Chapter 9
Modeling the Outcomes of Vote-Casting in Actual Elections

T. Nicolaus Tideman and Florenz Plassmann

9.1 Introduction

How often do events of interest to voting theorists occur in actual elections? For example, what is the probability of observing a voting cycle – an outcome in which no candidate beats all other candidates in pairwise comparison by majority rule? When there is a candidate who beats all others in such pairwise comparisons – a Condorcet winner – what is the probability that a voting method chooses this candidate? What is the probability that voters have an incentive to vote strategically – that is, cast their votes in ways that do not reflect their true preferences? Voting theorists have analyzed these questions in great detail, using a variety of statistical models that describe different distributions of candidate rankings. But there has been no systematic effort to determine which statistical model comes closest to describing the distribution of rankings of candidates in actual elections. Thus we know how often various voting events occur under different statistical models, but not how often voting events occur in actual elections. This chapter provides a framework for answering this question.

We consider elections in which each voter is asked to submit a strict ranking of \( m \) candidates. We interpret the rankings submitted by all voters as the outcome of a statistical model of vote-casting that yields a vector with \( m! \) components, representing the possible strict rankings of the \( m \) candidates. To assess the probabilities of voting events, we need to know the likelihoods of these vectors – are all vectors of rankings equally likely or are some more likely to occur than others? We identify a statistical model of vote-casting that comes very close to describing the distribution of vectors of rankings in actual three-candidate elections.
We find it intuitive to view the statistical model of vote-casting as a two-part process. One part describes the distribution of the expected shares of the voters who will report each of the $m!$ rankings. The other part describes the distribution of the observed shares of the voters who report each of the $m!$ rankings, given a vector that describes the expected shares of these rankings.

Voting theorists have proposed various models to describe possible distributions of the expected shares. However, these models were generally proposed for purposes other than describing rankings in actual elections, and we are not aware of a systematic investigation of whether any of these models is at all likely to describe rankings in actual vote-casting processes.

We consider, for the case of three candidates, nine models that have been proposed by others as well as three new models. We evaluate these 12 models with two sets of voting data. Our first data set was assembled by Nicolaus Tideman in 1987 and 1988, and it consists of individual ballot information for 87 elections that we transform into 883 three-candidate elections.\footnote{These data have been analyzed previously by Feld and Grofman (1990 and 1992), Felsenthal et al. (1993), Felsenthal and Machover (1995), Tideman and Richardson (2000), Regenwetter et al. (2002), and Tideman (2006).} Our second data set consists of 913 three-candidate “elections” that we construct from the “thermometer scores” that are part of the surveys conducted by the American National Election Studies (ANES).\footnote{ANES survey data have been used in several previous analyses of voting. For example, Chamberlin and Featherston (1986) use scores from ANES surveys administered in 1972, 1974, 1976, and 1978 to construct combinations of four candidates. Regenwetter et al. (2002 and 2003) analyze the thermometer scores of the three major candidates in the four ANES surveys administered in 1968, 1980, 1992, and 1996 to construct combinations of three candidates. Our method of constructing three-candidate “elections” is the same as that in these earlier analyses.} The results that we obtain from these two rather different data sets are very consistent; they indicate that the combination of a spatial model of voting to describe the distribution of the vector of expected shares and a multinomial model to describe the distribution of the vector of actual shares fits the observed results of three-candidate elections much better than any of the models that have so far been used in theoretical analyses of voting events, and well enough that it may be difficult to devise an alternative model that would fit actual election data significantly better.

While the spatial model fits the observed data very well, it is too complex to permit the type of calculation of the probabilities of voting events that theorists have undertaken so far (as, for example, in Gehrlein 2002). Thus we envisage using this model instead for Monte Carlo simulations, to generate data that have the same characteristics as data from actual elections. Such simulations would be unnecessary if there were enough data from actual elections in which voters rank the candidates to determine the frequencies of rare voting events with a reasonable degree of accuracy. But there are not nearly enough ranking data to undertake such a project. For example, Brams and Fishburn (2001) and Saari (2001) use data from a single election – the 1999 election for president of the Social Choice and Welfare Society – to illustrate and analyze the properties of different voting methods.