A Statistical Investigation of Social Choice Functions

Arkadii Slinko, Wayland Leung

Department of Mathematics
The University of Auckland
Private Bag 92019
Auckland NEW ZEALAND
fax: 64-9-3737457
emails: a.slinko@auckland.ac.nz
wayland_leung@mail.com

Abstract: We study social choice rules from the statistical point of view. We selected a set $\mathcal{S}$ of 19 most known social choice functions defined on the basis of a very diverse set of principles and we want to classify them on the basis of what they actually do. To this end, we generate a number of random profiles and, on each of them, we make a choice using all social choice functions one by one. This way, for each of the profiles, we get a partition of the set $\mathcal{S}$ into a set of disjoint classes such that in each class all rules determine the same winner while any two rules from two different classes choose different winners. Then all such partitions are averaged and this gives us a measure of closeness which reflects how often the rules were acting "similarly."

Running title: A Statistical Investigation of SCFs

1. Introduction

Given a group of individuals, group choice is usually understood as the reduction of individual preferences among objects in a given set to a single collective preference, or group preference. To do this one has to employ one of the existing social choice rules. The difficulty is that there are so many different social choice rules in the literature that the choice of a social choice rule has become difficult. One has to know which rules are more appropriate in their particular situation. The fact that various concepts are used in the definitions of the existing rules makes it difficult to compare them.

In this paper we study social choice rules from the statistical point of view. We selected a set $\mathcal{S}$ of 19 most known social choice functions defined on the basis of a very diverse set of principles. The existing classification that can be found in the literature is done on the basis of their defining concepts: point-scoring rules, rules based on the majority relation, rules based on the tournament matrix, etc. Instead, we want to classify them on the basis of what they actually do. To this end, we generate a number of random profiles and, on each of them, we make a choice using all social choice functions one by one. This way, for each of the profiles, we get a partition of the set $\mathcal{S}$ into a set of disjoint classes such that in each class all rules are unanimous in determining the winner while any two rules from two different classes choose different winners. Then all such partitions are averaged and this gives us a new partition based on the measure of closeness which reflects how often the rules are acting "similarly."

2. Social Choice Functions

We suggest that the society consists of a finite number n of individuals $S = \{v_1, \ldots, v_n\}$, whom we shall call voters. We shall consider their preferences on a finite set of alternatives $A = \{a_1, \ldots, a_m\}$ and we assume that they can be arbitrary linear orders. The set of all linear orders on A is denoted as $\mathcal{L}(A)$. All individual preferences taken together in an n-tuple $\overrightarrow{R} = (R_1, \ldots, R_n)$ form a profile. The set of all profiles will therefore be described by the elements of the cartesian product $\mathcal{L}(A)^n$.

Definition 1 *A Social Choice Function (SCF) is a mapping*

$$F \colon \mathcal{L}(A)^n \to A.$$

In other words, this is a function which assigns an alternative $a = F(\overrightarrow{R})$ *to any profile* $\overrightarrow{R}$*. Sometimes, mainly due to historical reasons, an SCF will be also called a rule.*

Since an SCF has to choose only one alternative while two or more alternatives might look similar to it, we need a certain rule to break ties. We suggest that a certain individual, let us call her the speaker, has a casting vote. This means that all ties are broken in favor of the alternative preferred by the speaker.

In Table 1 we list in alphabetical order all the 19 rules used in these experiments. These rules are commonly known or can be found in the book [7].

3. Probability Models for Voters' Preferences

We are going to look for similarities between the rules by means of computer experiments. So we need a society or a model of it to conduct the experiment, for the rules themselves do not assume any information indicating how individuals select their preferences. We have to adopt a certain priory assumptions i.e. the probability model to describe how individuals are distributed over the set of permissible linear orders. It is possible that the similarities sought for might depend on the model we select. The most popular models are based on each of the following conjectures. These are called:

- Impartial Culture Conjecture (IC)
- Anonimous Impartial Culture Conjecture (AIC)
- Real Society (RS)

In the first model voters are completely independent, in the second they are slightly dependent, in the third they are highly dependent. We choose the two extremes for our experiments, i.e IC and RS hypotheses. A good survey on probability models is [3].

Let us discuss our two models in some more details. Under the Impartial Culture Conjecture (IC) each voter chooses her preference order on the given set of alternatives A independently of the other voters and each voter is equally likely to hold any one of the $M = m!$ possible preference orders. In this case every profile from $\mathcal{L}(A)^n$ is equiprobable. This is a well-known model.

The next model which we will use is that of the "real society," where the voters' preferences are not independent because the voters are organized into a number of political parties.

To generate a set of voters belonging to the same party, we use the Metropolis algorithm [4]. We consider that a political party has a political platform, represented by a certain ranking σ_0 on the set of alternatives and we will assume that the members of this party have opinions scattered around this political platform. We will identify rankings with elements of the permutation group S_m.

In the preference studies a variety of nonuniform distributions on S_m is used. The Mallows model is often assumed. For this model a certain metric d on S_m is fixed and then the distribution has the form

$$\pi(\sigma) = \theta^{d(\sigma,\sigma_0)}/Z,$$

where $d(\cdot,\cdot)$ is the metric on S_m, σ_0 is the centering permutation, $0 < \theta \leq 1$ and $Z = Z(\theta, \sigma_0)$ is a normalizing constant. If $\theta = 1$, then we have a uniform distribution. If $\theta < 1$, then the distribution peaks at σ_0 and falls off exponentially as σ moves away from σ_0. A variety of metrics are used on S_m. The two most popular ones are: the Cayley distance, where $d(\sigma, \sigma_0)$ is the minimum number of transpositions required to bring σ to σ_0, and Spearman's footrule, for which

$$d(\sigma, \sigma_0) = \sum |\sigma(i) - \sigma_0(i)|.$$

We found that the last one is more adequate in describing opinions. For example, if in S_6 we have $\sigma_0 = \mathrm{id}$, then permutations $\sigma_1 = (12)$ and $\sigma_2 = (16)$ will be equally close to σ_0 in the Cayley metric as $d(\sigma_1, \sigma_0) = d(\sigma_2, \sigma_0) = 1$, while we will get $d(\sigma_1, \sigma_0) = 2$ and $d(\sigma_2, \sigma_0) = 10$ with the Spearman's footrule. The permutation σ_1 represents a swapping the first two preferences, which is not a radical change of opinion, while the permutation σ_2 means that the first and the last preferences were swapped, which is quite a radical change. Hence we used the Spearman's footrule as metric for S_m.

Since the normalising constant is impossible to calculate, the samples from π are normally drawn by the Metropolis algorithm from the base Markov chain of random transpositions. The chain is defined as follows. The initial state is σ_0. If the chain is currently at σ, then the chain proceeds by choosing a random transposition (ij) and computing $\sigma' = (ij)\sigma$. If $d(\sigma', \sigma_0) \leq d(\sigma, \sigma_0)$, the chain moves to σ'. If $d(\sigma', \sigma_0) > d(\sigma, \sigma_0)$, a coin is flipped with probability

$\theta^{d(\sigma',\sigma_0)-d(\sigma,\sigma_0)}$. If this comes out heads, the chain moves to σ'. Otherwise, the chain stays at σ. For $k \approx n \ln n$ the kth power of this Markov chain is very close to π (see [4], [5]). For example, for $m = 14$ we make 50 iterations of the chain to get a permutation from the distribution π.

Our "real society" consists of supporters of three political parties and independent voters. The two major political parties have 35% and 30% of the vote and the minor party has 20%. The other voters are independent.

To generate a profile belonging to such a "real society" we generate three random permutations σ_0, τ_0, ζ_0 and they are used as the centering permutations of the three distributions. We draw members of the three political parties from these distributions and we randomly draw independent voters from the uniform distribution.

4. Description of The Experimental Scheme

In this section we will describe how we determine which rules are close and which are not. Given a single profile, it is easy to decide this, namely, any two rules which suggest the same winner are close and any two rules which give different winners are not. Essentially what we obtain given a single profile is a partition of the rules. Having m alternatives we have m classes in this partition: in the ith class we gather all rules which suggest the ith alternative as the winner. Some of these classes may be empty.

But a single profile does not give us a complete understanding of what a rule does since the partition can be specific to this particular profile, for example all voters can be unanimous and all submit the same ranking. We need a representative set of profiles (whatever it may mean) to make conclusions about similarity or dissimilarity of the rules. Hence we need to try to use all our rules on each of the profiles of our representative set, obtain the respective partitions and then average them.

We considered that any set of 1000 profiles is representative. Although this number is rather a result of the limits imposed by our computers than a result of our theoretical considerations, experiments showed that for 1000 profiles the results are pretty stable and change very little, if at all, from one run to another.

The averaging of partitions may be performed, for example, as suggested in [8]. Instead of operating with partitions we operate with the matrices of their respective equivalence relations. Thus if the set $\mathcal{R}$ of rules is partitioned into m classes

$$\mathcal{R} = \mathcal{R}_1 \cup \ldots \cup \mathcal{R}_m$$

then we assign to this partition S an $m \times m$ matrix M for which $m_{ij} = 1$ if the ith and the jth rules are in the same class of this partition and $m_{ij} = 0$ otherwise.

Let φ_δ be the threshold function, i.e., $\varphi_\delta(x) = 0$, if $x < \delta$ and $\varphi_\delta(x) = 1$, if $x \geq \delta$. This threshold function can be extended to matrices: if $A = (a_{ij})$, then $\varphi_\delta(A) = (\varphi_\delta(a_{ij}))$.

Suppose now that we have N partitions $S_1, \ldots, S_N$ of $\mathcal{R}$ (each corresponds to a random profile) and we want to average them. Unlike the situation with numbers the resulting averaged partition will depend on the threshold parameter $0 \leq \delta \leq 1$. Let $M_1, \ldots, M_N$ be the matrices of the partitions $S_1, \ldots, S_N$. We form the matrix $T = \frac{1}{N}(M_1 + \ldots + M_N)$. Then $T_\delta = \varphi_\delta(T)$ corresponds to a certain relation and after taking the transitive closure of it we obtain an equivalence relation M_δ, which corresponds to a certain partition S_δ. Clearly, if $\delta_1 < \delta_2$, then S_{δ_2} is a refinement of S_{δ_1}, so, in fact, we obtain a tree of subsets in which every level represents a certain partition.

Computationally, M_δ can be computed as follows:

$$M_\delta = \left[\varphi_\delta \left(\frac{1}{N} \sum_{i=1}^{N} M_i \right) \right]^{r-1},$$

where r is the cardinality of $\mathcal{R}$, the addition of matrices is standard, and the multiplication is given by the rule: if $A = (a_{ij})$ and $B = (b_{ij})$, then $AB = (c_{ij})$, where

$$c_{ij} = \max_k \min(a_{ik}, b_{kj})$$

A few words must be said about the threshold parameter. In our analysis it plays the most crucial role. When we fix this parameter, we get a partition. But no one particular δ gives all the information. When it is very low, we will get all the rules in one cluster, when it is close to 1, we get 19 different clusters, each consisting of a single rule. When we increase δ from 0 to 1 we get all possible partitions, the whole tree of them.

Two different situations might happen. The first possibility is that at some stage the whole set of rules splits into two clusters of comparable size, then we may say that a meaningful clusterization is observed. The second possibility is that we will always have one big cluster and a collection of singletons which split from the main group one by one or in small unstable groups. In the latter case we observe something like a solar system: a certain "core" and outliers. In this case, the order in which rules split from the "core" and the values of the corresponding threshold parameters are of special interest as they show the "distance" of the rules from the "core." The first rule to split from the main group will be the most distant from the remaining rules, the rules that split late are the most "central" ones. We give an example of such an analysis in the Appendix.

5. Results of Computer Experiments

The summary of our computations are presented in the following Table 3. We discuss these results below.

1. Impartial Culture Conjecture. The first significant result of our computer experiments is that the given set of rules cannot be split into clusters, rather we have a set of core rules and several independent outliers which deviate from the core to a different degree. The Antiplurality rule is the first to split at the level of threshold parameter 0.26. This can be expected because the Antiplurality rule uses the information contained in the lower row of the profile (written as a table where the last row contains all last preferences of the voters) which has little correlation with the information contained in the upper half of the profile which most rules use. The next one to split at the level 0.38 is the Approval voting with random number of approvals. This can also be explained. The average number of approved alternatives will be $m/2 = 7$ and therefore the positions in the upper half of the table are more or less equivalent which leads to a low correlation with the importance of occupying top positions and hence low correlation with the other rules for which top positions are more important. The next to split at the level of 0.42 is a small cluster consisting of the Plurality and the Run-off Procedure which remains relatively stable and splits only at the level of 0.56. Before this cluster is split, there will be two other rules splitting from the main group, namely the Approval voting with 3 approvals at the level of 0.44 and and Hare's rule (STV) at 0.52. Coombs procedure is the next to go at 0.58. Then almost simultaneously the Majoritarian Compromise and the cluster of the Top Cycle and the Uncovered Set split at 0.61 and 0.62, respectively. The Top Cycle and the Uncovered Set keep relatively close which might partially be explained by the tie-breaking procedure involved. Indeed, as Bell proved [1, 2], when $m \to \infty$, $n \to \infty$, the probability that all alternatives belong to the Top Cycle is approaching 1. This means that at some stage, for very large m and n, the Top Cycle rule selects the first alternative with probability approaching 1. The same is probably true for the Uncovered Set.

The remaining "core" remain relatively stable up to the level 0.74, let us have a look at its composition:

- Black's rule;
- Borda rule;
- Copeland rule;
- Inverse Borda rule;
- Long path rule;
- Markov rule;
- Nanson's procedure;
- Simpson's procedure;
- Variant of Borda rule;

We see here three distinct groups here: the Borda cluster (Borda, Inverse Borda, Nanson's and Variant of Borda), the Condorcet cluster (Copeland, Long path, Markov), the Black's rule which belongs to both of these worlds and a separate Simpson's procedure. At the level of 0.75 the Simpson's procedure splits and at the level of 0.76 the whole core disintegrates into the Borda group (Black's, Borda, Variant of Borda), the Inverse Borda group (Inverse Borda and Nanson's) and the Condorcet group (Copeland, Long path, Markov). Black's rule joins the Borda cluster. It is consistent with the calculations of Gehrlein [6], where the probability of having a Condorcet winner for 14 alternatives and 101 voter was estimated as approximately 0.42. Our calculations give the value 0.44. It means that most of the time

Black's procedure will function as the Borda rule. These three groups remain relatively stable.

2. The "Real Society." As Williamson and Sargent showed [9] the probability of having a Condorcet winner increases dramatically when we deviate even slightly from the Impartial Culture assumption. Therefore we can expect the results to be different because of that. We see that Antiplurality splits even earlier this time at the level of 0.22 and the remaining rules remain in one group much longer, up to the level of 0.57, when the Approval voting with fixed number ($k = 3$) of approvals splits from it. This is rather surprising since under IC this rule splits later than the Plurality and the Run-off but for this case it splits earlier. But the greatest surprise we get when at the level of 0.62 the Majoritarian Compromise splits earlier than any other remaining rule. Under IC it split exactly at the same level (0.61) which is also unusual because normally in the latter case rules split at a higher level than under IC. Another anomaly is the behavior of the Plurality and the Run-off. They are no longer together. The Plurality now splits much earlier at 0.65 while the Run-off sits in the "core" up to the level of 0.7, when it splits together with the Hare's Rule. Thus we see that under IC the Run-off is similar to the Plurality while under RS its multistage elimination nature brings it closer to the Hare's rule.

At the level of $0.7-0.71$ the same stable "core" emerges as for IC. The only difference is the presence of the Coombs procedure there which splits now much much later (at 0.76. Apart from that the same pattern is maintained. The "core" is stable up to the level of 0.87 after which it disintegrates with forming three groups: the Borda group (Borda and Variant of Borda), the inverse Borda group (Inverse Borda, Nanson's) and the Condorcet group (Copeland, Long path, Markov). The Black's procedure is no longer a part of the Borda group which is consistent with the Williamson and Sargent results quoted earlier.

3. Conclusion The most significant outcome of our computer experiments is the observation that under both IC and RS hypotheses:

- the set of rules cannot be split into clusters, rather we observe a set of "core" rules and several independent outliers which deviate from the "core" to a different degree;

- The "core" consists of the rules which are 100% Condorcet efficient (i.e. select the Condorcet winner when there is one) and the value of the threshold parameter at which a particular rule splits highly correlates with the Condorcet efficiency of the rule. This link needs further investigation.

- Among the scoring rules the Borda rule and its variants are the closest to the "core."

The following rules behave differently under IC and RS hypotheses (the difference of the threshold parameters is high):

- Approval voting with random number of approvals -0.30;
- Run-off -0.28;
- Top Cycle -0.25;
- Uncovered set -0.25;
- Black's rule -0.24;
- Plurality -0.23.

The following rules are almost unaffected by switching from IC to RS (the difference of the threshold parameters is low):

- Markov rule 0.0;
- Copeland rule 0.00;
- Majoritarian compromise -0.01;
- Long path rule 0.03;
- Antiplurality rule 0.04.

The List of Rules Used in the Experiments
(in alphabetical order)

1. Antiplurality rule;
2. Black's rule;
3. Borda rule;
4. Coombs procedure;
5. Copeland rule;
6. Hare's rule (STV);
7. Inverse Borda rule;
8. Long path rule;
9. Majoritarian Compromise;
10. Markov rule;
11. Nanson's procedure;
12. Plurality rule;
13. Approval voting with random number of approvals;
14. Run-off procedure;
15. Simpson's procedure;
16. Top cycle rule;
17. Uncovered set rule;
18. Approval voting with fixed number ($k = 3$) of approvals;
19. Variant of Borda rule ($k = 3$);

Table 1

The Table of Threshold Parameters for IC and RS hypotheses

Rule	IC	RS	Difference	Condorcet Efficiency
Antiplurality	0.26	0.22	+0.04	0.25
Approval (random)	0.38	0.68	-0.30	0.42
Plurality	0.42	0.65	-0.23	0.30
Run-off	0.42	0.70	-0.28	0.49
Approval (fixed)	0.44	0.57	-0.13	0.48
Hare's rule (STV)	0.52	0.70	-0.18	0.70
Coombs procedure	0.58	0.76	-0.18	0.72
Majoritarian compromise	0.61	0.62	-0.01	0.63
Top cycle	0.62	0.87	-0.25	1.0
Uncovered set	0.62	0.87	-0.25	1.0
Simpson's procedure	0.74	0.91	-0.17	1.0
Black's rule	0.75	0.89	-0.24	1.0
Borda rule	0.75	0.88	-0.13	0.85
Variant of Borda rule	0.75	0.88	-0.13	0.82
Inverse Borda rule	0.76	0.91	-0.15	1.0
Nanson's procedure	0.76	0.91	-0.15	1.0
Long path rule	0.99	0.96	0.03	1.0
Copeland rule	0.99	0.99	0.00	1.0
Markov rule	0.99	0.99	0.00	1.0

Table 2

References

[1] Bell,C.V.: (1978) What happens when majority rule breaks down. Public Choice 33: 121–126.

[2] Bell,C.V.: (1981) A random voting graph almost surely has a Hamiltonian cycle when the number of alternatives is large. Econometrica 49: 1597–1603.

[3] Berg,S. and Lepelley,D.: (1994), On Probability Models in Voting Theory. Statist. Neerlandica 48: 133–146.

[4] Diakonis,P. and Saloff-Coste,L.: (1998) What do we know about the Metropolis Algorithm? Journal of Computer and System Sciences 57: 20–36.

[5] Diakonis,P.: (1988) Group Representations in Probability and Statistics. Lecture Notes - Monograph Series, 11, Institute of Mathematical Statistics, Hayward, California.

[6] Gehrlein,W.V.: (1983) Condorcet's Paradox. Theory and Decision 15: 161–197.

[7] Laslier, J-F. (1997) Tournament Solutions and Majority Voting. Springer-Verlag. Berlin-Heidelberg.

[8] Slinko,A.: (1986) Some algebraic operations over classification algorithms and their applications. USSR Comput. Math. and Math. Phys. 25: 169-177.

[9] Williamson,O.E. and Sargent,T.J.: (1967) Social Choice: A Probabilistic Approach. Economic Journal 77: 797–813.

6. Appendix: Example

Here we give a detailed example to show how the averaging of partitions works. Suppose we have the following five partitions of $\{1, 2, 3, 4\}$:

$$
\begin{aligned}
S_1 &: \{1,2,3,4\}, \emptyset, \emptyset, \emptyset \\
S_2 &: \{1,2,4\}, \{3\}, \emptyset, \emptyset \\
S_3 &: \{1\}, \{2,3,4\}, \emptyset, \emptyset \\
S_4 &: \{1,3\}, \{2,4\}, \emptyset, \emptyset \\
S_5 &: \{1\}, \{2\}, \{3,4\}, \emptyset
\end{aligned}
$$

Then $T = \frac{1}{5}(M_1 + M_2 + M_3 + M_4 + M_5)$ will look like

$$
T = \begin{pmatrix} 1 & 0.4 & 0.4 & 0.4 \\ 0.4 & 1 & 0.4 & 0.8 \\ 0.4 & 0.4 & 1 & 0.6 \\ 0.4 & 0.8 & 0.6 & 1 \end{pmatrix}.
$$

If $0 < \delta \leq 0.4$ we get

$$
T_\delta = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix},
$$

which is the matrix of the partition $\{1, 2, 3, 4\}$. For $0.4 < \delta \leq 0.6$ we get

$$
T_\delta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix},
$$

which is the relation which transitive closure has the matrix

$$
T_\delta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix},
$$

which is the matrix of the partition $\{1\}, \{2, 3, 4\}$. For $0.6 < \delta \leq 0.8$ we get

$$T_\delta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix},$$

which is the matrix of the partition $\{1\}, \{3\}, \{2, 4\}$. For $0.8 < \delta \leq 1$ we get

$$T_\delta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

which is the matrix of the partition $\{1\}, \{2\}, \{3\}, \{4\}$. As a result we get the tree presented below.