ARTICLE IN PRESS

European Journal of Operational Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/eor

Decision Support

An indifference result for social choice rules in large societies

Dezső Bednay [a], Balázs Fleiner [b], Attila Tasnádi [a,*]

[a] Department of Mathematics, Institute of Data Analytics and Information Systems, Corvinus University of Budapest, Fővám tér 13-15, 1093 Budapest, Hungary

[b] Department of Applied Quantitative Methods, Budapest Business University, Buzogány utca 10-12, 1149 Budapest, Hungary

ARTICLE INFO

Keywords:
Group decisions and negotiations
Distance-based aggregation
Kendall τ
Spearman rank correlation
Spearman footrule

ABSTRACT

Social choice rules can be defined or derived by minimizing distance-based objective functions. One problem with this approach is that any social choice rule can be derived by selecting an appropriate distance function. Another problem comes from the computational difficulty of determining the solution of some social choice rules. We provide a general positive indifference result when looking at expected average distances by showing that on 'average' each social choice rule performs equally well with respect to a very large class of distance functions if the number of voters is large. Our result applies also to the frequently employed Kendall τ, Spearman rank correlation and Spearman footrule 'distance functions'.

1. Introduction

The debate between Borda and Condorcet at the end of the 18th century demonstrated well the challenge in choosing a widely accepted social choice rule or more restrictively a voting rule, where the former one selects an ordering of the alternatives, while the latter one only a winning alternative (or a set of winning alternatives). From an axiomatic point of view (Arrow, 1951) settled the problem by terminating the search for an 'ideal' social choice rule (SCR). In his famous impossibility theorem he showed that if there are at least three alternatives (or candidates), there does not exist a SCR fulfilling four natural requirements.

In this paper we follow an alternative route to the axiomatic one, the 'operations research approach', which strives for selecting SCRs as solutions to appropriately defined distance minimization problems. The latter approach is either employed to define certain rules, like the Kemény–Young method (Kemény, 1959), or results in known SCRs, like the Borda count (Dwork et al., 2002). Other SCRs defined directly as solutions of optimization problems are Slater's and Dogson's rules both using the Kendall τ distance (see Eckert and Klamler (2011)).

If there is no natural distance function related to the given preference aggregation problem, then we can choose from plenty of distance functions. The problem was underlined by Lehrer and Nitzan (1985) and by Campbell and Nitzan (1986) who showed that basically any voting rule can be distance rationalized. This result can be regarded as a negative result like Arrow's impossibility theorem. The approach of minimizing the distance from a set of profiles with a clear winner such as the unanimous winner, the majoritarian winner, or the Condorcet winner has been developed further by Elkind et al. (2015) among others, and a survey of these results is given by Elkind and Slinko (2015). Bednay et al. (2017) offered a 'dual' approach based on distance maximization from the closest dictator in which the dictatorial rules are taken as the benchmark. The extension of Hadjibeyli and Wilson (2019) covered both voting rules and SCRs as two extreme cases in their framework, which allows as a social outcome a linear ordering of a subset of all alternatives. Kadziński et al. (2022) extended the preference aggregation problem to stochastic preferences.

From a more general point of view, a SCR provides a solution to the problem of finding a common ranking for a group, also known as the rank aggregation problem. Cook (2006) gave a general overview of the axiomatic and distance based preference aggregation problem. In addition, he established a 'crossover' link between rank aggregation and multi-criteria decision making. The main point is that preferences can be represented by pairwise comparison matrices (PCM), and then the bridge to methods in operations research relying on PCMs is evident. For instance, Saaty (1980) introduced his method of analytic hierarchy process and provided methods for weight assignments to alternatives based on PCMs, which can be used to derive a social ranking of alternatives. For example, Bozóki et al. (2016) determine a ranking of number one ranked male tennis players in the open era by employing incomplete PCMs.

Besides the axiomatic and distance based comparison of SCRs and voting rules there are other meaningful approaches. Burka et al. (2022) followed a quasi-experimental approach by teaching neural networks on a set of selected profiles. By finding the right and credible incentives

* Corresponding author.
E-mail addresses: dezso.bednay@uni-corvinus.hu (D. Bednay), fleinerb@gmail.com (B. Fleiner), attila.tasnadi@uni-corvinus.hu (A. Tasnádi).
URL: http://www.uni-corvinus.hu/~tasnadi (A. Tasnádi).

https://doi.org/10.1016/j.ejor.2024.09.018
Received 8 February 2024; Accepted 9 September 2024
Available online 12 September 2024

for the participants (Ambuehl & Bernheim, 2024) compared voting rules in an experiment. Both works found that the Borda count is the most salient voting rule. Montes et al. (2020) established a link between voting rules and stochastic orderings.

In this paper we address the expected average performance of SCRs in case of a large number of voters. We find that for a large class of a combination of distance functions and probability distributions over the set of profiles the expected objective function value of any anonymous SCR tends to the same value. Hence, in contrast to the result in the deterministic setting stating that different distance functions lead to different SCRs we arrive at the conclusion that under appropriate conditions considering the expected average asymptotic performance all SCRs are equally good. We give a necessary and sufficient condition for our 'indifference result', where by indifference we mean that the selection of the SCR does not really matter if the number of voters is sufficiently large. We can interpret this result in favor of the very simple plurality rule, which is the most frequently employed one in elections, while based on its axiomatic properties it is one of the most widely criticized rules. The negative opinion of social choice theorists about the plurality rule are highlighted in 'And the looser is ... plurality voting', by Laslier (2011).

Even if our indifference result does not hold we can bound the expected performances of anonymous SCRs by the expected performances of two appropriately selected constant rules, where a SCR is a constant rule if it selects the same linear ordering for any profile.

From the computational point of view the approximation results of SCRs by each other are related to the worst case analysis. This question is also important since some rules are difficult to calculate, like the Kemény–Young method, and therefore the best approximating polynomial time rule can serve as a replacement. Fagin et al. (2016) showed in their elaborate contribution that the Borda count, the median rank based SCR, the Copeland method and the Kemény–Young method are constant factor approximations of each other, while this is not the case for the plurality rule, the single transferable vote rule and the Simpson–Kramer minmax method. In their work they considered the Kendall τ, the Spearman footrule and the Spearman rank correlation distances as many other studies (e.g. Diaconis & Graham, 1977; Monjardet, 1997). These three distance functions are included in the large class of distance functions for which our result holds. In contrast to the worst case analysis in this paper we carry out an average case analysis. Theorem 2 provides an implicit necessary and sufficient condition for our indifference result to hold, which requires that the expected average distances to all constant SCRs have to be the same. Two sufficient conditions for our indifference result to hold are given in Corollaries 1 and 2. The first one works for the impartial culture (IC) case in which each preference profile is equally likely and the Kendall τ distance among other distances, while the second one works for any culture in which the distribution of alternatives is (at least asymptotically) uniform and the distance function is a function of the rank differences of alternatives. Note that among many other distributions and distance functions Corollary 2 is applicable for the two most frequently applied cultures, the IC case and the impartial anonymous culture (IAC) case in which each anonymous preference profile is equally likely, and also for both the Spearman footrule and the Spearman rank correlation distances.

We would like to emphasize that SCRs are used, far more widely than in the traditional social choice context, especially in areas in which rankings have to be produced based on various inputs. In general, aggregating rankings can be also regarded as combining inputs from multiple sources like in automated decision making, machine learning (e.g. Volkovs & Zemel, 2014) or database middleware (e.g. Masthoff, 2004), or in the determination of the results in sport competitions (e.g. Csató, 2023, and Ausloos, 2024). The problem also arises in coding theory since the alternatives can be regarded as letters and the rankings as strings, and the distance function can be utilized in error detection (Bortolussi et al., 2012). Nowadays, the problem of aggregating rankings also emerges in the link analysis in networks like the world wide web, which lie at the heart of web search algorithms (Borodin et al., 2005). Therefore, a partially parallel literature emerged, which sometimes uses similar procedures without referring to the procedures already known in social choice (e.g. the MedRank algorithm and the Bucklin rule) and that, more importantly, in a pragmatic way the problem at hand determines the appropriate choice of a distance, while the social choice literature is more focused on general principles.

From another point of view in the social choice context the number of voters is usually significantly larger than the number of candidates, while in the applications in computer science and operations research the number of alternatives is frequently larger than the number of experts (or voters).

The structure of the paper is as follows. Section 2 introduces the basic notations, Section 3 presents our main asymptotic result and Section 4 contains some relevant examples. Finally, Section 5 concludes.

2. The framework

Let $A = \{a_1, \dots, a_m\}$ be the set of alternatives, where $m \geq 2$, and $N = \{1, \dots, n\}$ be the set of voters. We shall denote by $\mathcal{P}$ the set of all linear (or strict preference) orderings (irreflexive, transitive and total binary relations) on A and by $\mathcal{P}^n$ the set of all preference profiles. If $\Pi = (\succ_1, \dots, \succ_n) \in \mathcal{P}^n$ and $i \in N$, then $\succ_i$ is the preference ordering of voter i over A.

Definition 1. A mapping $F : \mathcal{P}^n \to \mathcal{P}$ that selects a linear ordering is called a *social choice rule*, henceforth, SCR.

At this point we introduce the Borda count and the constant rule to illustrate the notions of SCRs. The *Borda count*, briefly denoted by BC, orders the alternatives based on the sum of their ranks. In particular, an alternative with a lower sum of ranks is preferred over an alternative with a higher sum of ranks. We shall denote by $rk[a, \succ]$ the *rank* of alternative a in the ordering $\succ \in \mathcal{P}$ (i.e. $rk[a, \succ] = 1$ if a is the top alternative in the ranking $\succ$, $rk[a, \succ] = 2$ if a is second-best, and so on). Clearly, there exists profiles for which some alternatives have the same sums of ranks. Therefore, the essential part in the definition of the Borda count does not determine a strict preference for all profiles. Hence, to arrive at a Borda count type SCR we employ for simplicity a so-called lexicographic tie-breaking rule that orders alternatives with the same sum of ranks following an exogenously given linear ordering of alternatives τ. Of course, ties can be broken in many other ways. Then the SCR BC is the *Borda count* if for all $\Pi \in \mathcal{P}^n$ and all pairs of distinct alternatives a and b we have

$$a \; BC_\tau(\Pi) \; b \Leftrightarrow \sum_{i=1}^{n} rk[a, \succ_i] < \sum_{i=1}^{n} rk[b, \succ_i] \text{ or } \sum_{i=1}^{n} rk[a, \succ_i] = \sum_{i=1}^{n} rk[b, \succ_i] \text{ and } a \tau b.$$

The trivial *constant rule*, denoted by CR, assigns to each profile the same fixed preference relation. Formally, let $\succ^* \in \mathcal{P}$ be a fixed linear ordering and we define the *constant rule* by $CR(\Pi) = \succ^*$ for all $\Pi \in \mathcal{P}^n$.

Let $\mathcal{F} = \mathcal{P}^{\mathcal{P}^n}$ be the set of SCRs and $\mathcal{F}^{an} \subset \mathcal{F}$ be the set of anonymous SCRs. Since for anonymous SCRs only the numbers $n_1, n_2, \dots, n_{m!} \in \mathbb{N}$ of occurrences of the respective enumeration of the orderings $o_1, o_2, \dots, o_{m!}$ in a preference profile matter, where $n_1 + n_2 + \cdots + n_{m!} = n$, the necessary information of an anonymous preference profile $\Pi \in \mathcal{P}^n$ to determine $F(\Pi)$ is contained in the anonymous profile $\pi = (n_1, n_2, \dots, n_{m!}) \in \mathbb{N}_n^{m!}$, which we briefly also call a profile and we write $F(\pi) = F(\Pi)$ with a slight abuse of notation. We shall denote by $\mathcal{AP}^n = \mathbb{N}_n^{m!}$ the set of anonymous preference profiles.

We shall denote by $d : \mathcal{P} \times \mathcal{P} \to \mathbb{R}_+$ a quasidistance function or quasimetric on the set of preferences, henceforth briefly a distance function, which satisfies (i) $d(\succ, \succ') = 0$ if and only if $\succ = \succ'$, (ii) $d(\succ, \succ') = d(\succ', \succ)$ for any $\succ, \succ' \in \mathcal{P}$, and (iii) $d(\succ, \succ') \leq c(d(\succ, \succ''$

$) + d(\succ'', \succ'))$ for any $\succ, \succ' \succ'' \in \mathcal{P}$ for a given $c > 0$ (also called a c-relaxed triangle inequality). The three most frequently employed distance functions in the preference aggregation framework are

$$d_K(\succ, \succ') = \#\{(a_i, a_j) \in A^2 \mid a_i \succ a_j \text{ and } a_j \succ' a_i\}, \tag{2.1}$$

$$d_1(\succ, \succ') = \sum_{i=1}^{m} |rk[a_i, \succ] - rk[a_i, \succ']|, \tag{2.2}$$

$$d_2(\succ, \succ') = \sum_{i=1}^{m} (rk[a_i, \succ] - rk[a_i, \succ'])^2, \tag{2.3}$$

where d_K and d_1 are metrics ($c = 1$) and d_2 is just a quasimetric with $c = 2$. The distance functions defined by (2.1), (2.2) and (2.3) are known in the literature as the Kendall τ (or Kemény), the Spearman footrule and the Spearman rank correlation distances.

We extend d to distances between profiles and preferences in an additive way:

$$d(\Pi, \succ) = \frac{1}{n} \sum_{i=1}^{n} d(\succ_i, \succ),$$

where $\Pi = (\succ_1, \dots, \succ_n)$. The average distance between an anonymous profile $\pi \in \mathcal{AP}^n$ and a preference relation o_k equals

$$d(\pi, o_k) = \sum_{j=1}^{m!} \frac{n_j}{n} d(o_j, o_k).$$

We shall denote by p_{ij} the probability that voter $i \in N$ chooses ordering o_j. We assume that the probability distributions of the voters are independent. Let

$$p_j = \frac{1}{n} \sum_{i=1}^{n} p_{ij}$$

be the probability of the occurrence of ordering o_j. Note that $\sum_{i=1}^{n} \sum_{j=1}^{m!} p_{ij} = n$, and therefore $(p_j)_{j=1}^{m!}$ specifies a probability distribution.

3. Limits for general distance functions

In this Section we prove our main theorem about the limit of expected average distances of anonymous SCRs.

Let random variable $X_j^{(n)} : \mathcal{P}^n \to [0, 1]$ stand for the proportion of voters with ordering o_j in a profile. The vector of these random variables will be denoted by $\mathbf{X}^{(n)}$ for a given n. For notational convenience we assume that the probability distribution of the orderings $(o_j)_{j=1}^{m!}$ is the same for any n, and therefore we do not need superscripts for denoting p_j.[1]

Proposition 1. *Let $\varepsilon > 0$ be given. Then for any $\delta > 0$ there exists an n_0 such that for any $n \geq n_0$ we have*

$$P\left(|X_1^{(n)} - p_1| < \varepsilon \text{ and } |X_2^{(n)} - p_2| < \varepsilon \text{ and } \dots \text{ and } |X_{m!}^{(n)} - p_{m!}| < \varepsilon\right) \geq 1 - \delta.$$

Proof. We show that the probability of the complementary event is close to zero. Since

$$P\left(|X_1^{(n)} - p_1| \geq \varepsilon \text{ or } |X_2^{(n)} - p_2| \geq \varepsilon \text{ or } \dots \text{ or } |X_{m!}^{(n)} - p_{m!}| \geq \varepsilon\right) \leq \sum_{j=1}^{m!} P\left(|X_j^{(n)} - p_j| \geq \varepsilon\right)$$

it is sufficient to show that $P\left(|X_j^{(n)} - p_j| \geq \varepsilon\right)$ is close to zero since m and $m!$ are fixed.

We shall denote by $X_{ij}^{(n)}$ the indicator random variable showing whether voter i chooses ordering o_j or not. Then $X_j^{(n)} = (\sum_{i=1}^{n} X_{ij}^{(n)})/n$, where we are taking the sum of independent random variables since the voters are selecting their orderings independently. Clearly,

$$EX_j^{(n)} = p_j \quad \text{and} \quad VarX_j^{(n)} = \frac{\sum_{i=1}^{n} p_{ij}(1 - p_{ij})}{n^2} \leq \frac{1}{4n},$$

which in turn establishes that $P\left(|X_j^{(n)} - p_j| \geq \varepsilon\right)$ is close to zero for sufficiently large n by Chebyshev's inequality. In particular,

$$P\left(|X_j^{(n)} - p_j| \geq \varepsilon\right) \leq \frac{1}{4n\varepsilon^2}. \quad \square$$

Proposition 1 implies that if the number of voters is large, we only have to care about the 'average' distribution of orderings because only a few distributions deviate from it.

The expected average distance for a SCR $F \in \mathcal{F}$ equals

$$E(d(\mathbf{X}^{(n)}, F(\mathbf{X}^{(n)}))) = \sum_{(o_{j_1}, \dots, o_{j_n}) \in \mathcal{P}^n} p_{1j_1} \cdots p_{nj_n} \times d((o_{j_1}, \dots, o_{j_n}), F((o_{j_1}, \dots, o_{j_n}))).$$

Let $d_{\max}$ be the maximum distance between two orderings associated with distance d. Since the constant functions play a special role we define the expected distance of the constant SCR $F(\Pi) = o_k$ for all $\Pi \in \mathcal{P}^n$ from the set of all profiles in the limit by

$$\overline{d}(o_k) = \lim_{n \to \infty} \sum_{\Pi \in \mathcal{P}^n} p_{1j_1} \cdots p_{nj_n} d(\Pi, o_k), \tag{3.4}$$

which, as the following derivation shows, can be determined more concisely without requiring the evaluation of a limit:

$$\begin{aligned}
\overline{d}(o_k) &= \lim_{n \to \infty} \sum_{\Pi \in \mathcal{P}^n} p_{1j_1} \cdots p_{nj_n} d(\Pi, o_k) \\
&= \lim_{\varepsilon \to 0} \lim_{n \to \infty} \sum_{\Pi \in A_\varepsilon^{(n)}} p_{1j_1} \cdots p_{nj_n} d(\Pi, o_k) \\
&= \lim_{\varepsilon \to 0} \lim_{n \to \infty} \sum_{\Pi \in A_\varepsilon^{(n)}} p_{1j_1} \cdots p_{nj_n} \sum_{i=1}^{m!} \frac{n_j}{n} d(o_j, o_k) \\
&= \lim_{\varepsilon \to 0} \lim_{n \to \infty} \sum_{\Pi \in A_\varepsilon^{(n)}} p_{1j_1} \cdots p_{nj_n} \sum_{i=1}^{m!} (p_j \pm \varepsilon) d(o_j, o_k) \\
&= \lim_{\varepsilon \to 0} \lim_{n \to \infty} \sum_{\Pi \in A_\varepsilon^{(n)}} p_{1j_1} \cdots p_{nj_n} \sum_{i=1}^{m!} p_j d(o_j, o_k) \pm \\
&\quad \lim_{\varepsilon \to 0} \lim_{n \to \infty} \sum_{\Pi \in A_\varepsilon^{(n)}} p_{1j_1} \cdots p_{nj_n} \sum_{i=1}^{m!} \varepsilon d(o_j, o_k) \\
&= \lim_{\varepsilon \to 0} \lim_{n \to \infty} \sum_{\Pi \in A_\varepsilon^{(n)}} p_{1j_1} \cdots p_{nj_n} \sum_{i=1}^{m!} p_j d(o_j, o_k) \pm \lim_{\varepsilon \to 0} 1 \cdot \varepsilon \cdot m! \cdot d_{max} \\
&= \sum_{i=1}^{m!} p_j d(o_j, o_k) \cdot \lim_{\varepsilon \to 0} \lim_{n \to \infty} \sum_{\Pi \in A_\varepsilon^{(n)}} p_{1j_1} \cdots p_{nj_n} = \sum_{j=1}^{m!} p_j d(o_j, o_k),
\end{aligned} \tag{3.5}$$

where the second equality follows from Proposition 1 and the introduction of $\pm$ means that the initial sequence can be bounded from below and above, which implies that the limit defining $\overline{d}(o_k)$ exists since the two bounding sequences tend to the same limit. To summarize, we can also write

$$\overline{d}(o_k) = \sum_{j=1}^{m!} p_j d(o_j, o_k). \tag{3.6}$$

We shall denote by $\succ^*$ and $\succ^{**}$ the preferences (or the respective constant functions) with the smallest and largest expected distances from the set of all profiles in the limit. More formally,

$$\underline{d} = \min_{\succ \in \mathcal{P}} \sum_{j=1}^{m!} p_j d(o_j, \succ) \quad \text{and} \quad \overline{d} = \max_{\succ \in \mathcal{P}} \sum_{j=1}^{m!} p_j d(o_j, \succ),$$

[1] It would be sufficient to assume that the sequence of probability distributions of $\mathbf{X}^{(n)}$ has a limit distribution.

where these distances are minimized and maximized at $\succ^*$ and $\succ^{**}$, respectively.

Strictly speaking, we have for each n different distance functions and SCRs. At least in case of distance functions for the sequence $\mathcal{P}^n$ of domains we are obtaining the sequence of distance functions in an additive way from a fixed distance function defined on the set of preferences. For the case of anonymous SCRs we have in mind that for different n we are employing the same type of formula. For instance, we can take for all n in the sequence the Borda count. However, the next proposition does not even require a 'consistent selection' of anonymous SCRs for different numbers of voters.

Theorem 1. *The distance between an arbitrary family of anonymous SCRs $(F : \mathcal{P}^n \to \mathcal{P})_{n=1}^{\infty}$ is bounded by the constant SCRs $\succ^*$ and $\succ^{**}$ as n tends to infinity in the following way:*

$$\underline{d} \le \liminf_{n\to\infty} E(d(\mathbf{X}^{(n)}, F(\mathbf{X}^{(n)}))) \le \limsup_{n\to\infty} E(d(\mathbf{X}^{(n)}, F(\mathbf{X}^{(n)}))) \le \overline{d}, \tag{3.7}$$

where the inequalities are tight.

Proof. For a given n let

$$A_\varepsilon^{(n)} = \left\{ \Pi \in \mathcal{P}^n \mid \max_{l=1,\dots,m!} \left| X_l^{(n)}(\Pi) - p_l \right| < \varepsilon \right\}$$

be the set of those profiles for which the frequency of any ordering in the profile just differs from its probability (proportion) in $\mathbf{X}^{(n)}$ by less than $\varepsilon > 0$. From Proposition 1 it follows that $P(A_\varepsilon^{(n)}) \to 1$ as $n \to \infty$. In addition, let $A_{\varepsilon,k}^{(n)} = \{\Pi \in A_\varepsilon^{(n)} \mid F(\Pi) = o_k\}$, where $k = 1, \dots, m!$.

We show that if $A_{\varepsilon,k}^{(n)} \neq \emptyset$ for infinitely many n, then $\overline{d}(o_k)$ can be determined in an alternative way. Informally, if n is large enough, then the ratios of the ordering o_k remain roughly the same within a profile in $A_\varepsilon^{(n)}$ as within a profile in $A_{\varepsilon,k}^{(n)}$. Let

$$q_{\varepsilon,k}^{(n)} = \sum_{\Pi \in A_{\varepsilon,k}^{(n)}} p_{1j_1} \cdots p_{nj_n}.$$

Clearly, $\mathbf{q}_\varepsilon^{(n)} = (q_{\varepsilon,1}^{(n)}, \dots, q_{\varepsilon,m!}^{(n)})$ is a bounded sequence for any given $\varepsilon > 0$. Pick an arbitrary convergent subsequence of $\mathbf{q}_\varepsilon^{(n)}$ and for notational convenience let us assume that $\mathbf{q}_\varepsilon^{(n)}$ is already convergent, which tends to $\mathbf{q}_\varepsilon$. Note that by Proposition 1 its limit specifies a probability distribution. Then through similar calculations as in (3.5) we get

$$\begin{aligned}
\lim_{\varepsilon\to 0}\lim_{n\to\infty} \sum_{\Pi \in A_{\varepsilon,k}^{(n)}} p_{1j_1}\cdots p_{nj_n} d(\Pi, o_k) &= \lim_{\varepsilon\to 0}\lim_{n\to\infty} \sum_{\Pi \in A_{\varepsilon,k}^{(n)}} p_{1j_1}\cdots p_{nj_n} \sum_{i=1}^{m!} \frac{n_j}{n} d(o_j, o_k) \\
&= \lim_{\varepsilon\to 0}\lim_{n\to\infty} \sum_{\Pi \in A_{\varepsilon,k}^{(n)}} p_{1j_1}\cdots p_{nj_n} \sum_{i=1}^{m!} (p_j \pm \varepsilon) d(o_j, o_k) \\
&= \lim_{\varepsilon\to 0}\lim_{n\to\infty} \sum_{\Pi \in A_{\varepsilon,k}^{(n)}} p_{1j_1}\cdots p_{nj_n} \sum_{i=1}^{m!} p_j d(o_j, o_k) \pm \\
&\quad \lim_{\varepsilon\to 0}\lim_{n\to\infty} \sum_{\Pi \in A_{\varepsilon,k}^{(n)}} p_{1j_1}\cdots p_{nj_n} \sum_{i=1}^{m!} \varepsilon d(o_j, o_k) \\
&= \lim_{\varepsilon\to 0}\lim_{n\to\infty} \sum_{\Pi \in A_{\varepsilon,k}^{(n)}} p_{1j_1}\cdots p_{nj_n} \sum_{i=1}^{m!} p_j d(o_j, o_k) \pm \\
&\quad \lim_{\varepsilon\to 0} 1 \cdot \varepsilon \cdot m! \cdot d_{max} \\
&= \sum_{i=1}^{m!} p_j d(o_j, o_k) \cdot \lim_{\varepsilon\to 0}\lim_{n\to\infty} \sum_{\Pi \in A_{\varepsilon,k}^{(n)}} p_{1j_1}\cdots p_{nj_n} \\
&= q_k \sum_{j=1}^{m!} p_j d(o_j, o_k) = q_k \overline{d}(o_k),
\end{aligned} \tag{3.8}$$

where $q_k = \lim_{\varepsilon\to 0} q_{\varepsilon,k}$, which exists since $q_{\varepsilon,k}$ is decreasing in ε. Furthermore, $q_1, \dots, q_{m!}$ specifies a probability distribution.

From the law of total expectation we have

$$E(d(\mathbf{X}^{(n)}, F(\mathbf{X}^{(n)}))) = P(A_\varepsilon^{(n)}) E(d(\mathbf{X}^{(n)}, F(\mathbf{X}^{(n)})) \mid \mathbf{X}^{(n)} \in A_\varepsilon^{(n)}) + P(\overline{A}_\varepsilon^{(n)}) E(d(\mathbf{X}^{(n)}, F(\mathbf{X}^{(n)})) \mid \mathbf{X}^{(n)} \in \overline{A}_\varepsilon^{(n)}). \tag{3.9}$$

By Proposition 1 $P(A_\varepsilon^{(n)})$ and $P(\overline{A}_\varepsilon^{(n)})$ tend to 1 and 0, respectively, as n tends to infinity. Since the second conditional expected value in (3.9) is bounded the second summand is zero. We turn to the evaluation of the first conditional expected value in (3.9). Then

$$\begin{aligned}
E(d(\mathbf{X}^{(n)}, F(\mathbf{X}^{(n)})) \mid \mathbf{X}^{(n)} \in A_\varepsilon^{(n)}) &= \frac{\sum_{\Pi \in A_\varepsilon^{(n)}} p_{1j_1}\cdots p_{nj_n} d(\Pi, F(\Pi))}{\sum_{\Pi \in A_\varepsilon^{(n)}} p_{1j_1}\cdots p_{nj_n}} \\
&= \frac{\sum_{k=1}^{m!} \sum_{\Pi \in A_{\varepsilon,k}^{(n)}} p_{1j_1}\cdots p_{nj_n} d(\Pi, o_k)}{\sum_{\Pi \in A_\varepsilon^{(n)}} p_{1j_1}\cdots p_{nj_n}} \\
&= \frac{\sum_{k=1}^{m!} q_{\varepsilon,k}^{(n)} \sum_{l=1}^{m!} \frac{n_k}{n} d(o_l, o_k)}{\sum_{k=1}^{m!} q_{\varepsilon,k}^{(n)}}
\end{aligned} \tag{3.10}$$

from which we can see that the conditional expected value in (3.10) tends to the weighted average of the expected distances from the $m!$ constant functions. Therefore, the limit points of the sequence in (3.9) cannot be smaller than $\underline{d}$ and cannot be larger than $\overline{d}$. □

From Theorem 1 we can see that in general the limits of the expected distances of different anonymous SCRs may differ. Though our next statement is a simple corollary of Theorem 1 we formulate it as a theorem since it is our main indifference result.

Theorem 2. *The limits of the expected distances between any arbitrary family of anonymous SCRs $(F : \mathcal{P}^n \to \mathcal{P})_{n=1}^{\infty}$ and any constant SCR tend to zero if and only if $\underline{d} = \overline{d}$.*

Since $\underline{d} = \overline{d}$ is an implicit condition we present two important settings under which it is satisfied.

Corollary 1. *If $(p_j)_{j=1}^{m!}$ is uniformly distributed and the distance function is symmetric in the alternatives (i.e., the relabeling of the alternatives would not change the distances), then $\underline{d} = \overline{d}$.*

For instance, Corollary 1 assures for the Kendal τ distance under an IC that the expected distances of all anonymous SCRs tend to the same limit.

The next corollary admits generalizations of the Spearman footrule and the Spearman rank correlation.

Corollary 2. *Assume that the distance between a profile and a preference relation equals the sum of a function of rank differences, i.e.*

$$d(\Pi, \succ) = \sum_{i=1}^{n} \sum_{j=1}^{m} f(rk[a_j, \succ_i] - rk[a_j, \succ]),$$

and that for each alternative the probability that an alternative is ranked lth equals $1/m$ (that is we have uniform distributions above the set of ranks). Then the expected distances of all anonymous SCRs tend to the same limit.

Proof. The statement follows from the observation that each rank difference occurs with the same probability for each alternative, and that therefore the functional form of f does not play a role. □

Note that the IC and the IAC assumptions both imply a uniform distribution over the set of alternatives, and therefore Corollary 2 covers both cultures.

Finally, we conclude this section with an example showing that we can have $\underline{d} = \overline{d}$ even in case of a non-uniform distribution. Let $n = 3$, the probability of the occurrence of ordering o_1 be $1/4$ and the probabilities of the occurrences of the remaining five orderings be $3/20$. Furthermore, let $d(o_1, o_k) = 6/5$ for all $k = 2, \dots, 6$ and $d(o_l, o_k) = 1$ for all $l \neq k$ and $l, k = 2, \dots, 6$. Then

$$\overline{d}(o_1) = 5 \cdot \frac{3}{20} \cdot \frac{6}{5} = 0.9 \quad \text{and}$$

$$\overline{d}(o_k) = 1 \cdot \frac{1}{4} \cdot \frac{6}{5} + 4 \cdot \frac{3}{20} \cdot 1 = 0.9$$

for all $k = 2, \dots, 6$.

4. Examples

Considering the Kendall τ, the Spearman rank correlation and the Spearman footrule distances, we determine the expected normalized average distances for the constant rule. In our examples we assume the IC case.

It will be helpful for us that Diaconis and Graham (1977) provide the maximum values for the Kendall τ, the Spearman footrule and the Spearman rank correlation distances between two linear orderings, which equal $(m-1)m$, $\lfloor m^2/2 \rfloor$ and $(m^3-m)/3$, respectively. For normalization purposes we let

$$C_K = (m-1)m, C_1 = \lfloor m^2/2 \rfloor \text{ and } C_2 = (m^3-m)/3.$$

First, considering the Kendall τ distance under the assumption that all preference profiles are equally likely, it can be easily verified that the normalized expected average distance of any constant rule equals $1/2$ since for any given preference relation $\succ$, when determining its expected Kendall τ distance, in exactly half of the cases we have that two distinct alternatives are ordered in the same way by $\succ$ as by the preferences in all profiles.

Second, turning to Spearman's rank correlation under the assumption of uniformly distributed ranks for each alternative as required in Corollary 2, we determine the normalized expected average distance for the constant rule when employing the distance d_2. Note that any alternative a_j is ranked lth with probability $1/m$ by any voter i. Without loss of generality we can assume that the constant rule is given by $a_1 \succ^* a_2 \succ^* \cdots \succ^* a_m$. The possible rank distances from alternative a_l are $l-1, l-2, \ldots, 1, 0, 1, \ldots, m-l$. If we take all alternatives into consideration, then there are $2(m-1), 2(m-2), \ldots, 2(m-l), \ldots, 2\cdot 2, 2\cdot 1$ ways such that the rank distances of an alternative in $\succ^*$ and $\succ_i$ equal $1, 2, \ldots, l, \ldots, m-2, m-1$, respectively. Therefore,

$$\begin{aligned}
\frac{\overline{d}}{C_2} &= \frac{1}{n}\sum_{i\in N}\frac{1}{(m!)^n C_2}\sum_{j=1}^{m}\sum_{\succ\in\mathcal{P}^n}(rk[a_j,\succ^*]-rk[a_j,\succ_i])^2 \\
&= \frac{1}{n}\sum_{i\in N}\frac{1}{(m!)C_2}\sum_{j=1}^{m}\sum_{\succ_i\in\mathcal{P}}(j-rk[a_j,\succ_i])^2 \\
&= \frac{1}{mC_2}2\left((m-1)1^2+\cdots+(m-l)l^2+\cdots+1(m-1)^2\right) \\
&= \frac{1}{mC_2}2\left(m-1+m2^2-2^3+\cdots+ml^2-l^3+\cdots\right. \\
&\quad \left.+m(m-1)^2-(m-1)^3\right) \\
&= \frac{1}{mC_2}2\left(m(1^2+2^2+\cdots+(m-1)^2)-(1^3+2^3+\cdots+(m-1)^3)\right) \\
&= \frac{1}{mC_2}2\left(m\frac{(m-1)m(2m-1)}{6}-\frac{(m-1)^2m^2}{4}\right) \\
&= \frac{3}{m+1}\left(\frac{2m-1}{3}-\frac{m-1}{2}\right)=\frac{1}{2}.
\end{aligned} \tag{4.11}$$

Finally, we determine the normalized expected average distance of the constant rule for the distance d_1 under the assumption of uniformly distributed ranks for each alternative as required in Corollary 2. Without loss of generality we can assume that $a_1 \succ^* a_2 \succ^* \cdots \succ^* a_m$. The possible rank distances from alternative a_l are $l-1, l-2, \ldots, 1, 0, 1, \ldots, m-l$. If we take all alternatives into consideration, then there are $2(m-1), 2(m-2), \ldots, 2(m-l), \ldots, 2\cdot 2, 2\cdot 1$ ways such that the rank distances of an alternative in $\succ^*$ and $\succ_i$ equal $1, 2, \ldots, l, \ldots, m-2, m-1$, respectively. Moreover, $\succ^*$ is fixed, while $\succ_i$ are drawn independently. Furthermore, each alternative is ranked lth with probability $1/m$ in $\succ_i$.

$$\begin{aligned}
\frac{\overline{d}}{C_1} &= \frac{1}{C_1}\frac{1}{m}2\left((m-1)+(m-2)2+\cdots+(m-l)l+\cdots+1(m-1)\right) \\
&= \frac{1}{C_1 m}2\left(m-1+m2-2^2+\cdots+ml-l^2+\cdots+m(m-1)-(m-1)^2\right) \\
&= \frac{1}{C_1 m}2\left(m(1+2+\cdots+(m-1))-(1^2+2^2+\cdots+(m-1)^2)\right) \\
&= \frac{1}{C_1 m}2\left(m\frac{1}{2}m(m-1)-\frac{(m-1)m(2m-1)}{6}\right) \\
&= \frac{m-1}{C_1}\left(m-\frac{2m-1}{3}\right)=\frac{m^2-1}{3\lfloor m^2/2\rfloor},
\end{aligned} \tag{4.12}$$

which equals $2/3$ if m is odd and $2/3(1-1/m^2)$ if m is even.

5. Concluding remarks

In this paper we followed the distance based approach to SCRs. Through our analysis we contributed, besides the mentioned approach, to the selection of a SCR in case of many voters. In general we provided an upper and a lower bound for the expected average distance of an anonymous SCR by appropriately chosen constant rules. Under further assumptions on the distribution of profiles we found an 'indifference result' by which we mean that in the limit all anonymous SCRs tend to the same expected average distance. We gave two sufficient conditions for our indifference result to hold. First, a symmetric distance function and the IC case is a sufficient condition. Second, assuming distance functions obtained through the sum of a function of rank differences under cultures which result in a uniform distribution over the set of alternatives is also sufficient. The second sufficient condition covers both the IC and IAC cases.

We outline possible further research directions. In our analysis we presented our results for a given number of alternatives m and let the number of voters n vary. Clearly, this is the more interesting case in the social choice context. However, in computer science applications the other case in which we fix n and let m vary can be equally interesting. Since m alternatives imply $m!$ rankings, this analysis is far less tractable. Already determining the Kemény–Young ranking is NP-hard (Bartholdi et al., 1989). Therefore, we expect fewer results from this research direction. Furthermore, additional limits, in which we tend only with m to infinity or with both m and n, may be determined.

CRediT authorship contribution statement

Dezső Bednay: Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Balázs Fleiner:** Formal analysis, Writing – review & editing. **Attila Tasnádi:** Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing.

Acknowledgments

We are grateful to the Editor-in-Chief, three anonymous reviewers, Ágnes Cseh and Andreas Orland for many helpful comments and suggestions on an earlier version of this paper.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Ambuehl, S., & Bernheim, B. D. (2024). *Interpreting the will of the people: social preferences over ordinal outcomes*: *Working paper, No. 395*, Zurich: University of Zurich, Department of Economics, http://dx.doi.org/10.5167/uzh-206634.

Arrow, K. (1951). *Social choice and individual values*. Wiley.

Ausloos, M. (2024). Hierarchy selection: New team ranking indicators for cyclist multi-stage races. *European Journal of Operational Research*, *314*, 807–816. http://dx.doi.org/10.1016/j.ejor.2023.10.044.

Bartholdi, J., Tovey, C. A., & Trick, M. A. (1989). Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare*, *6*, 157–165. http://dx.doi.org/10.1007/BF00303169.

Bednay, D., Moskalenko, A., & Tasnádi, A. (2017). Does avoiding bad voting rules lead to the good ones? *Operations Research Letters*, *45*, 448–451. http://dx.doi.org/10.1016/j.orl.2017.07.001.

Borodin, A., Roberts, G. O., Rosenthal, J. S., & Tsaparas, P. (2005). Link analysis ranking: Algorithms, theory, and experiments. *ACM Transactions on Internet Technology*, *5*, 231–297. http://dx.doi.org/10.1145/1052934.1052942.

Bortolussi, L., Dinu, L. P., & Sgarro, A. (2012). Spearman permutation distances and Shannon's distinguishability. *Fundamenta Informaticae*, *118*, 245–252. http://dx.doi.org/10.3233/FI-2012-712.

Bozóki, S., Csató, L., & Temesi, J. (2016). An application of incomplete pairwise comparison matrices for ranking top tennis players. *European Journal of Operational Research*, *248*, 211–218. http://dx.doi.org/10.1016/j.ejor.2015.06.069.

Burka, D., Puppe, C., Szepesváry, L., & Tasnádi, A. (2022). Voting: A machine learning approach. *European Journal of Operational Research*, *299*, 1003–1017. http://dx.doi.org/10.1016/j.ejor.2021.10.005.

Campbell, D. E., & Nitzan, S. I. (1986). Social compromise and social metrics. *Social Choice and Welfare*, *3*, 1–16. http://dx.doi.org/10.1007/BF00433520.

Cook, W. D. (2006). Distance-based and ad hoc consensus models in ordinal preference ranking. *European Journal of Operational Research*, *172*, 369–385. http://dx.doi.org/10.1016/j.ejor.2005.03.048.

Csató, L. (2023). A comparative study of scoring systems by simulations. *Journal of Sports Economics*, *24*, 526–545. http://dx.doi.org/10.1177/15270025221134241.

Diaconis, P., & Graham, R. L. (1977). Spearman's footrule as a measure of disarray. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *39*, 262–268. http://dx.doi.org/10.1111/j.2517-6161.1977.tb01624.x.

Dwork, C., Kumar, R., Naor, M., & Sivakumar, D. (2002). Rank aggregation revisited. http://www.cse.msu.edu/cse960/Papers/games/rank.pdf, (Accessed 6 February 2024).

Eckert, D., & Klamler, C. (2011). Distance-based aggregation theory. In E. Herrera-Viedma, J. L. García-Lapresta, J. Kacprzyk, M. Fedrizzi, H. Nurmi, & S. Zadrozny (Eds.), *Consensual processes, studies in fuzziness and soft computing* (pp. 3–22). Springer, http://dx.doi.org/10.1007/978-3-642-20533-0_1.

Elkind, E., Faliszewski, P., & Slinko, A. (2015). Distance rationalization of voting rules. *Social Choice and Welfare*, *45*, 345–377. http://dx.doi.org/10.1007/s00355-015-0892-5.

Elkind, E., & Slinko, A. (2015). Rationalizations of voting rules. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, & A. D. Procaccia (Eds.), *Handbook of computational social choice* (pp. 169–196). Cambridge University Press, http://dx.doi.org/10.1017/CBO9781107446984.009.

Fagin, R., Kumar, R., Mahdian, M., Sivakumar, D., & Vee, E. (2016). An algorithmic view of voting. *SIAM Journal on Discrete Mathematics*, *30*, 1978–1996. http://dx.doi.org/10.1137/15M1046915.

Hadjibeyli, B., & Wilson, M. C. (2019). Distance rationalization of anonymous and homogeneous voting rules. *Social Choice and Welfare*, *52*, 559–583. http://dx.doi.org/10.1007/s00355-018-1156-y.

Kadziński, M., Miebs, G., Grynia, D., & Słowiński, R. (2022). Aggregation of stochastic rankings in group decision making. In T. Szapiro, & J. Kacprzyk (Eds.), *Collective decisions: theory, algorithms and decision support systems* (pp. 83–101). Springer, http://dx.doi.org/10.1007/978-3-030-84997-9_4.

Kemény, J. (1959). Mathematics without numbers. *Daedalus*, *88*, 577–591.

Laslier, J. F. (2011). And the loser is ... plurality voting. In D. S. Felsenthal, & M. Machover (Eds.), *Electoral systems – paradoxes, assumptions, and procedures* (pp. 327–351). Springer.

Lehrer, E., & Nitzan, S. (1985). Some general results on the metric rationalization for social decision rules. *Journal of Economic Theory*, *37*, 191–201. http://dx.doi.org/10.1016/0022-0531(85)90036-5.

Masthoff, J. (2004). Group modeling: Selecting a sequence of television items to suit a group of viewers. In L. Ardissono, A. Kobsa, & M. Maybury (Eds.), *Personalized digital television* (pp. 93–141). Kluwer Academic Publishers, http://dx.doi.org/10.1007/1-4020-2164-X_5.

Monjardet, B. (1997). Concordance between two linear orders: The Spearman and Kendall coefficients revisited. *Journal of Classification*, *14*, 269–295. http://dx.doi.org/10.1007/s003579900013.

Montes, I., Rademaker, M., Pérez-Fernández, R., & De Baets, B. (2020). A correspondence between voting procedures and stochastic orderings. *European Journal of Operational Research*, *285*, 977–987. http://dx.doi.org/10.1016/j.ejor.2020.02.038.

Saaty, T. L. (1980). *The analytic hierarchy process: planning, priority setting, resource allocation*. McGraw-Hill.

Volkovs, M. N., & Zemel, R. S. (2014). New learning methods for supervised and unsupervised preference aggregation. *Journal of Machine Learning Research*, *15*, 981–1022.