COMSOC HW 3: Report on Question 5

Max Rapp, Haukur Jónsson, Silvan Hungerbühler

1 Introduction

It is clear from the Gibbard-Satterthwaite Theorem that any reasonable voting rule is in principle prone to manipulation. That is, there will be profiles for which some voters stand to improve the outcome of the vote by reporting dishonest preferences. It is unclear, however, how often such situations will arise in practice. We therefore assess and compare the *Borda*, *Plurality* and *Copeland* rules with respect to their practical proneness to manipulability.

There are various factors that could potentially influence the prevalence of manipulability: the number of candidates partaking in the election, the voting rule employed, the number of voters in the profile and assumptions concerning the distribution over possible profiles. Due to time constraints we are unable to tinker with the last parameter, so we have not analyzed it. We stick to using the impartial culture assumption throughout this report. In order to individually account for the remaining three factors we set up a logistic regression model. This allows us to separately account for each factor's effect on the expected probability of a given profile-cum-voting rule being manipulable. Moreover, it provides us with confidence intervals for the effects we observe; that is, we obtain some measure of how robust our results are.

Using algorithms, described in Section 2, we generated a data set in the following way: First, we randomly generated 40000 profiles for 3, 6, 8 and 10 candidates. For each number of candidates 100 profiles with between 10 and 1000 voters (in interval of 10). Secondly, we determined the winner in all of the profiles under each voting rule. Then we checked for each profile and voting rule whether some voter in the profile under consideration a) stands to gain from successful manipulation and b) has the powers to induce such a beneficial manipulation. If both points are satisfied, then we declare the outcome manipulable, otherwise, we declare it non-manipulable. For each profile-voting rule pair we thus get a binary verdict: manipulable or not. The fraction of manipulable profiles under some voting rule is the voting rule's practical proneness to manipulation.

With the data at hand we first give a descriptive overview over the respective proneness of each voting rule for different numbers of voters. Specifically, we consider the four types of elections: 3, 6, 8 and 10 candidates. For each of the four types we compute the fraction of profiles that are manipulable out of 100 profiles for all such profile-bundles with between 10 and 1000 voters (in intervals of 10). Ensuingly, we undertake a logistic regression analysis where we try to disentangle which of the three factors we identified contributes to which degree to likely manipulability. The main focus of our analysis, however, still lies on the effect the number of voters has on the manipulability of a given voting rule.

The report is structured as follows: Section 2 provides some detail on the algorithms we used to determine the winners for a given profile-voting rule pair and whether said outcome can be manipulated. Section 3 presents a descriptive analysis of the data we generated and a logistic regression analysis to account for different factors separately. Also, it discusses the findings from and concludes the report.

2 Determining Winners & Manipulability Under Each Voting Rule

This Section sketches the algorithms we used to establish which candidate wins an election under a given voting rule for some profile and whether manipulation can occur.

2.1 Determining Winners

Given some profile **R** we determined the winner by assigning each candidate a counter on which we put points. The candidate with most points wins, ties are broken lexicographically. The points were awarded thus for the three rules:

- *Plurality*: A candidate's points is the number of times the candidate shows up in the first spot. Essentially, this is a PSR with the scoring vector $s = (1, 0, ..., 0)$.
- *Borda*: The points of a candidate are her Borda-score, assuming the standard Borda scoring vector.
- *Copeland*: A candidate's points are the number of pair-wise matchups a candidate wins minus the number of pair-wise matchups a candidate loses. (We view this rule effectively as a positional scoring rule. The scoring vector depends on the number of candidates: $|X| = x$. It looks as follows: $s = (x-1, x-3, x-5, ..., -(x-3), -(x-1))$.)

2.2 Determinining Manipulability

The algorithms to determine the manipulability run thus:

- *Plurality*:
  1. Compute the respective margin of victory (MOV) between *actual winner* and all other candidates. If MOV>1 for all candidates, then declare the profile-rule pair non-manipulable. (One invdividual voter whose first option is not the winner can at most 'push' another candidate by 1 point. If the MOV is larger than this she can not manipulate successfully.) Put all the candidates for whom MOV<2 in a set of *potential winners*. These are the candidates in 'striking distance' who could still be manipulated into winning.
  2. For all candidates in *potential winners* look at each voter's profile and see if two conditions hold:
     a) The *potential winner* is strictly preferred to the *actual winner* by the voter under consideration.
     b) The *potential winner* was not already ranked in the first position by the

voter.
If both conditions hold for some voter, then declare manipulability. Otherwise, declare non-manipulability.

- *Borda*:

  1. Compute the MOV for between all candidates and the *actual winner*. If MOV$> x-2$ for all candidates, then declare the profile-rule pair non-manipulable. (One invdividual voter whose first option is not the winner can at most 'push' another candidate by $x-2$ point. If the MOV is larger than this she can not manipulate successfully.) Put all the candidates for whom MOV$< x-1$ in a set of *potential winners*. These are the candidates in 'striking distance'.
  2. For all candidates in *potential winners* look at each voter's profile and see if three conditions hold:
  a) The *potential winner* is strictly preferred to the *actual winner* by the voter under consideration.
  b) The *potential winner* was not already ranked in the first position by the voter.
  If some condition fails to hold, move to next candidate. Otherwise, proceed in the following way:
  Move the *potential winner* all the way to the top of the ballot under consideration. Put all the hopeless candidates (those who neither win nor are part of the *potential winners* right below the candidate we are currently trying to promote. At the bottom of the ranking try all possible permutations of ranking the remaining candidates. For each possible permutation compute the winner. If the *potential winner* manages to win instead of the *actual winner*, declare the profile manipulable. If for all candidates and all voters thus considered such a permutation cannot be found, declare non-manipulability.
  [1]
  3. If we find one candidate in the *possible winners* that a) some voter prefers to actual winner b) will become the winner by moving him to the top and the *actual winner* at some spot behind what he previously occupied, then declare the profile manipulable. Otherwise, declare it non-manipulable.

[1] Perhaps by way of clarification, we cannot simply put the *actual winner* all the way to the bottom when checking for manipulability but need to proceed by considering the permutations one-by-one because of situations like the following:
Assume there are five candidates. The three candidates at the top of the Borda-score ranking - A,B,C - have this many points (α being some positive integer):
A: α
B: $\alpha-2$
C: $\alpha-1$
Voter i has the following ranking:
some candidate $\succ B \succ A \succ$ some candidate $\succ C$
By moving B to the top, B now has $\alpha-1$ many points. If i moves A back one spot, then A now has $\alpha-1$ points too. We define this as having manipulated successfully, since the winner now depends on the tie breaking we perform. But if i were to move A further back still, then C would have to move up, thus obtaining α many points and winning clearly. But C is dispreferred by i, so this manipulation is not desirable for her.
In order to exclude cases like this one, we therefore need to check at each step what happens with the other candidates that could also become winners. On the other hand, we can move B up with no further considerations, because this will at most hurt other candidates' Borda-score.

- *Copeland*: This works essentially the same as what we use for Borda-manipulability. The only difference is that, instead of the Borda scoring vector, we use the one we induced to determine the winner for the specific profile under consideration.

3 Descriptive Analysis & Logistic Regression

Figures 1, 2 and 3 show the frequencies of manipulable profiles for 100 profiles with between 10 and 1000 voters.

3.1 Descriptive Analysis

As the plots already give away, there is no difference discernible between the frequency of Borda and Copeland manipulability. We therefore represent them with the same data points. The table following immediately after summarizes the three plots depicted, plus one more which we did not present here, by giving the average fraction of manipulable profiles across all number of voters for the voting rules.

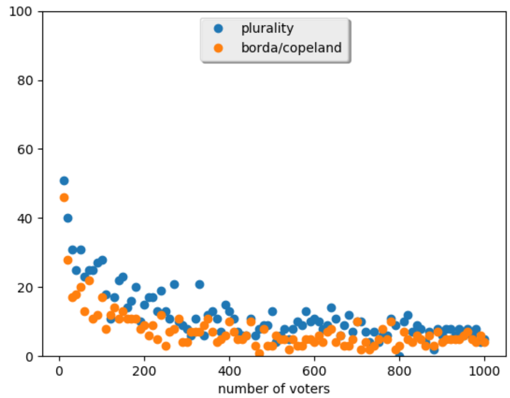

Figure 1: Number of manipulable profiles on the x-axis plotted against the number of voters on the y-axis, for each voting rule. Each datapoint represents 100 profiles for 3 candidates.

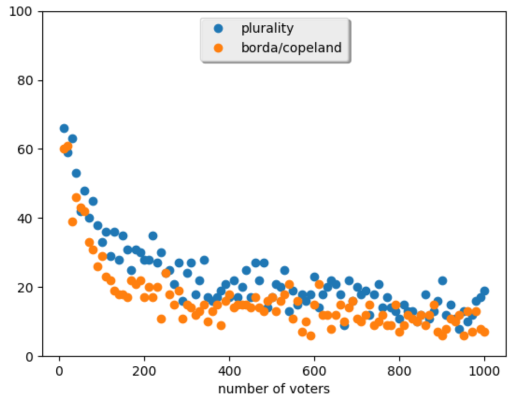

Figure 2: Number of manipulable profiles on the x-axis plotted against the number of voters on the y-axis, for each voting rule. Each datapoint represents 100 profiles for 6 candidates.

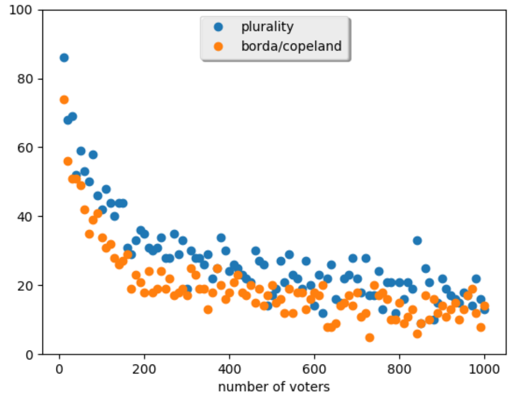

Figure 3: Number of manipulable profiles on the x-axis plotted against the number of voters on the y-axis, for each voting rule. Each datapoint represents 100 profiles for 8 candidates.

Table 1: Average Manipulable Fraction of Profiles for 3, 6, 8 and 10 Candidates

Average Manipulable Fraction	3 Candidates	6 Candidates	8 Candidates	10 Candidates
Plurality	0.1195	0.2264	0.2752	0.3137
Borda/Copeland	0.075	0.1634	0.1983	0.219

The data suggests that the more voters a profile has, the less prone to manipulation it becomes. This is indicated by both Figure 1 and 2 where the fraction of manipulable profiles clearly decreases for all voting rules as we increase the number of voters. Intuitively this accords well our choice to focus on the scoring-aspect of voting rules in our construction of algorithms in Section 2. The more voters there are, the more points can be handed out. The more points distributed across the candidates, the bigger the variance between the candidates' final score is likely to be. But if these distances are great, it becomes harder for individual voters to manipulate the outcomes. Thus, we see a decrease in manipulability for higher number of voters; or so our attempt at explaining the data goes.

A second observation concerns the effect the number of candidates has on manipulability. The Table 1 provides some, albeit very limited, evidence that more candidates imply increased manipulability. Whereas for 3 candidates the average fraction of manipulable profiles under the plurality rule is at roughly 0.12, for 6 candidates the number is almost around 10 percentage points higher. For Borda and Copeland the difference is even more marked. For 6 candidates the fraction is more than double as large. However, by comparing diifferences betwen 6, 8 and 10 candidates we can gauge that the marginal increase in manipulability is positive, yet decreasing. We conjecture that it will converge to some maximal fraction of manipulability for each voting rule eventually. Unfortunately, we were unable to pursue this hypothesis further, because the computations needed to generate our data quickly jumped to unmanageable levels as we increased the number of candidates .

Lastly, our data at this stage gives good evidence that Copeland and Borda manipulability are highly correlated, if not completely equivalent. At least for the data generated and analyzed, a profile was always manipulable under either both Borda and Copeland or under none of the two. This either points to a fundamental similarity of the two or to a flaw in the construction of our algorithms.

3.2 Logistic Regression

We used the algorithms described earlier to generate a large amount of data on manipulability. Next, we would like to know how probable it is that a profile with a given number of candidates and voters is manipulable under our three voting rules. Furthermore, we would like to compare the impact the choice of voting rule has on manipulability and how high our confidence in this impact should be as compared to pure accident. Finally, we would like to compare the importance of voting rules with respect to manipulability to the impact of other factors such as properties of the electorate.

The adequate tool to answer these questions is regression analysis. Since our dependent variable (manipulable? yes/no) is a binary, categorical variable, we should not use simple linear regression since it is unlikely to yield a good fit. Instead the method of choice is logistic regression.

Logistic regression makes use of the standard logistic function to turn binary output into real output. Namely

$$F(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+...)}}$$

yields a regression model where F(x) can be interpreted as the probability of the dependent variable given the a linear combination of the independent variables.

Via the inverse "logit"-function we regain a linear regression model

$$\beta_0+\beta_1 x_1+... = \log(\frac{F(x)}{1-F(x)}) = \beta^T x$$

Regression can then be carried out for example via gradient descent using

$$\beta_j := \beta_j - \frac{\alpha}{m}\sum_{i=1}^{m}(\frac{1}{1+e^{-(\beta^T x)}} - y_i) * x_{ij}$$

where α is a suitable learning rate. The results obtained using this method largely confirm our hypotheses derived from the plots: Table 1 shows that plurality appears to be the voting rule most vulnerable to manipulation. A slight anomaly is that the Borda rule is predicted to be more vulnerable than the Copeland rule by the regression model which is at odds with our earlier findings of same outcomes for those two rules.

Table 2: Probability of Manipulability Conditioned on Voting Rules

Voting Rule	P(1)	95% Confidence Interval
Plurality	0.217	0.209-0.224
Borda	0.183	0.179-0.187
Copeland	0.154	0.147-0.160

An interesting result is that the number of voters seems to have a stronger influence than the choice of voting rule if it becomes large enough as can be seen in table 2. This suggests that large elections are somewhat less easy to manipulate. However, as table 3 shows, this effect is offset if additional candidates are introduced: as suspected, more candidates lead to more possibilites for manipulation.

Finally, the highest resistance to manipulation was achieved for large electorates with few candidates using the Copeland rule. On the other hand, the combination of plurality vote with a large number of candidate, as is frequently used in modern

Table 3: Probability of Manipulability Conditioned on Number of Voters

#Voters	P(1)	95% Confidence Interval
10	0.3357	0.324-0.348
...	...	...
1000	0.0789	0.074-0.084

democracies, proved to be particularly vulnerable to manipulation. In general, the importance of voting rules pales in comparison to properties of the electorate: Table 4 shows that the highest and lowest vulnerability values are almost achieved when only considering the numbers of voters and candidates.

Table 4: Probability of Manipulability Conditioned on Number of Candidates

#Candidates	P(1)	95% Confidence Interval
3	0.097	0.094-0.100
6	0.167	0.165-0.170
8	0.234	0.230-0.239

Table 5: Highest&Lowest Vulnerability to Manipulation With and Without Voting Rule

Voting Rule	#Voters	#Candidates	P(1|Voting Rule, #Voters, #Candidates	P(1|#Voters, #Candidates)
Plurality	10	8	0.469	0.4122
Copeland	1000	3	0.029	0.0374