Exploring pathological scenarios of preferential voting elections

Ken Gene Quah, supervised by Damjan Vukcevic

Introduction

A recent study introduced hypothetical elections where one candidate slowly accumulates a majority of the votes as their opponents get eliminated [1]. Such a scenario can occur under the preferential voting system. This type of election is challenging for statistical inference and hence difficult to verify the election result.

In this project, we explore these ‘pathological’ scenarios analytically and through simulations. The simulations evaluate the winner of the hypothetical elections after a small perturbation of the ballots has been performed.

Why pathological?

We look at elections where almost every voter is of one of three types:

1. $[A]$ (Only puts one preference: Candidate A)
2. $[B]$ (Only puts one preference: Candidate B)
3. $[C_i, B]$ (Puts two preferences: First preference is some other candidate not A or B, then second preference is Candidate B)

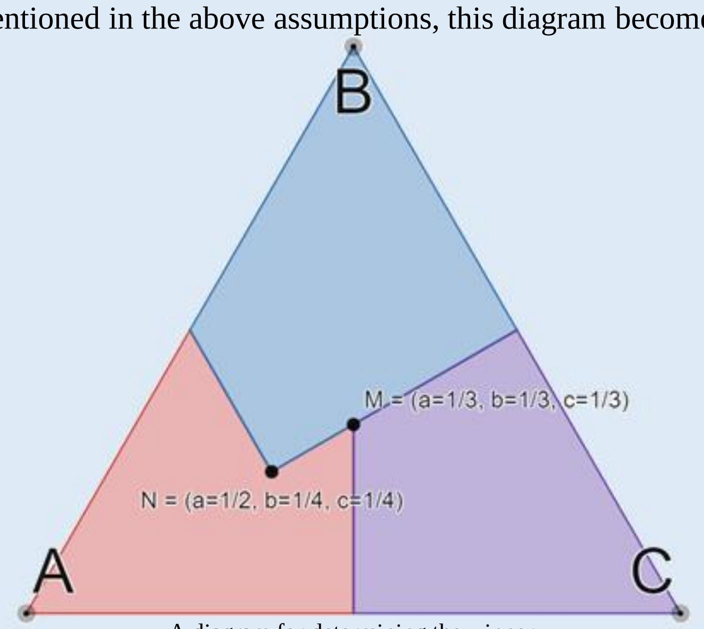
In addition, assume that the numbers of $[B]$, $[C_1, B]$, $[C_2, B]$,...etc. are approximately equal.

If B is eliminated before any C is eliminated, then the election result will be determined by the ballots’ first preferences. However, if any C is eliminated while B remains, then B receives that eliminated C’s votes, and now B is no longer at risk of being eliminated (until the final two). B eventually accumulates all the votes from all C’s as they get eliminated.

Visualisation

(for 3 candidates)

In [2], Eggers introduces a diagram for describing the winner of a 3-candidate election as a function of a, b, c. These are the proportions of first preference ballots candidates A, B, C receives respectively. With the situation mentioned in the above assumptions, this diagram becomes:



A diagram for determining the winner

To use this diagram, take a weighted average of the points A, B, C with weights a, b, c and mark this point on the diagram. The region it falls into tells us which candidate wins (or which candidates are tied).

Above are marked two interesting points M and N. In the neighbourhoods of M and N, the areas are split 1:3:2 and 3:1 respectively.

Model

Suppose there k ways to fill out a ballot. Then we can model the ballots received as a k-dimensional vector $\vec{p} = (p_1, \dots, p_k)$, with each component non-negative and $\sum_{i=1}^{k} p_i = 1$. Each p_i is the proportion of ballots of type i received. $\vec{p}$ determines which candidate wins. We also limit voters to put a maximum of 3 preferences to significantly reduce the number of dimensions of $\vec{p}$ (the analysis in [1] does this too).

Method

First we choose a value of n (number of candidates) and an initial proportion (Scenario M or N). The n candidates are $A, B, C_1, C_2, \dots, C_{n-2}$.

Next we perform a large number of random trials (100000+). In each trial,

1. Select radius $r \in \mathrm{Unif}(0, 0.1)$
2. Uniformly sample $\vec{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_k)$ on the set $\{\sum_{i=1}^{k} \varepsilon_i = 0, ||\vec{\varepsilon}|| = r$, each component of $\vec{p} + \vec{\varepsilon}$ is non-negative$\}$
3. Determine the election result for the vector $\vec{p} + \vec{\varepsilon}$.

Scenario M:

$[A] = \frac{1}{n}$

$[B] = \frac{1}{n}$

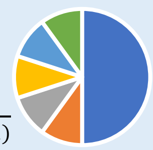
$[C_i, B] = \frac{1}{n}$ for each $1 \le i \le n-2$

Scenario N:

$[A] = \frac{1}{2}$

$[B] = \frac{1}{2(n-1)}$

$[C_i, B] = \frac{1}{2(n-1)}$ for each $1 \le i \le n-2$



$||\vec{\varepsilon}||$ represents the size of the perturbation. We will try both:

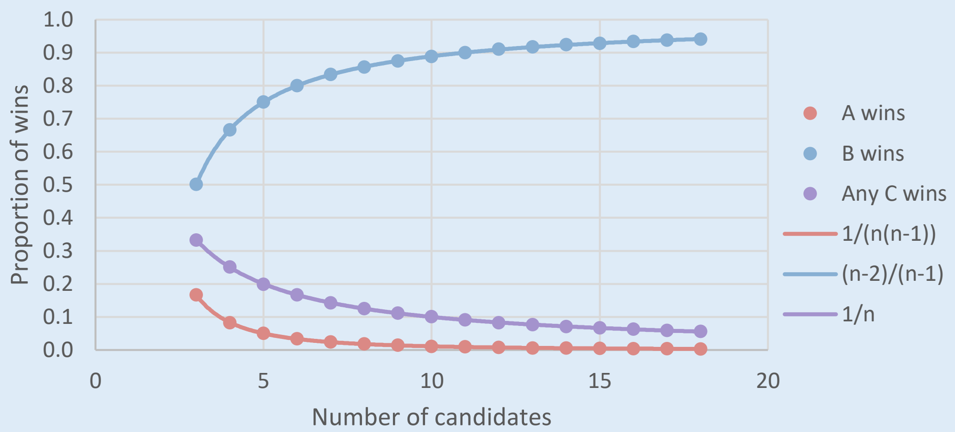
1-norm (Manhattan distance): $||\vec{\varepsilon}||_1 = \sum_{i=1}^{k} |\varepsilon_i|$ and

2-norm (Euclidean distance): $||\vec{\varepsilon}||_2 = \sqrt{\sum_{i=1}^{k} \varepsilon_i^2}$

Results

Only the results using Manhattan distance are shown, since using Euclidean distance yielded very similar results. In both scenarios we test all n from 3 to 18.

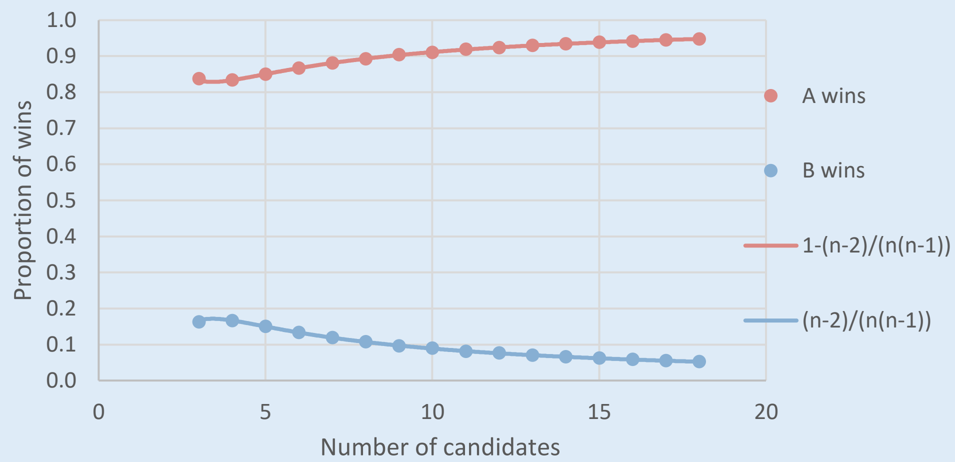
Scenario M



In scenario M, it appears that the win rates are as follows:

	Win rate
A	$\frac{1}{n(n-1)}$
B	$\frac{n-2}{n-1}$
Any C	$\frac{1}{n}$

Scenario N

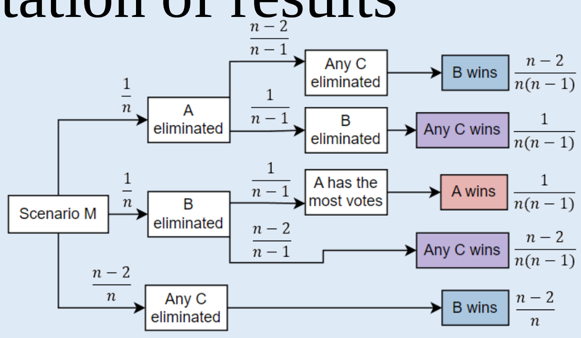


In scenario N, it appears that the win rates are as follows:

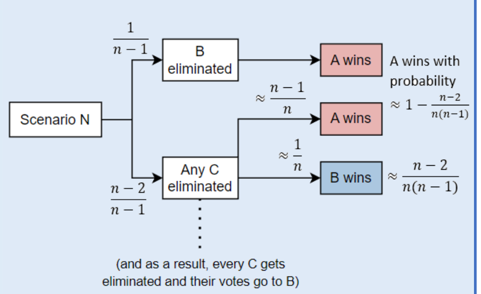
	Win rate
A	$1 - \frac{n-2}{n(n-1)}$
B	$\frac{n-2}{n(n-1)}$

Interpretation of results

On the right are probability tree diagrams, explaining the above win rates and trends.



B wins with probability $\frac{n-2}{n-1}$ and any C wins with probability $\frac{1}{n}$



References

[1] Everest, F., Blom, M., Stark, P.B., Stuckey, P.J., Teague, V., Vukcevic, D. Ballot-Polling Audits of Instant-Runoff Voting Elections with a Dirichlet-Tree Model (2022). https://doi.org/10.48550/arXiv.2209.03881

[2] Eggers, A.C. A diagram for analyzing ordinal voting systems. *Soc Choice Welf* 56, 143–171 (2021). https://doi.org/10.1007/s00355-020-01274-y