The Probability of Safe Manipulation

Mark C. Wilson and Reyhaneh Reyhani

Abstract

The concept of safe manipulation has recently been introduced by Slinko and White. We show how to compute the asymptotic probability that a safe manipulation exists for a given scoring rule under the uniform distribution on voting situations (the so-called Impartial Anonymous Culture). The technique used is computation of volumes of convex polytopes. We present explicit numerical results in the 3-candidate case.

1 Introduction

The Gibbard-Satterthwaite theorem [2, 5] shows that for each nondictatorial social choice function allowing unrestricted preferences of voters over alternatives and such that the $m \geq 3$ alternatives can each win in some profile, there always exists a profile which is unstable. In other words, in the voting game with ordinal utilities given by the voter preferences of that profile, the strategy where all voters express their sincere preference is not a Nash equilibrium, so that at least one voter has incentive to deviate unilaterally by expressing an insincere preference. For common social choice functions, the probability that a single individual can succeed in changing the election result converges to zero as n, the number of voters, tends to ∞. Thus the question of coalitional manipulation is more interesting.

Coalitions must be of fairly large size in order to manipulate effectively. For example, under the IC hypothesis (uniform distribution on profiles) the manipulating coalitions are typically of order $\sqrt{n}$, while they can be considerably larger under other preference distributions [7, 6]. Thus the question of coalition formation becomes important, because there are substantial coordination difficulties to be overcome in order to manipulate successfully.

Slinko and White [8] proposed a simple model for coalition formation, whereby a "leader" publicizes a strategic vote and voters sharing the leader's preference order decide whether to follow this strategy or vote sincerely. As a topic for further research, [8] lists the study of the probability that such an attempt succeeds sometimes and the coalition members never fare worse than with the sincere outcome. The present paper studies this topic for a well-known preference distribution, namely the Impartial Anonymous Culture.

2 Definitions and basic properties

Let $m \geq 1$ be an integer and let $\mathcal{C}$ be a set of size m, the set of **alternatives** (or **candidates**). Let $n \geq 1$ be an integer and let $\mathcal{V}$ be a set of size n, the set of **agents** (or **voters**). Each agent is assumed to have a total order of the alternatives, the agent's **preference order**. An agent a **strongly prefers** alternative i to alternative j if and only if i is strictly above j in a's preference order; if we also allow the possibility $i = j$ then we just use the term **prefers**. There are $M := m!$ possible such preference orders, which we call **types**. We denote the set of all types by $\mathcal{T}$ and the set of all agents of type t by $\mathcal{V}_t$. A multiset from $\mathcal{T}$ with total weight n is a **voting situation**, whereas a function taking $\mathcal{V}$ to $\mathcal{T}$ is a **profile**. Each voting situation corresponds naturally to several profiles, corresponding to the different permutations of the multiset.

Let F be a social choice function, a map that associates an element of $\mathcal{C}$ to each profile. If this map depends only on the voting situation, then the rule is called **anonymous**.

The **Impartial Anonymous Culture** (IAC) is the uniform probability distribution on the set of voting situations. If F is anonymous, then we can compute the probability of an event under IAC simply by counting voting situations. Since voting situations can be encoded by tuples of natural numbers $(n_1, \ldots, n_M)$ with $\sum_i n_i = n$, this amounts to counting lattice points in a subset of a dilated standard simplex.

In the following definitions it is assumed that agents not mentioned continue to vote sincerely.

Definition 1. *A voting situation is **manipulable** if there is some subset X of voters such that, if all members of X vote insincerely, the result is strongly preferred by all members of X to the sincere outcome. Such a set X is called a **manipulating coalition**.*

*A voting situation is **safe** for voters of type t if there is some type t' such that for all x with $0 \leq x \leq n_t$, whenever x agents of type t change their vote to t', these agents weakly prefer the resulting outcome to the sincere outcome.*

*A voting situation is **safely manipulable** by voters of type t if it is safe for them, and there is some value of x for which the agents concerned strongly prefer the resulting outcome to the sincere outcome.*

There are three main points in the definition of safe manipulation:

- the manipulating coalition consists only of voters of a single type;
- the manipulating strategy is the same for all coalition members;
- the size of this coalition is unknown and there must be no risk of obtaining a worse outcome than the sincere one (through "undershooting" or "overshooting").

Overshooting occurs when the following situation holds. If some number x change from t to t', the result is strongly preferred to the sincere one, but if some number $y > x$ change, the sincere result is strongly preferred to the latter outcome. **Undershooting** is the same, but with $y > x$ replaced by $y < x$. Examples in [8] show that both phenomena can occur. In fact they can both occur in the same voting situation as shown by the following example.

Example 1. *Let $m = 5$ and consider the voting situation with 3 voters having each of the possible preference orders, except the order 12345 which has 4 voters. The scoring rule (see Section 3 for definitions if necessary) with weights $(55, 39, 33, 21, 0)$ yields scores that induce an overall ordering 12345 (meaning candidate 1 wins, candidate 2 is second, etc). Consider voters of type 53124 and the strategy of voting 35241. If 1 voter switches to this strategy, the new winner is candidate 2; if 2 voters switch, then the new winner is candidiate 3; if 3 voters switch, the new winner is candidate 4. This shows that undershooting and overshooting can be possible for the same type and choice of insincere strategy in the same voting situation.*

Remark 1. *We can consider a game in which the set T of types of voters is partitioned into two subsets, T', T''. The set T'' consists of all types of voters whose only action is to vote sincerely, while voters corresponding to types in T' have all possible votes open to them (we do not allow abstention). In the case where $T'' = \emptyset$ and this is common knowledge, we have a fully strategic game. A situation is manipulable if and only if it is not a strong Nash equilibrium of this game.*

When $T' = T_i$ for some fixed type T_i, there is a different game that is easier to analyse. A situation is safe for members of T' if and only if there exists a pure strategy that weakly dominates the sincere strategy, and safely manipulable if and only if there exists a pure strategy that dominates the sincere strategy.

Remark 2. *Note that for each type of voter that ranks the sincere winner lowest, every situation is safe (in fact a stronger statement is true: such voters have nothing to lose by strategic voting, no matter what T' and T'' are). On the other hand, types that rank the sincere winner highest can never manipulate.*

3 Algorithms and polytopes

We restrict to scoring rules. However the method described works more generally (for some rules, much more care may be needed when considering ties).

Scoring rules

Definition 2. *Let $w = (w_1, \ldots, w_m)$ be such that all $w_i \geq 0$, $w_1 \geq w_2 \ldots w_m$ and $w_1 > w_m$. The **scoring rule defined by** w gives the following score to each candidate c:*

$$|c| = \sum_{t \in \mathcal{T}} n_t w_{r(c,t)}$$

*where $r(c,t)$ denotes the rank of c according to type t. The candidates with largest score are the winners. The scores give a **social ordering** of candidates (the value of the associated social welfare function).*

Remark 3. *If a tie occurs for largest score, then a separate tiebreaking procedure is needed in order to obtain a social choice function. This can be a difficult issue, but fortunately when considering asymptotic results under IAC as in this paper, we do not need to consider it further. This is because the set of tied situations has measure zero in the limit as $n \to \infty$.*

We now impose an order on the candidates, and write $\mathcal{C} = \{c_1, c_2, \ldots, c_m\}$. The types are then identified with permutations of $\{1, \ldots, m\}$ and can be written in the usual way. We describe the scores by the **scoreboard**, the tuple $s = (|c_1|, \ldots, |c_m|)$ of scores. The group of types acts on the scoreboard w via permuting candidates and we denote the action of t on w by w^t. In terms of our current notation, we have

$$s = \sum_{t \in \mathcal{T}} n_t w^{t^{-1}}.$$

Example 2. *Let $m = 3$ and consider the voting situation in which 6 agents have preference order 312 and 2 agents have order 213. Under the plurality rule given by $w = (1, 0, 0)$, the scoreboard is $(0, 2, 6)$ and c_3 wins, the social ordering being 321. Under the Borda rule given by $(2, 1, 0)$, the scoreboard is $(8, 4, 12)$ and the order of second and third place is reversed, the social ordering being 312. Under the antiplurality rule given by $w = (1, 1, 0)$, the scoreboard is $(8, 2, 6)$ and social ordering is 132. There is no weight vector for which c_2 can win, as c_3 always has a higher score.*

Without loss of generality we assume from now on that the sincere social ordering is $123 \ldots m$.

3.1 When t and t' are specified

Fix types t and t' until further notice. We now describe the set S of safely manipulable voting situations. S is the union $\bigcup_{t \in T} S_t$, where S_t is the set of situations that are safely manipulable by voters of type t. This can be further refined to $S = \bigcup_{t \neq t'} S_{t,t'}$ where $S_{t,t'}$ is the set of situations that are safely manipulable by voters of type t using strategy t'.

To describe $S_{t,t'}$, we use the following basic observations.

Let x denote the number of members in a coalition of type t who vote insincerely and suppose they vote t'. Then the new and old scoreboards are related by

$$s' - s = x\left(w^{(t')^{-1}} - w^{t^{-1}}\right).$$

For brevity we refer to those candidates ranked above candidate 1 by agents of type t as **good**, and those ranked below 1 as **bad**. For example, when $m = 3$ and the social ordering is 123, then according to an agent of type 213, c_2 is good and c_3 is bad. The new outcome is preferred by type t agents if and only if no bad candidate is the new winner. It is strongly preferred if and only if some good candidate is the new winner.

Proposition 1. *When $m = 3$, undershooting can never occur, and overshooting occurs if and only if some bad candidate wins when $x = n_t$.*

Proof. First note that as a function of x, the differences in scores of each alternative between the sincere and strategic voting situation are (linearly) either increasing or decreasing. Thus if candidate i is above candidate j for some x but below for some larger value of x, it will remain below for all even larger values of x. For types 123 and 132, no better result can be achieved by strategic voting; for types 231 and 321, no worse result. The only other cases are types 213 and 312. In each case there is only one good and one bad candidate: once one overtakes the other and the sincere winner, it stays ahead and cannot be subsequently beaten by another candidate of the opposite type. □

Proposition 2. *The following algorithm solves the decision problem for safe manipulation for scoring rules, and runs in polynomial time provided the tiebreaking procedure does.*

Let $|c|_x$ denote the score of candidate c when x agents have switched from t to t', and let L be the set of points of intersection of the graphs of the functions $x \mapsto |c|_x$ for $0 \leq x \leq n_t$. Sort the elements of L. For each interval formed by successive elements, compute the maximum score B of all bad candidates, and the maximum score G of all good candidates. If $B > G$ for any interval (or $B = G$ and the tiebreaking procedure says that a valid manipulation in favour of a bad candidate has occurred) then safe manipulation is not possible; otherwise it is possible.

Proof. The winner is constant on each interval, so we need only check one point in each interval, plus endpoints to deal with ties. There are at most $m(m-1)/2$ intersections of the lines which are the graphs of the functions $x \mapsto |c|_x$ for $0 \leq x \leq n_t$. The condition on maximum good and bad scores can be checked for each interval in time proportional to m. □

Corollary 1. *When $m = 3$, we need only calculate which candidate wins when $x = n_t$, and safe manipulation is possible if and only if the winner is good.*

3.2 The general case

When at least one of t and t' is not specified, there are obviously more possibilities, and a brute force approach that simply tries each pair (t, t') in turn will work. However, we can clearly do better than this.

There are some values of t for which S_t is empty. This means that no matter what the situation and the differences in the sincere scores, safe manipulation is impossible by type t. For example, every t for which the sincere winner 1 is ranked first has no incentive to manipulate. Other types have incentive but as we see in Example 3, S_t may still be empty.

For those t for which S_t is nonempty, we can still remove strategies t' for which $S_{t,t'}$ is empty. Similarly, we can express the union defining S_t with as few terms as possible. This is done by discarding dominated strategies (in any particular voting situation, even more strategies may be dominated, but we consider here those that are never worth including for any situation). For example, any type that ranks a bad candidate ahead of a good one is dominated by the type that differs only by transposing those two candidates. Thus all good candidates should be ranked ahead of all bad ones. The sincere winner should not be ranked ahead of any good candidate for the same reason. Furthermore, each strategy that does not allow some good candidate to catch the sincere winner should be rejected, as should each strategy that further advantages a bad candidate higher in the social ordering over all good candidates.

Example 3 ($m = 3$)**.** *Consider type* 312*. The only possibly undominated strategy that we need to consider, according to the above discussion, is* 321*. However* 321 *cannot lead to successful manipulation, as it increases the score of* 2 *and not of* 3*. Thus type* 312 *cannot manipulate at all, let alone safely. Types* $231, 213$ *and* 321 *have respectively the strategies* $321, 231, 231$ *available.*

Example 4. *When* $m = 4$*, the strategies that are worth considering in some situation are as follows. For any type starting with* 1*, only the sincere strategy. For any type ending with* 1*, any strategy that keeps* 1 *at the bottom. For types starting* 41*, only the sincere strategy; for types starting* 31*, any strategy that lowers* 1 *while keeping* 3 *at the top and not promoting* 2*; for types starting* 21*, any strategy that lowers* 1*, keeping* 2 *first. For types ranking* 1 *third, transpose the two good candidates.*

When there are very few distinct entries in w, there are many fewer strategies to consider. The extreme cases are plurality ($w = (1, 0, \ldots, 0)$) and antiplurality $(1, 1, \ldots, 1, 0)$). For plurality (respectively antiplurality), safe manipulation is possible by a type t voter if and only if it is possible by ranking some good candidate first (respectively some bad candidate last). The player is indifferent between the different strategies satisfying this criterion (if the good candidate is fixed) and the analysis does not distinguish between them, so we can assume that any such voter uses a standard strategy that makes a chosen good candidate the favoured one and orders the others by increasing value of index. Thus, for example, for $m = 3$ under plurality we consider 213 and 312 as possible values for t'.

We have so far expressed S_t in terms of a union of $S_{t,t'}$ which is as small as possible. However the terms in the union may not be disjoint. For example, with $m \geq 4$ a voter of type ranking c_1 last may use any of the $(m-1)! - 1$ insincere strategies that leave c_1 at the bottom (when $m = 3$ there is only one such strategy).

To compute the final probability of safe manipulation, we need to compute the volume of the union of all S_t. This union is in general not disjoint even for $m = 3$, as the following example shows.

Example 5. *Let* $m = 3$ *and consider the voting situation with* 3 *agents having preference* 123*,* 2 *having preference* 231 *and* 2 *having preference* 321*. Under the plurality rule, the last two types can each manipulate safely.*

We use inclusion-exclusion to compute the volume of the union. The number of terms in the inclusion-exclusion formula is $2^p - 1$ where p is the number of types involved.

4 Numerical results

We restrict to $m = 3$ and some selected scoring rules including the commonly studied plurality, Borda ($w = (2, 1, 0)$), and antiplurality.

For a situation in which the sincere result is 123, types $123, 132$ and 312 cannot manipulate safely. We need deal with only the remaining types, each of which has only one insincere strategy to consider. The linear systems in question are as follows. We denote $w_i - w_j$ by w_{ij}.

The fact that 123 is the sincere result is expressed as $|c_1| \geq |c_2| \geq |c_3|$. This translates to

$$\begin{aligned} 0 &\leq n_1 w_{12} + n_2 w_{13} + n_3 w_{21} + n_4 w_{31} + n_5 w_{23} + n_6 w_{32} \\ 0 &\leq n_1 w_{23} + n_2 w_{32} + n_3 w_{13} + n_4 w_{12} + n_5 w_{31} + n_6 w_{21} \\ n_i &\geq 0 \text{ for all } i \\ n &= n_1 + \cdots + n_6. \end{aligned}$$

For type 213, safe manipulation is possible if and only the following additional conditions are satisfied.

$$\begin{aligned} |c_2| &\geq |c_1| - n_3 w_{23} \\ |c_2| &\geq |c_3| + n_3 w_{23} \end{aligned}$$

which simplifies to the following system.

$$\begin{aligned} 0 &\geq n_1 w_{12} + n_2 w_{13} + n_3 w_{31} + n_4 w_{31} + n_5 w_{23} + n_6 w_{32} \\ 0 &\leq n_1 w_{23} + n_2 w_{32} + n_3 w_{12} + n_4 w_{12} + n_5 w_{31} + n_6 w_{21} \end{aligned}$$

Every voting situation can be represented in this way up to a permutation of alternatives.

Thus the asymptotic probability under IAC that type 213 can safely manipulate is given by the ratio of the volume of the "strategic" polytope to that of the "sincere" polytope. A completely analogous method works for other types. The volumes can be computed using standard software as described in [9, 3].

The results for several voting rules are shown in Table 1. The column labelled "P(manip)" gives the asymptotics probability of a voting situation begin manipulable (possibly by a coalition of more than one type) and was computed using the methods in [4] (note that the results for plurality, antiplurality and Borda have been computed exactly elsewhere [9]). Note that the ordering of rules according to their susceptibility to manipulation and the corresponding order for safe manipulation differ. Also the entries in the last column, giving conditional probabilities, are decreasing. This last fact is not surprising in hindsight and probably not dependent on the culture IAC. For example, plurality allows only one type of member in a minimal manipulating coalition, and such members have nothing to lose, so manipulation is possible if and only if it is safely possible. At the other extreme, only one type of voter can manipulate under antiplurality, but whether this is safe or not depends strongly on the voting situation.

The Borda rule is often criticized for its susceptibility to manipulation. While it is still the most manipulable here by both measures, it is clear that many manipulable situations under Borda require unsafe manipulations. The plurality rule seems the least manipulable when complicated coalitions are used, but its advantage disappears when safety is considered. These results, which of course depend on the particular distribution IAC, nevertheless indicate that when communication is restricted, traditional ratings of voting rules may need to be revised.

Table 2 shows the probability that a given rule is safely manipulable by all of the individual types listed. We see for example that type 213 has the most manipulating power under the $(3, 2, 0)$ rule, whereas 231 and 321 are strongest under plurality. Note that, for

Table 1: Asymptotic probability under IAC of a situation being (safely) manipulable.

scoring rule	P(manip)	P(safely)	P (safely | manip)
Plurality	0.292	0.292	1.00
(3,1,0)	0.422	0.322	0.76
Borda	0.502	0.347	0.69
(3,2,0)	0.535	0.330	0.62
(10,9,0)	0.533	0.264	0.49
Antiplurality	0.525	0.222	0.42

Table 2: Asymptotic probability under IAC of safe manipulation by various types

scoring rule	213	231	321	213, 231	213, 321	231, 321	213, 231, 321
Plurality	0.0000000	0.156250	0.246528	0.000000	0.000000	0.111111	0.000000
(3,1,0)	0.178369	0.086670	0.216913	0.000080	0.104229	0.053084	0.000067
Borda	0.225000	0.047950	0.196759	0.000033	0.093542	0.027400	0.000024
(3,2,0)	0.239297	0.020019	0.152812	0.000007	0.070438	0.010926	0.000005
(10,9,0)	0.234375	0.001687	0.051107	0.000000	0.022681	0.000866	0.000000
Antiplurality	0.2222222	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

example, there is an appreciable probability that both types 213 and 321 can manipulate safely. If each proceeds, ignoring the other, the result may no longer be safe. On the other hand, if both 231 and 321 try simultaneously to manipulate safely, the cancellation effect means that they are less likely to be disappointed.

5 Further discussion

The uniform distribution on profiles (the Impartial Culture hypothesis) has been used in many analyses in voting theory, because of its analytical tractability. However, for the asymptotic study of safe manipulation it seems less useful. This is because under IC for scoring rules, much weight is placed on situations that are nearly tied: a typical situation has almost equal numbers of each type, and the differences between the scores are of order $\sqrt{n}$. Thus as $n \to \infty$ the probability that, for example, a voter of type 321 can safely manipulate will approach 1 rapidly, while the probability that a type 213 can do so will approach 0 rapidly.

The inclusion-exclusion procedure used is probably exponential in m, since the number p of types used seems to grow linearly in m (we have not formally proved this). Thus a better algorithm is needed for large m.

As pointed out by the referee, the argument of Section 3.2 involve a monotonicity property that should be satisfied by more than just the scoring rules, but we have not pursued such a generalization here, leaving it for possible future work.

The literature on safe manipulation is very small still - our literature search turned up only one preprint of unknown publication status, dealing with complexity issues (though a similar idea was apparently used in [1] without explicit mention). However the basic model is attractive and some obvious generalizations should be investigated. For example, we can use a probability distribution to model the number of followers, instead of considering

the worst case outcome, and thereby consider whether strategic voting even with lack of coordination can lead to better outcomes in the sense of expected utility.

We thank the referee for several useful comments.

References

[1] Pierre Favardin and Dominique Lepelley. Some further results on the manipulability of social choice rules. *Soc. Choice Welf.*, 26(3):485–509, 2006.

[2] Allan Gibbard. Manipulation of voting schemes: a general result. *Econometrica*, 41:587–601, 1973.

[3] Dominique Lepelley, Ahmed Louichi, and Hatem Smaoui. On Ehrhart polynomials and probability calculations in voting theory. *Soc. Choice Welf.*, 30(3):363–383, April 2008.

[4] Geoffrey Pritchard and Mark C. Wilson. Exact results on manipulability of positional voting rules. *Soc. Choice Welf.*, 29(3):487–513, 2007.

[5] Mark Allen Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *J. Econom. Theory*, 10(2):187–217, 1975.

[6] Arkadii Slinko. How large should a coalition be to manipulate an election? *Math. Social Sci.*, 47(3):289–293, 2004.

[7] Arkadii Slinko. How the size of a coalition affects its chances to influence an election. *Soc. Choice Welf.*, 26(1):143–153, 2006.

[8] Arkadii Slinko and Shaun White. Is it ever safe to vote strategically? Technical Report 563, University of Auckland Department of Mathematics Report Series, 2008.

[9] Mark C. Wilson and Geoffrey Pritchard. Probability calculations under the IAC hypothesis. *Math. Social Sci.*, 54(3):244–256, 2007.

Mark C. Wilson
Department of Computer Science
University of Auckland
Email: `mcw@cs.auckland.ac.nz`

Reyhaneh Reyhani
Department of Computer Science
University of Auckland
Email: `rrey015@aucklanduni.ac.nz`