 1)
\end{aligned}
$$

Then, $[r, \tilde{r}, \hat{r}]$ according to Kemeny and Snell but not according to Cook and Seiford. We have $d^K(r, \hat{r}) = 24 = 2 + 22 = d^K(r, \tilde{r}) + d^K(\tilde{r}, \hat{r})$ as requested by Axiom 4. However, it is not requested that $d^{CS}(r, \hat{r}) = d^{CS}(r, \tilde{r}) + d^{CS}(\tilde{r}, \hat{r})$ and, as a matter of fact, we have $d^{CS}(r, \hat{r}) = 16 < 2 + 16 = d^{CS}(r, \tilde{r}) + d^{CS}(\tilde{r}, \hat{r})$.

Example 3 Let:

$$
\begin{aligned}
r &= (3, 2, 1, 4, 5, 6) \\
\tilde{r} &= (3, 4, 1, 2, 5, 6) \\
\hat{r} &= (6, 5, 1, 2, 3, 4)
\end{aligned}
$$

---

[8]Curiously, Cook and Seiford do not say a word on this difference and the reason for it.

Then, $[r, \tilde{r}, \hat{r}]$ according to Cook and Seiford but not according to Kemeny and Snell. Indeed, $\tilde{r}_1 < \tilde{r}_2$, $r_1 > r_2$, and $\hat{r}_1 > \hat{r}_2$. We have $d^{CS}(r, \hat{r}) = 12 = 4 + 8 = d^{CS}(r, \tilde{r}) + d^{CS}(\tilde{r}, \hat{r})$ as requested by Axiom 4. However, it is not requested that $d^K(r, \hat{r}) = d^K(r, \tilde{r}) + d^K(\tilde{r}, \hat{r})$ and, as a matter of fact, we have $d^K(r, \hat{r}) = 12 < 6 + 10 = d^K(r, \tilde{r}) + d^K(\tilde{r}, \hat{r})$.

In Example 3, removing alternative 3 and rescaling or not the ranks of the other alternatives would change none of the distances computed with d^{CS} and d^K. This is in accordance with Axiom 6.

Example 4 Let $r = (1, 2, 3)$, $\tilde{r} = (3, 1, 2)$ and $\hat{r} = (3, 2, 1)$. Note that $[r, \tilde{r}, \hat{r}]$ according to Kemeny and Snell but not according to Cook and Seiford. Then, $d^{CS}(r, \tilde{r}) = 4$, $d^{CS}(r, \hat{r}) = 4$ and $d^{CS}(\tilde{r}, \hat{r}) = 2$; $d^K(r, \tilde{r}) = 4$, $d^K(r, \hat{r}) = 6$ and $d^K(\tilde{r}, \hat{r}) = 2$. Note that $d^{CS}(r, \tilde{r}) = d^{CS}(r, \hat{r})$ but $d^K(r, \tilde{r}) < d^K(r, \hat{r})$.[9]

3.3 The Minimum Variance Rule

Cook and Seiford (1982) define an l^2 metric on $\mathcal{R}$ by:

$$d^2(r, \hat{r}) = \sum_{s=1}^{m} (r_s - \hat{r}_s)^2$$

This yields

$$\delta^2(r, X) = \sum_{j=1}^{n} d^2(r, x^j)$$

and another aggregation rule, which Cook and Seiford (1982) call the *Minimum Variance Method*:

$$MV(X) = \arg\min_{r \in \mathcal{R}} \delta^2(r, X)$$

Members of $MV(X)$ are called *mean rankings*, with respect to d^2.

Remark 4 Let $\bar{b}(s) = \dfrac{1}{n} \sum_{j=1}^{n} x_s^j$. A Borda ranking can be defined as a weak order r such that:

$$\forall s, t \in A : r_s \leq r_t \Leftrightarrow \bar{b}(s) \leq \bar{b}(t)$$

In plain words, a Borda ranking is defined from the average ranks given by the voters. In statistics, it is well known that the sample mean is the point estimate that minimizes the

---

[9]We leave open the question as to if it is possible to have $d^{CS}(r, \tilde{r}) < d^{CS}(r, \hat{r})$ and $d^K(r, \tilde{r}) < d^K(r, \hat{r})$.

squared errors. This led Kendall (1970, p. 101 and 114) to claim and "show" that a Borda ranking is actually a mean ranking as defined above. Using geometry, Cook and Seiford (1982) show that Kendall's claim and proof is true only if ties are ruled out, that is, if $\mathcal{R}$ is replaced by $\mathcal{L}$. This can be seen with the following example.

Example 5 Let $m = 3$, $n = 3$, $r^1 = (2, 3, 1)$, $r^2 = (3, 2, 1)$, and $r^3 = \left(\dfrac{3}{2}, \dfrac{3}{2}, 3\right)$. Thus, $\bar{b} = \left(\dfrac{13}{6}, \dfrac{13}{6}, \dfrac{5}{3}\right)$ and the Borda ranking is $r^* = (2.5, 2.5, 1)$. However, $MV(X) = \{\tilde{r}\} = \{(2, 2, 2)\}$. Indeed, $\sum_{s=1}^{m} \left(\bar{b}(s) - r_s^*\right)^2 = \dfrac{4}{6}$ while $\sum_{s=1}^{m} \left(\bar{b}(s) - \tilde{r}_s\right)^2 = \dfrac{1}{6}$. [10]

4   From metrics to loss functions

A metric may be used as a loss function in an expected loss approach. That is, the number $d(r, \hat{r})$ may be seen as the loss resulting from the selection of the ranking $\hat{r}$ when r is the true ranking. Then, given a profile X, a posterior distribution function $\pi(\cdot \mid X)$ on the set $\mathcal{R}$, the *expected loss* resulting from the selection of a ranking $\hat{r}$ is given by: $\sum_{r \in \mathcal{R}} \pi(r \mid X) d(r, \hat{r})$. An optimal ranking minimizes this expected loss. The aggregation rule that yields these optimal rankings has more statistical flavour than the ones reviewed in the previous section. Such a rule is defined in Gordon and Truchon (2007) and given the name *minimum expected loss* (MEL) rule.

Another question of interest is how different ranking rules Γ, taken form a set $\mathcal{G}$, could fare in repeated uses. In other words, what is the *ex ante expected loss* or the *risk* of these rules? This is the question addressed in Truchon and Gordon (2006). Instead of starting with a given profile and the associated posterior distribution of the true rankings, they take as given the true ranking r and its associated distribution, say $f(\cdot \mid r)$, of the profiles $X \in \mathcal{L}^n$ and compute the risk $\sum_{X \in \mathcal{L}^n} f(X \mid r) d(r, \Gamma(X))$ of each aggregation rule $\Gamma \in \mathcal{G}$.

Two types of metric on $\mathcal{R}$ have been presented in the previous section. Their basic ingredient are respectively the terms $|\hat{r}_s - r_s|$ and $\sum_{t \in A} \gamma_{st}(r, \hat{r})$, $s \in A$. The latter is the number of reversals in the position of s in $\hat{r}$ with respect to other alternatives ranked after s in r.[11] Any of these metrics could be used in an expected loss approach. However, they

---

[10]r^* is also the unique ranking satisfying the Extended Condorcet Criterion defined in Truchon (2004). Thus, the Minimum Variance Method does not satisfy this criterion.

[11]Note that if $r = (1, 2, 3, 4)$ is the true ranking on $\{a, b, c, d\}$, then $\sum_{t \in \{a,b,c,d\}} \gamma_{2t}(r, \hat{r}) = 2$ with $\hat{r} = (3, 4, 2, 1)$ as well as with $\hat{r} = (4, 3, 2, 1)$. Thus, the fact that alternatives a and b are better ranked in the

all have a drawback for this purpose. All errors of the same kind are assigned the same loss whether they concern a top ranked or a bottom ranked alternative in the reference order.

In many situations, an error on a bottom ranked alternative in the true ranking would be of less consequence than the same error on a top ranked alternative. For example, if $abcd$ is the true ranking on $A = \{a, b, c, d\}$ then, an error on alternative a represents a greater loss than the same error on alternative b, and an error on alternative b represents a greater loss than the same error on alternative c, and so on. Thus, a good loss function should be decreasing with respect to r_s itself, that is, the rank of alternative s in the true ranking, for a fixed value of $|\hat{r}_s - r_s|$ or $\sum_{t \in A} \gamma_{st}(r, \hat{r})$. In addition to being increasing with respect to $|\hat{r}_s - r_s|$ or $\sum_{t \in A} \gamma_{st}(r, \hat{r})$ and decreasing with respect to r_s, we may wish that d be concave, that is, increase at a decreasing rate, with respect to $|\hat{r}_s - r_s|$ or $\sum_{t \in A} \gamma_{st}(r, \hat{r})$, or alternatively be convex. We may have similar desiderata for the change with respect to r_s.

We define two classes of parametric functions that offer much latitude in this respect. However, they are not necessarily metrics on $\mathcal{R}$. The first is the class of functions $d_{\eta\theta} : \mathcal{R}^2 \to \mathbb{R}_+$ defined by:

$$d_{\eta\theta}(r, \hat{r}) = \sum_{s \in A} |\hat{r}_s - r_s|^\eta (m - r_s + 1)^\theta, \ \eta \geq 0, \ \theta \geq 0$$

The second class is given by the functions $d_{\eta\theta}^K$ defined as follows:

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

With these functions, the rate at which the loss increases with the term $|\hat{r}_s - r_s|$ or the partial sum $\sum_{s \in A} \gamma_{st}(r, \hat{r})$ is controlled by the parameter η. Moreover, these numbers are weighted by the term $(m - r_s + 1)^\theta$, which takes into account the place in the true ranking where an error occurs.

The members of the first class are concave or convex with respect to $|\hat{r}_s - r_s|$ if $\eta < 1$ or $\eta > 1$ respectively. They are concave or convex with respect to r_s if $\theta < 1$ or $\theta > 1$ respectively. We also know that these functions are quasiconcave. An increase in $|\hat{r}_s - r_s|$ may be compensated by an increase in r_s but the rate at which this compensation takes place diminishes with the size of $|\hat{r}_s - r_s|$.[12] The same remark, with $|\hat{r}_s - r_s|$ replaced by $\sum_{s \in A} \gamma_{st}(r, \hat{r})$, applies to the members of the second class.

---

first $\hat{r}$ while their order is reversed in the second is not taken into account in this partial sum. However, it will be accounted for in the complete sum.

[12] Of course, this language is completely correct with continuous instead of discrete arguments.

Remark 5 The Kemeny metric is a particular case of $d_{\eta\theta}^K$. Indeed, $d^K = d_{1,0}^K$. With the convention $0^0 = 0$, the naive function 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 $\kappa > 0$, which is the implicit loss function in the maximum likelihood approach, may be seen as a particular case of both $d_{\eta\theta}$ and $d_{\eta\theta}^K$, namely:

$$d^N = d_{00} = d_{00}^K$$

Remark 6   • All members of these two classes satisfy Axiom 5 of a Kemeny metric defined in the previous section. Indeed, let σ be a permutation of the elements of r and $\hat{r}$. Clearly,

$$\sum_{s=1}^{m} |\hat{r}_s - r_s|^\eta (m - r_s + 1)^\theta = \sum_{s=1}^{m} |\hat{r}_{\sigma(s)} - r_{\sigma(s)}|^\eta (m - r_{\sigma(s)} + 1)^\theta$$

and similarly for $d_{\eta\theta}^K$.

• It is easy to check that they also satisfy Axiom 6.

• On the other hand, Axiom 2 of a metric does not hold, except when $\theta = 0$, since interchanging r and $\hat{r}$ change the term $(m - r_s + 1)$.

• Axiom 3 is akin to subadditivity, which is implied by concavity. Thus, both $d_{\eta\theta}$ and $d_{\eta\theta}^K$ satisfy Axiom 3 when $\theta = 0$ and $\eta \leq 1$. However, Axiom 3 is violated for $\theta = 0$ and $\eta > 1$. For $\theta > 0$, the picture is more complex. Thus, it may well be that Axiom 3 is violated even with $\eta \leq 1$.

• Axiom 4 does not hold even when $\theta = 0$ and $\eta \leq 1$. Finally, there is no assurance that the minimum positive distance is 1. Thus, Axiom 7 does not hold. Of course, this normalization could be restored but it is not important.

Remark 7 The metrics d^{SC} and d^2 defined in sections 3.2 and 3.3 are particular cases of the above functions: $d^{SC} = d_{10}$, $d^2 = d_{20}$. However, d^{2K} is not a particular case of $d_{\eta\theta}^K$ since the exponent in d^{2K} applies to $(2\gamma_{st}(r, \hat{r}))$ and not to the partial sum.

Tables 3 to 8 of the Appendix illustrate the behavior of $d_{\eta\theta}$ and $d_{\eta\theta}^K$ for the case of four alternatives and for twelve pairs of values of the parameters, those corresponding to the cases of Table 2. The true order is taken as $(1, 2, 3, 4)$. Without loss of generality, the

distance between any order and the true order is divided by the maximum distance. Thus, the distance range is always $[0, 1]$, which facilitates the comparisons. For each pair of values of the parameters, the orders have been ranked in order of increasing distance with respect to the true order.

With $d_{\eta\theta}^K$, the maximum distance is always attained for $(4, 3, 2, 1)$. Actually, this is true for any number of alternatives. That is, if $(1, \ldots, m)$ is the true order, then the maximum distance is always attained for $(m, \ldots, 1)$, whatever the values of the parameters. Indeed, if $(1, \ldots, m)$ is the true order, then $\sum_{t \neq s} \gamma_{st}(r, \hat{r}) = \sum_{t > s} \gamma_{st}(r, \hat{r})$ and for any s, the last term is maximal for $\hat{r} = (m, \ldots, 1)$. The picture is different with $d_{\eta\theta}$ as can be seen from Tables 3 to 5.

We shall argue against the adoption of $d_{\eta\theta}$. First, for $\eta = \frac{1}{2}$, it can be noted that simply inverting the last two alternatives or the second and the third alternatives in the true ranking produces a loss larger than 0.21. This is too large a value, compared to the largest loss of 1. In addition, $(4, 3, 2, 1)$ is not the error that receives the maximum penalty. Thus, we believe that this function, with $\eta = \frac{1}{2}$, should be discarded. It should also be discarded for the other values of the parameters for which $(4, 3, 2, 1)$ does not produce the largest loss, that is, for $(\eta, \theta) = \left(1, \frac{1}{2}\right), (1, 1), (1, 2), (2, 1), (2, 2)$. For $(\eta, \theta) = (1, 0), (2, 0)$, the function is not very discriminating. It produces many ties.

For $\left(2, \frac{1}{2}\right)$, the functions behaves strangely. For example, let $\{a, b, c, d\}$ be the set of alternatives and $abcd$ be the true order and consider the four orders in Table 1, together with their distance from the true one, according to $d_{2, \frac{1}{2}}$ and $d_{2, \frac{1}{2}}^K$. They are taken from Tables 5 and 8 respectively. With $d_{2, \frac{1}{2}}$, it appears that inverting c and d is worst than inverting a and b. The function $d_{2, \frac{1}{2}}^K$ has a better behaviour in this respect.

	r	$d_{2, \frac{1}{2}}$	$d_{2, \frac{1}{2}}^K$
$cabd$	$(2, 3, 1, 4)$	0.311	0.142
$cbad$	$(3, 2, 1, 4)$	0.453	0.369
$dabc$	$(2, 3, 4, 1)$	0.469	0.195
$dbac$	$(3, 2, 4, 1)$	0.611	0.423

Table 1: Comparison of $d_{2, \frac{1}{2}}$ and $d_{2, \frac{1}{2}}^K$

There are other anomalies of this kind with $d_{2, \frac{1}{2}}$. For all these reasons, it seems preferable to use $d_{\eta\theta}^K$ in the expected loss approach. One can always find some strange behavior with

this function as for $d_{\eta\theta}$. Yet, as a whole, $d^K_{\eta\theta}$ behaves better than $d_{\eta\theta}$, at least for some values of the parameters.

Table 2 suggests eight pairs of values, indicated by an X, that could be appropriate in the expected loss approach. They were retained in Truchon and Gordon (2006) together with $(0,0)$. Table 2 includes $(\eta,\theta) = (1,0)$, because these values correspond to the Kemeny metric. Note however that $d^K_{\eta\theta}$ is not very discriminating for $\theta = 0$, producing many ties. Thus, $(\eta,\theta) = (2,0)$ is not among the suggestions. For $\eta = \frac{1}{2}$, simply inverting the last two alternatives or the second and the third alternatives in the true ranking produces a loss larger than 0.2. We have objected to the use of $d_{\eta\theta}$ with $\eta = \frac{1}{2}$ for this reason. Yet, retaining $d^K_{\eta\theta}$ with the pair $\left(\frac{1}{2}, 1\right)$ could be interesting for the sake of comparisons.

Increasing η above 1, for a given value of θ, reduces the relative cost of making a particular inversion error. This is because increasing η increases the cost of completely reversing the true order proportionally more than is the case with any other error. Thus, values of η greater than 2 are clearly implausible; the costs of making a mistake at the top end of the ranking are higher than what large values of η would seem to imply. Maybe that even $\eta = 2$ is too high; Gordon and Truchon (2007) did not retain this value.

Low values of θ are also probably less plausible in the expected loss approach. Intermediate values for these parameters, for example $\frac{1}{4}$, $\frac{3}{2}$, are probably not necessary to verify the robustness of any result to the specification of the parameters.

		θ			
		0	$\frac{1}{2}$	1	2
η	$\frac{1}{2}$			X	
	1	X	X	X	X
	2		X	X	X

Table 2: Suggested values for the parameters (indicated by an X)

References

Ali, I., Cook, W.D., and M. Kress (1986): "On the Minimum Violations Ranking of a Tournament," *Management Science,* **32**, 660-972.

Arrow, K.J. (1963): *Social Choice and Individual Values*, second edition, New York: Wiley.

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

Bogart, K.P. (1973): "Preference Structures I: Distances Between Transitive Preference Relations," *Journal of Mathematical Sociology,* **3**, 49-67.

Bogart, K.P. (1975): "Preference Structures II: Distance Between Asymmetric Relations," *SIAM Journal of Applied Mathematics,* **29**, 254-262.

Cook, W.D. and L.M. Seiford (1978): "Priority Ranking and Consensus Information," *Management Science,* 24, 1721-1732.

Cook, W.D. and L.M. Seiford (1982): "On the Borda-Kendall Consensus Method for Priority Ranking Problems," *Management Science*; **28**, 621-637.

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.

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.

Kendall, M.G. (1970): *Rank Correlation Methods,* fourth edition, London: Griffin.

Truchon, M. (2004): "Aggregation of Rankings in Figure Skating," Cahier de Recherche, 0402, Département d'Économique, Université Laval.

Truchon, M. and S. Gordon (2006): "Statistical Comparison of Aggregation Rules for Votes," Cahier de Recherche 06-25, Centre interuniversitaire sur le risque, les politiques économiques et l'emploi (CIRPÉE). http://132.203.59.36/CIRPEE/cahierscirpee/2006/2006.htm

Appendix: Behavior of loss functions

$\eta = \frac{1}{2}$ Order	$\theta = 0$ $d_{\eta\theta}$
$(1,2,3,4)$	0.000
$(1,2,4,3)$	0.354
$(1,3,2,4)$	0.354
$(2,1,3,4)$	0.354
$(1,4,3,2)$	0.500
$(3,2,1,4)$	0.500
$(1,3,4,2)$	0.604
$(1,4,2,3)$	0.604
$(2,3,1,4)$	0.604
$(3,1,2,4)$	0.604
$(4,2,3,1)$	0.612
$(2,1,4,3)$	0.707
$(2,4,3,1)$	0.733
$(3,2,4,1)$	0.733
$(4,1,3,2)$	0.733
$(4,2,1,3)$	0.733
$(2,3,4,1)$	0.837
$(4,1,2,3)$	0.837
$(2,4,1,3)$	0.854
$(3,1,4,2)$	0.854
$(4,3,2,1)$	0.966
$(3,4,2,1)$	0.983
$(4,3,1,2)$	0.983
$(3,4,1,2)$	1.000

$\eta = \frac{1}{2}$ Order	$\theta = \frac{1}{2}$ $d_{\eta\theta}$
$(1,2,3,4)$	0.000
$(1,2,4,3)$	0.278
$(1,3,2,4)$	0.362
$(2,1,3,4)$	0.429
$(1,4,3,2)$	0.445
$(1,3,4,2)$	0.525
$(3,2,1,4)$	0.555
$(1,4,2,3)$	0.560
$(4,2,3,1)$	0.598
$(2,3,1,4)$	0.659
$(3,1,2,4)$	0.687
$(3,2,4,1)$	0.687
$(2,1,4,3)$	0.707
$(2,4,3,1)$	0.711
$(4,2,1,3)$	0.744
$(4,1,3,2)$	0.761
$(2,3,4,1)$	0.791
$(3,1,4,2)$	0.850
$(2,4,1,3)$	0.857
$(4,1,2,3)$	0.876
$(4,3,2,1)$	0.960
$(3,4,2,1)$	0.969
$(4,3,1,2)$	0.991
$(3,4,1,2)$	1.000

$\eta = \frac{1}{2}$ Order	$\theta = 1$ $d_{\eta\theta}$
$(1,2,3,4)$	0.000
$(1,2,4,3)$	0.212
$(1,3,2,4)$	0.353
$(1,4,3,2)$	0.399
$(1,3,4,2)$	0.453
$(2,1,3,4)$	0.494
$(1,4,2,3)$	0.511
$(3,2,1,4)$	0.599
$(4,2,3,1)$	0.611
$(3,2,4,1)$	0.663
$(2,3,1,4)$	0.694
$(2,4,3,1)$	0.704
$(2,1,4,3)$	0.706
$(3,1,2,4)$	0.752
$(2,3,4,1)$	0.757
$(4,2,1,3)$	0.759
$(4,1,3,2)$	0.800
$(2,4,1,3)$	0.852
$(3,1,4,2)$	0.852
$(4,1,2,3)$	0.912
$(3,4,2,1)$	0.962
$(4,3,2,1)$	0.964
$(3,4,1,2)$	0.998
$(4,3,1,2)$	1.000

$\eta = \frac{1}{2}$ Order	$\theta = 2$ $d_{\eta\theta}$
$(1,2,3,4)$	0.000
$(1,2,4,3)$	0.114
$(1,3,2,4)$	0.297
$(1,4,3,2)$	0.323
$(1,3,4,2)$	0.329
$(1,4,2,3)$	0.405
$(2,1,3,4)$	0.571
$(3,2,1,4)$	0.646
$(3,2,4,1)$	0.648
$(4,2,3,1)$	0.673
$(2,1,4,3)$	0.685
$(2,4,3,1)$	0.696
$(2,3,1,4)$	0.700
$(2,3,4,1)$	0.702
$(4,2,1,3)$	0.785
$(2,4,1,3)$	0.808
$(3,1,2,4)$	0.814
$(3,1,4,2)$	0.846
$(4,1,3,2)$	0.871
$(3,4,2,1)$	0.938
$(4,1,2,3)$	0.953
$(3,4,1,2)$	0.969
$(4,3,2,1)$	0.969
$(4,3,1,2)$	1.000

Table 3: Distance from the true order with the loss function $d_{\eta\theta}$

$\eta = 1$ $\theta = 0$		$\eta = 1$ $\theta = \frac{1}{2}$		$\eta = 1$ $\theta = 1$		$\eta = 1$ $\theta = 2$	
Order	$d_{\eta\theta}$	Order	$d_{\eta\theta}$	Order	$d_{\eta\theta}$	Order	$d_{\eta\theta}$
$(1,2,3,4)$	0.00	$(1,2,3,4)$	0.000	$(1,2,3,4)$	0.000	$(1,2,3,4)$	0.000
$(1,2,4,3)$	0.25	$(1,2,4,3)$	0.192	$(1,2,4,3)$	0.143	$(1,2,4,3)$	0.075
$(1,3,2,4)$	0.25	$(1,3,2,4)$	0.250	$(1,3,2,4)$	0.238	$(1,3,2,4)$	0.194
$(2,1,3,4)$	0.25	$(2,1,3,4)$	0.297	$(1,3,4,2)$	0.333	$(1,3,4,2)$	0.224
$(1,3,4,2)$	0.50	$(1,3,4,2)$	0.410	$(2,1,3,4)$	0.333	$(1,4,3,2)$	0.299
$(1,4,2,3)$	0.50	$(1,4,3,2)$	0.435	$(1,4,3,2)$	0.381	$(1,4,2,3)$	0.343
$(1,4,3,2)$	0.50	$(1,4,2,3)$	0.468	$(1,4,2,3)$	0.429	$(2,1,3,4)$	0.373
$(2,1,4,3)$	0.50	$(2,1,4,3)$	0.489	$(2,1,4,3)$	0.476	$(2,1,4,3)$	0.448
$(2,3,1,4)$	0.50	$(2,3,1,4)$	0.522	$(2,3,1,4)$	0.524	$(2,3,4,1)$	0.478
$(3,1,2,4)$	0.50	$(3,2,1,4)$	0.544	$(2,3,4,1)$	0.571	$(2,3,1,4)$	0.493
$(3,2,1,4)$	0.50	$(3,1,2,4)$	0.569	$(3,2,1,4)$	0.571	$(2,4,3,1)$	0.552
$(2,3,4,1)$	0.75	$(2,3,4,1)$	0.649	$(2,4,3,1)$	0.619	$(3,2,4,1)$	0.582
$(2,4,1,3)$	0.75	$(3,2,4,1)$	0.670	$(3,1,2,4)$	0.619	$(3,2,1,4)$	0.597
$(2,4,3,1)$	0.75	$(2,4,3,1)$	0.674	$(3,2,4,1)$	0.619	$(2,4,1,3)$	0.642
$(3,1,4,2)$	0.75	$(4,2,3,1)$	0.717	$(2,4,1,3)$	0.714	$(3,1,2,4)$	0.672
$(3,2,4,1)$	0.75	$(3,1,4,2)$	0.728	$(3,1,4,2)$	0.714	$(3,1,4,2)$	0.701
$(4,1,2,3)$	0.75	$(2,4,1,3)$	0.740	$(4,2,3,1)$	0.714	$(4,2,3,1)$	0.761
$(4,1,3,2)$	0.75	$(4,1,3,2)$	0.775	$(4,1,3,2)$	0.810	$(3,4,2,1)$	0.851
$(4,2,1,3)$	0.75	$(4,2,1,3)$	0.782	$(4,2,1,3)$	0.810	$(4,2,1,3)$	0.851
$(4,2,3,1)$	0.75	$(4,1,2,3)$	0.808	$(4,1,2,3)$	0.857	$(4,1,3,2)$	0.881
$(3,4,1,2)$	1.00	$(3,4,2,1)$	0.946	$(3,4,2,1)$	0.905	$(3,4,1,2)$	0.896
$(3,4,2,1)$	1.00	$(4,3,2,1)$	0.967	$(3,4,1,2)$	0.952	$(4,1,2,3)$	0.925
$(4,3,1,2)$	1.00	$(3,4,1,2)$	0.979	$(4,3,2,1)$	0.952	$(4,3,2,1)$	0.955
$(4,3,2,1)$	1.00	$(4,3,1,2)$	1.000	$(4,3,1,2)$	1.000	$(4,3,1,2)$	1.000

Table 4: Distance from the true order with the loss function $d_{\eta\theta}$

$\eta = 2$	$\theta = 0$
Order	$d_{\eta\theta}$
$(1,2,3,4)$	0.0
$(1,2,4,3)$	0.1
$(1,3,2,4)$	0.1
$(2,1,3,4)$	0.1
$(2,1,4,3)$	0.2
$(1,3,4,2)$	0.3
$(1,4,2,3)$	0.3
$(2,3,1,4)$	0.3
$(3,1,2,4)$	0.3
$(1,4,3,2)$	0.4
$(3,2,1,4)$	0.4
$(2,4,1,3)$	0.5
$(3,1,4,2)$	0.5
$(2,3,4,1)$	0.6
$(4,1,2,3)$	0.6
$(2,4,3,1)$	0.7
$(3,2,4,1)$	0.7
$(4,1,3,2)$	0.7
$(4,2,1,3)$	0.7
$(3,4,1,2)$	0.8
$(3,4,2,1)$	0.9
$(4,2,3,1)$	0.9
$(4,3,1,2)$	0.9
$(4,3,2,1)$	1.0

$\eta = 2$	$\theta = \frac{1}{2}$
Order	$d_{\eta\theta}$
$(1,2,3,4)$	0.000
$(1,2,4,3)$	0.080
$(1,3,2,4)$	0.104
$(2,1,3,4)$	0.124
$(2,1,4,3)$	0.204
$(1,3,4,2)$	0.237
$(1,4,2,3)$	0.310
$(2,3,1,4)$	0.311
$(1,4,3,2)$	0.363
$(3,1,2,4)$	0.370
$(3,2,1,4)$	0.453
$(2,3,4,1)$	0.469
$(3,1,4,2)$	0.502
$(2,4,1,3)$	0.517
$(2,4,3,1)$	0.595
$(3,2,4,1)$	0.611
$(4,1,2,3)$	0.735
$(4,1,3,2)$	0.787
$(3,4,1,2)$	0.816
$(4,2,1,3)$	0.818
$(3,4,2,1)$	0.841
$(4,2,3,1)$	0.896
$(4,3,1,2)$	0.975
$(4,3,2,1)$	1.000

$\eta = 2$	$\theta = 1$
Order	$d_{\eta\theta}$
$(1,2,3,4)$	0.000
$(1,2,4,3)$	0.059
$(1,3,2,4)$	0.098
$(2,1,3,4)$	0.137
$(1,3,4,2)$	0.176
$(2,1,4,3)$	0.196
$(1,4,2,3)$	0.294
$(2,3,1,4)$	0.294
$(1,4,3,2)$	0.314
$(2,3,4,1)$	0.353
$(3,1,2,4)$	0.412
$(3,2,1,4)$	0.471
$(2,4,1,3)$	0.490
$(2,4,3,1)$	0.490
$(3,1,4,2)$	0.490
$(3,2,4,1)$	0.529
$(3,4,2,1)$	0.765
$(3,4,1,2)$	0.784
$(4,1,2,3)$	0.824
$(4,1,3,2)$	0.843
$(4,2,1,3)$	0.882
$(4,2,3,1)$	0.882
$(4,3,2,1)$	0.980
$(4,3,1,2)$	1.000

$\eta = 2$	$\theta = 2$
Order	$d_{\eta\theta}$
$(1,2,3,4)$	0.000
$(1,2,4,3)$	0.029
$(1,3,2,4)$	0.075
$(1,3,4,2)$	0.098
$(2,1,3,4)$	0.145
$(2,1,4,3)$	0.173
$(2,3,4,1)$	0.220
$(1,4,3,2)$	0.231
$(1,4,2,3)$	0.237
$(2,3,1,4)$	0.237
$(2,4,3,1)$	0.353
$(2,4,1,3)$	0.399
$(3,1,2,4)$	0.445
$(3,2,4,1)$	0.445
$(3,2,1,4)$	0.462
$(3,1,4,2)$	0.468
$(3,4,2,1)$	0.653
$(3,4,1,2)$	0.694
$(4,2,3,1)$	0.884
$(4,1,3,2)$	0.908
$(4,1,2,3)$	0.913
$(4,2,1,3)$	0.931
$(4,3,2,1)$	0.960
$(4,3,1,2)$	1.000

Table 5: Distance from the true order with the loss function $d_{\eta\theta}$

$\eta = \frac{1}{2}$	$\theta = 0$		$\eta = \frac{1}{2}$	$\theta = \frac{1}{2}$		$\eta = \frac{1}{2}$	$\theta = 1$		$\eta = \frac{1}{2}$	$\theta = 2$
Order	$d^K_{\eta\theta}$		Order	$d^K_{\eta\theta}$		Order	$d^K_{\eta\theta}$		Order	$d^K_{\eta\theta}$
$(1,2,3,4)$	0.000		$(1,2,3,4)$	0.000		$(1,2,3,4)$	0.000		$(1,2,3,4)$	0.000
$(1,2,4,3)$	0.241		$(1,2,4,3)$	0.193		$(1,2,4,3)$	0.152		$(1,2,4,3)$	0.090
$(1,3,2,4)$	0.241		$(1,3,2,4)$	0.236		$(1,3,2,4)$	0.228		$(1,3,2,4)$	0.203
$(2,1,3,4)$	0.241		$(2,1,3,4)$	0.273		$(2,1,3,4)$	0.304		$(1,4,2,3)$	0.286
$(1,4,2,3)$	0.341		$(1,4,2,3)$	0.334		$(1,4,2,3)$	0.322		$(1,3,4,2)$	0.293
$(3,1,2,4)$	0.341		$(3,1,2,4)$	0.386		$(1,3,4,2)$	0.380		$(2,1,3,4)$	0.360
$(4,1,2,3)$	0.418		$(1,3,4,2)$	0.429		$(3,1,2,4)$	0.429		$(1,4,3,2)$	0.376
$(1,3,4,2)$	0.482		$(2,1,4,3)$	0.466		$(2,1,4,3)$	0.456		$(2,1,4,3)$	0.450
$(2,1,4,3)$	0.482		$(4,1,2,3)$	0.473		$(1,4,3,2)$	0.474		$(3,1,2,4)$	0.509
$(2,3,1,4)$	0.482		$(2,3,1,4)$	0.509		$(4,1,2,3)$	0.526		$(2,3,1,4)$	0.563
$(1,4,3,2)$	0.582		$(1,4,3,2)$	0.527		$(2,3,1,4)$	0.531		$(3,1,4,2)$	0.599
$(2,4,1,3)$	0.582		$(3,1,4,2)$	0.579		$(3,1,4,2)$	0.581		$(4,1,2,3)$	0.624
$(3,1,4,2)$	0.582		$(2,4,1,3)$	0.607		$(2,4,1,3)$	0.626		$(2,4,1,3)$	0.646
$(3,2,1,4)$	0.582		$(3,2,1,4)$	0.622		$(3,2,1,4)$	0.657		$(2,3,4,1)$	0.653
$(4,1,3,2)$	0.659		$(4,1,3,2)$	0.666		$(4,1,3,2)$	0.678		$(3,2,1,4)$	0.712
$(4,2,1,3)$	0.659		$(2,3,4,1)$	0.702		$(2,3,4,1)$	0.683		$(4,1,3,2)$	0.714
$(3,4,1,2)$	0.682		$(4,2,1,3)$	0.709		$(3,4,1,2)$	0.752		$(2,4,3,1)$	0.736
$(2,3,4,1)$	0.724		$(3,4,1,2)$	0.720		$(4,2,1,3)$	0.754		$(3,4,1,2)$	0.796
$(4,3,1,2)$	0.759		$(2,4,3,1)$	0.800		$(2,4,3,1)$	0.778		$(3,2,4,1)$	0.802
$(2,4,3,1)$	0.823		$(4,3,1,2)$	0.807		$(3,2,4,1)$	0.809		$(4,2,1,3)$	0.826
$(3,2,4,1)$	0.823		$(3,2,4,1)$	0.815		$(4,3,1,2)$	0.848		$(3,4,2,1)$	0.886
$(4,2,3,1)$	0.900		$(4,2,3,1)$	0.902		$(3,4,2,1)$	0.903		$(4,3,1,2)$	0.910
$(3,4,2,1)$	0.923		$(3,4,2,1)$	0.913		$(4,2,3,1)$	0.906		$(4,2,3,1)$	0.916
$(4,3,2,1)$	1.000		$(4,3,2,1)$	1.000		$(4,3,2,1)$	1.000		$(4,3,2,1)$	1.000

Table 6: Distance from the true order with the loss function $d^K_{\eta\theta}$

$\eta = 1 \quad \theta = 0$	
Order	$d^K_{\eta\theta}$
$(1,2,3,4)$	0.000
$(1,2,4,3)$	0.167
$(1,3,2,4)$	0.167
$(2,1,3,4)$	0.167
$(1,3,4,2)$	0.333
$(1,4,2,3)$	0.333
$(2,1,4,3)$	0.333
$(2,3,1,4)$	0.333
$(3,1,2,4)$	0.333
$(1,4,3,2)$	0.500
$(2,3,4,1)$	0.500
$(2,4,1,3)$	0.500
$(3,1,4,2)$	0.500
$(3,2,1,4)$	0.500
$(4,1,2,3)$	0.500
$(2,4,3,1)$	0.667
$(3,2,4,1)$	0.667
$(3,4,1,2)$	0.667
$(4,1,3,2)$	0.667
$(4,2,1,3)$	0.667
$(3,4,2,1)$	0.833
$(4,2,3,1)$	0.833
$(4,3,1,2)$	0.833
$(4,3,2,1)$	1.000

$\eta = 1 \quad \theta = \frac{1}{2}$	
Order	$d^K_{\eta\theta}$
$(1,2,3,4)$	0.000
$(1,2,4,3)$	0.130
$(1,3,2,4)$	0.159
$(2,1,3,4)$	0.184
$(1,3,4,2)$	0.289
$(2,1,4,3)$	0.314
$(1,4,2,3)$	0.318
$(2,3,1,4)$	0.343
$(3,1,2,4)$	0.368
$(1,4,3,2)$	0.448
$(2,3,4,1)$	0.473
$(3,1,4,2)$	0.498
$(2,4,1,3)$	0.502
$(3,2,1,4)$	0.527
$(4,1,2,3)$	0.552
$(2,4,3,1)$	0.632
$(3,2,4,1)$	0.657
$(4,1,3,2)$	0.682
$(3,4,1,2)$	0.686
$(4,2,1,3)$	0.711
$(3,4,2,1)$	0.816
$(4,2,3,1)$	0.841
$(4,3,1,2)$	0.870
$(4,3,2,1)$	1.000

$\eta = 1 \quad \theta = 1$	
Order	$d^K_{\eta\theta}$
$(1,2,3,4)$	0.00
$(1,2,4,3)$	0.10
$(1,3,2,4)$	0.15
$(2,1,3,4)$	0.20
$(1,3,4,2)$	0.25
$(1,4,2,3)$	0.30
$(2,1,4,3)$	0.30
$(2,3,1,4)$	0.35
$(1,4,3,2)$	0.40
$(3,1,2,4)$	0.40
$(2,3,4,1)$	0.45
$(2,4,1,3)$	0.50
$(3,1,4,2)$	0.50
$(3,2,1,4)$	0.55
$(2,4,3,1)$	0.60
$(4,1,2,3)$	0.60
$(3,2,4,1)$	0.65
$(3,4,1,2)$	0.70
$(4,1,3,2)$	0.70
$(4,2,1,3)$	0.75
$(3,4,2,1)$	0.80
$(4,2,3,1)$	0.85
$(4,3,1,2)$	0.90
$(4,3,2,1)$	1.00

$\eta = 1 \quad \theta = 2$	
Order	$d^K_{\eta\theta}$
$(1,2,3,4)$	0.000
$(1,2,4,3)$	0.057
$(1,3,2,4)$	0.129
$(1,3,4,2)$	0.186
$(2,1,3,4)$	0.229
$(1,4,2,3)$	0.257
$(2,1,4,3)$	0.286
$(1,4,3,2)$	0.314
$(2,3,1,4)$	0.357
$(2,3,4,1)$	0.414
$(3,1,2,4)$	0.457
$(2,4,1,3)$	0.486
$(3,1,4,2)$	0.514
$(2,4,3,1)$	0.543
$(3,2,1,4)$	0.586
$(3,2,4,1)$	0.643
$(4,1,2,3)$	0.686
$(3,4,1,2)$	0.714
$(4,1,3,2)$	0.743
$(3,4,2,1)$	0.771
$(4,2,1,3)$	0.814
$(4,2,3,1)$	0.871
$(4,3,1,2)$	0.943
$(4,3,2,1)$	1.000

Table 7: Distance from the true order with the loss function $d^K_{\eta\theta}$

$\eta = 2$	$\theta = 0$		$\eta = 2$	$\theta = \frac{1}{2}$		$\eta = 2$	$\theta = 1$		$\eta = 2$	$\theta = 2$
Order	$d^K_{\eta\theta}$		Order	$d^K_{\eta\theta}$		Order	$d^K_{\eta\theta}$		Order	$d^K_{\eta\theta}$
$(1,2,3,4)$	0.000		$(1,2,3,4)$	0.000		$(1,2,3,4)$	0.00		$(1,2,3,4)$	0.000
$(1,2,4,3)$	0.071		$(1,2,4,3)$	0.054		$(1,2,4,3)$	0.04		$(1,2,4,3)$	0.022
$(1,3,2,4)$	0.071		$(1,3,2,4)$	0.066		$(1,3,2,4)$	0.06		$(1,3,2,4)$	0.049
$(2,1,3,4)$	0.071		$(2,1,3,4)$	0.076		$(2,1,3,4)$	0.08		$(1,3,4,2)$	0.071
$(1,3,4,2)$	0.143		$(1,3,4,2)$	0.119		$(1,3,4,2)$	0.10		$(2,1,3,4)$	0.087
$(2,1,4,3)$	0.143		$(2,1,4,3)$	0.130		$(2,1,4,3)$	0.12		$(2,1,4,3)$	0.109
$(2,3,1,4)$	0.143		$(2,3,1,4)$	0.142		$(2,3,1,4)$	0.14		$(2,3,1,4)$	0.136
$(2,3,4,1)$	0.214		$(2,3,4,1)$	0.195		$(2,3,4,1)$	0.18		$(2,3,4,1)$	0.158
$(1,4,2,3)$	0.286		$(1,4,2,3)$	0.263		$(1,4,2,3)$	0.24		$(1,4,2,3)$	0.196
$(3,1,2,4)$	0.286		$(3,1,2,4)$	0.304		$(1,4,3,2)$	0.28		$(1,4,3,2)$	0.217
$(1,4,3,2)$	0.357		$(1,4,3,2)$	0.317		$(2,4,1,3)$	0.32		$(2,4,1,3)$	0.283
$(2,4,1,3)$	0.357		$(2,4,1,3)$	0.339		$(3,1,2,4)$	0.32		$(2,4,3,1)$	0.304
$(3,1,4,2)$	0.357		$(3,1,4,2)$	0.357		$(2,4,3,1)$	0.36		$(3,1,2,4)$	0.348
$(3,2,1,4)$	0.357		$(3,2,1,4)$	0.369		$(3,1,4,2)$	0.36		$(3,1,4,2)$	0.370
$(2,4,3,1)$	0.429		$(2,4,3,1)$	0.393		$(3,2,1,4)$	0.38		$(3,2,1,4)$	0.397
$(3,2,4,1)$	0.429		$(3,2,4,1)$	0.423		$(3,2,4,1)$	0.42		$(3,2,4,1)$	0.418
$(3,4,1,2)$	0.571		$(3,4,1,2)$	0.567		$(3,4,1,2)$	0.56		$(3,4,1,2)$	0.543
$(3,4,2,1)$	0.643		$(3,4,2,1)$	0.620		$(3,4,2,1)$	0.60		$(3,4,2,1)$	0.565
$(4,1,2,3)$	0.643		$(4,1,2,3)$	0.683		$(4,1,2,3)$	0.72		$(4,1,2,3)$	0.783
$(4,1,3,2)$	0.714		$(4,1,3,2)$	0.737		$(4,1,3,2)$	0.76		$(4,1,3,2)$	0.804
$(4,2,1,3)$	0.714		$(4,2,1,3)$	0.749		$(4,2,1,3)$	0.78		$(4,2,1,3)$	0.832
$(4,2,3,1)$	0.786		$(4,2,3,1)$	0.803		$(4,2,3,1)$	0.82		$(4,2,3,1)$	0.853
$(4,3,1,2)$	0.929		$(4,3,1,2)$	0.946		$(4,3,1,2)$	0.96		$(4,3,1,2)$	0.978
$(4,3,2,1)$	1.000		$(4,3,2,1)$	1.000		$(4,3,2,1)$	1.00		$(4,3,2,1)$	1.000

Table 8: Distance from the true order with the loss function $d^K_{\eta\theta}$