Risk-limiting Audits and the Margin of Victory in Nonplurality Elections

Anand Sarwate
TTI Chicago

Stephen Checkoway
Johns Hopkins University

Hovav Shacham
UC San Diego

Abstract

Post-election audits are an important method for verifying the outcome of an election. Recent work on risk-limiting, post-election audits has focused almost exclusively on plurality elections. Several organization and municipalities use nonplurality methods such as range voting, the Borda count, and instant-runoff voting (IRV). We believe that it is crucial to develop effective methods of performing risk-limiting, post-election audits for these methods. We define a general notion of the margin of victory and develop risk-limiting auditing procedures for these nonplurality methods. For positional or scored systems, we show how to adapt methods from plurality auditing. For IRV, the situation is markedly different. We provide a risk-limiting method for auditing the candidate elimination order. We provide a more efficient audit for the elections in which the margin of the IRV election can be efficiently calculated or bounded. We provide efficiently computable upper and lower bounds on the margin and, where possible, compare them to the exact margins for a large number of real elections.

1 Introduction

In political elections, plurality voting (also known as first-past-the-post) is the most widely-used system for determining the winner. The drawbacks of plurality voting, such as the "spoiler effect," are well-documented (Black, 1958, Poundstone, 2008, Saari, 2001, Szpiro, 2010), and recently, several groups and municipalities have adopted alternative voting systems. The most common alternative systems are instant-runoff voting (IRV) or, its multi-winner version, the single transferable vote (STV), but other systems such as approval voting (Brams and Fishburn, 2005), range voting, and Borda counts are used in both political and nonpolitical elections. A major component of election certification is a *post-election audit* which is a procedure that samples ballots and compares the electronic vote tallies with paper ballots in order to validate the reported outcome. These tallies may be formed by scanning the paper ballots in an optical-scan machine, or from voting machines that produce paper ballot "receipts" as part of a voter-verified paper audit trail (VVPAT). A *risk-limiting audit* is one for which there is a known probability (over the sampling), or risk level, of certifying an outcome that is incorrect. Although many audit mechanisms have been proposed for plurality voting, to our knowledge few audit mechanisms have been proposed for alternative voting systems. We contend that auditing is integral to properly certifying elections; our contribution in this paper is to describe risk-limiting audit mechanisms for a range of nonplurality voting systems.

Many political systems allocate resources to parties based on their popularity in elections; they can receive funding and recognition if they capture the first preferences of a certain portion of the electorate (Roberts, 2010). Many activists, especially in the United States, feel that plurality voting entrenches two-party systems. Proponents of various alternatives to plurality voting have successfully changed the voting systems used by municipalities and professional societies, and there are some recent empirical studies of how these systems work in practice (Brams and Fishburn, 2005, Farvaque, Jayet, and Ragot, 2009).

We divide alternatives to plurality voting into two classes. In *scored systems* voters assign points to each candidate; examples include range voting, approval voting, and Borda counts; these include positional systems (Saari, 1995). In *ranked systems*, voters rank some or all of the candidates; IRV and Condorcet methods (Fishburn, 1977, Schulze, 1997, Tideman, 1987, Woodall, 1997) are ranked systems. Approval and range voting are used by some professional societies but not yet in major political elections (Brams and Fishburn, 2005), whereas Borda counts are used for some political elections in Slovenia (Consortium for Elections and Political Process Strengthening, 2011) and the Pacific island nations of Nauru and Kiribati (Reilly, 2002), as well as in sports — e.g., the Heisman Trophy (The Heisman Memorial Trophy, 2011) — and academic professional societies (Brams

and Fishburn, 2005). A Condorcet method proposed by Schulze (2011) is used by the Swedish Pirate Party (for primaries), the Wikimedia Foundation, the Debian project, and the Gentoo project.[1]

By far the most popular alternatives to plurality voting are STV or IRV (sometimes known also as Ranked Choice Voting (RCV)). In an IRV election, candidates are eliminated sequentially, beginning with the candidate receiving the fewest first-ranked votes.[2] The ballots whose first-ranked candidate was eliminated are assigned to their second-ranked candidates. A more formal description of IRV is given in Section 5.3. The Australian House of Representatives uses STV (Australian Electoral Commission, 2011), as does the Republic of Ireland for all public elections including presidential elections and elections to Dáil Éireann — the lower house of parliament (Consortium for Elections and Political Process Strengthening, 2011). The California cities of Berkeley, Oakland, San Francisco, and San Leandro use IRV for some elections.

California law requires a 1% manual tally of each election that is then reported to the Secretary of State. Officials may then compare the paper ballots to machine records to determine if there are anomalies. Municipal elections in San Francisco use IRV. For these elections, the following manual tally procedure is employed (San Francisco Voting Systems Task Force, 2011). First, in each randomly chosen precinct, the paper ballots are examined to determine the number of first-choice, second-choice, and third-choice votes each candidate received;[3] these totals are compared against the corresponding totals claimed in the original machine count. Second, an IRV elimination election is run with only the ballots from the tallied precinct, and the winner of this mini-election is noted.

There is no reason to believe that the San Francisco tally of IRV elections is a risk-limiting audit for any particular risk level. Indeed, the San Francisco Voting Systems Task Force gives an example election in which two sets of ballots that are identical under the tally procedure produce two different election outcomes (San Francisco Voting Systems Task Force, 2011, Appendix A).[4] In this example election, running San Francisco's manual tally and finding no discrepancies does not increase our confidence that the reported and actual winner are the same! By contrast, a 1% manual tally of a plurality election *can* provide a risk-measuring audit, though the risk level depends on the election margin.

[1] Wikipedia lists 60 organizations which use the Schulze Method in some form. `http://en.wikipedia.org/w/index.php?title=Schulze_method&oldid=434396935#Use_of_the_Schulze_method` Accessed 2011-06-15.

[2] This is Hare's rule for ballot transfers (Tideman, 1995).

[3] San Francisco allows voters to rank no more than three of the candidates for each race.

[4] In a presentation at the EVN 2011 conference, Emily Shen gave another such example.

Our contributions

In this paper, we take steps towards developing audit mechanisms for nonplurality voting systems. We propose a generalized definition of the *margin of victory* for an election. Our new margin is the minimum level of *ballot errors* — unintentional or otherwise — that must have been introduced to change the final tallies from a situation with a different outcome (i.e., a winner different from the reported winner) to the reported tallies. The key to this definition is that ballot errors are measured differently under different voting rules. Margins are discussed in more detail in Sections 4.1 and 5.1.

We show how to audit scored systems by adapting risk-limiting audits for plurality elections. For Condorcet methods, we can reduce the auditing problem to one with multiple-contest plurality elections; these can be audited by adapting the method of Stark (2010). We propose two approaches for risk-limiting audits for IRV based on the margin of victory or a bound on the margin. Although it is possible to compute the margin exactly (Section 6), these methods may be computationally costly, especially when the number of candidates is large. As an alternative, we provide low-complexity upper and lower bounds on the margin which can be used to evaluate the difficulty of auditing a particular election. Furthermore, we analyze real election data from IRV elections to compare our bounds with the exact margin. Intriguingly, we find that in these real elections, the IRV winner is almost always a Condorcet winner.

2 Related work

Auditing non-plurality systems is connected to both the statistical literature on post-election auditing and the mathematical literature on social choice theory.

2.1 Risk-limiting audits

An audit consists of sampling ballots, comparing the paper ballots with the cast vote records (CVRs), and deciding whether to continue sampling, stop and certify the reported winner of the election, or demand a full hand count. A single-ballot audit samples individual ballots from all those cast in the election. For ballots that are sampled, we assume that the auditor can determine both the intent of the voter and how that vote was counted. Based on the randomness used to sample the ballots, if the auditor can find a number α such that based on the evidence, certifying the election will be incorrect with probability at most α, then the procedure is called a *risk-limiting audit at risk level* α. Statistical, post-election audits were first proposed

by Saltman (1975), and a recent survey is available in Lindeman and Stark (2012).[5] In recent years, the problem of providing strong guarantees for the election outcome's correctness has been studied along two orthogonal axes.[6]

The first axis concerns exactly what the audit seeks to (statistically) guarantee. Earlier work focused on finding evidence of a single miscounted vote (see Dopp (2008) and the references therein for the history of these methods). These audits do not certify the outcome unless no errors are found. Unfortunately, almost every election has some miscounted ballots, due to human or machine error, tampering with voting machines, or tampering with election software. In contrast, Stark (2008a) proposed the first complete audit procedure that specifies what to do when miscounts are discovered. Stark's procedure looks for evidence that the reported *outcome* is incorrect rather than looking for incorrect tallies. Follow up work produced procedures that are easier to understand and, simultaneously, statistically more powerful (Stark, 2008b,c, 2009a,b,d, 2010). Checkoway et al. (2010) proposed an auditing method based on convex optimization with the same basic goal of finding evidence of incorrect outcomes.

The second axis of study concerns the size of each sample to be audited. Most early auditing procedures operate at the granularity of a precinct as that is the granularity at which most results are tabulated (Aslam, Popa, and Rivest, 2007, 2008, Higgins, Rivest, and Stark, 2011, Stanislevic, 2006, Stark, 2008a,b,c, 2009a,b,c,d). The traditional organization of elections into precincts makes this a natural model; however, Calandrino, Halderman, and Felten (2007), Johnson (2004), Neff (2003); and Sturton, Rescorla, and Wagner (2009) note that the statistical power of post-election audits would be greatly increased by reducing the unit of an audit to a single ballot.[7] One challenge with ballot-level auditing is that the system must be able to associate the CVR with the physical ballot. This can be done by printing a unique serial number on each ballot as they are being counted (Calandrino et al., 2007) or by weighing stacks of ballots (Sturton et al., 2009). Because the gain in statistical power is so great, most recent algorithms use ballot-level auditing (Benaloh, Jones, Lazarus, Lindeman, and Stark, 2011, Checkoway et al., 2010, Stark, 2010).

[5] A simpler form of auditing simply recounts ballots to confirm the winner, called a ballot polling audit in Lindeman and Stark (2012).

[6] The remainder of this subsection is adapted from the authors' earlier work on risk-limiting, post-election audits (Checkoway, Sarwate, and Shacham, 2010).

[7] Intermediate sub-precinct audit units, such as individual voting machines, appear to provide little gain in statistical power, but may reduce the cost of locating the ballots to audit.

2.2 Social choice, strategic voting, and manipulation

The field of social choice theory deals with how to aggregate individual preferences into a societal preference. Starting with the seminal work of Arrow (1951), researchers have investigated what is and what isn't possible in terms of the properties of social choice aggregation functions. Popular accounts can be found in recent books of Poundstone (2008) and Szpiro (2010), and more technical introductions in books by Black (1958) and Saari (2001). Social choice theory is primarily interested in the structure of how individual choices are aggregated, and not in how to measure and correct errors for a given social choice function. While plurality voting is hardly beloved, scholars have demonstrated theoretical flaws in many voting systems, including STV (Brams and Fishburn, 1983, Doron and Kronick, 1977, Dummett, 1984).

Some work has been done in computational social choice on the problem of strategic voting (Gardenfors, 1976, Gibbard, 1973, Satterthwaite, 1975). Strategic voting arises because voters have an incentive to cast ballots that do not reflect their true preferences. However, from the auditor's perspective, the voters' true preferences are irrelevant; a post-election audit is concerned with making sure that the voters' expressed preferences are counted correctly. A question common to both strategic voting and auditing is the following: Given the ballots cast in an election, how large a subset must an adversary control in order to force a particular outcome of the election? From the perspective of strategic voting, this subset is a coalition of strategic voters. From the perspective of auditing, the subset is the minimum number of errors required to change the outcome of the election.

Bartholdi III and Orlin (1991) showed that under STV, it is NP-complete for a manipulator, knowing all other ballots, to find a preference order for themselves to ensure the election of a favored candidate. For scored systems like plurality voting the manipulation problem can be easier for a single manipulator (Bartholdi III, Tovey, and Trick, 1989, 1992), but recent work has shown that it is NP-hard to find multiple manipulators for a Borda count (Betzler, Niedermeier, and Woeginger, 2011, Davies, Katsirelos, Narodytska, and Walsh, 2011). In these works the complexity is measured as a function of the number of candidates. Later work has focused on the computational hardness results when the number of candidates is fixed (Conitzer, Sandholm, and Lang, 2007).

These hardness results show that an instance of an NP-hard problem is equivalent to a particular election manipulation problem, but do not show that a given election is hard to manipulate. This has led to several alternative ways of thinking about the complexity of manipulation, for example by extending the types of manipulation (Faliszewski, Hemaspaandra, and Hemaspaandra, 2011), designing approximation algorithms (Brelsford, Faliszewski, Hemaspaandra, Schnoor, and

Schnoor, 2008), using average-case complexity (Procaccia and Rosenschein, 2007), or random models for errors (Friedgut, Kalai, and Nisan, 2008, Isaksson, Kindler, and Mossel, 2010). Some researchers believe that the hardness of manipulation is a desirable property, especially in elections done automatically by computer agents (Faliszewski, Hemaspaandra, and Hemaspaandra, 2010).

These hardness results do not immediately imply that computing the number of ballots needed to manipulate a given election is itself hard. Firstly, in many real elections the number of candidates is small. Secondly, most real elections will not look like the NP-hard manipulation instances used in computational social choice.

3 Risk-limiting audits and the margin of victory

Let k denote the number of candidates in the election and let $[k] = \{1, 2, \ldots, k\}$ denote the set of candidates. Let n be the total number of ballots cast in the election. We think of each ballot as a pair of values $(\mathbf{x}_i, \mathbf{y}_i)$, where $\mathbf{x}_i$ is the true marking of the ballot by voter i and $\mathbf{y}_i$ is the marking of the ballot as reported by the election tabulation system. In some cases $\mathbf{x}_i \neq \mathbf{y}_i$; our underlying assumption is that some of the ballots may be miscounted either due to human error, machine error, or adversarial tampering.

Let $\Omega(\cdot)$ be the function that calculates the winner of the election and let $w_r = \Omega(\{\mathbf{y}_i : i \in [n]\})$ be the *reported winner* of the election. The function $\Omega(\cdot)$ represents the particular voting system used (e.g., plurality, Borda count, IRV). The *actual winner* of the election is $w_a = \Omega(\{\mathbf{x}_i : i \in [n]\})$, which is the $\Omega(\cdot)$ function applied to the true ballot values. The reported outcome is correct if $w_r = w_a$; otherwise, it is incorrect.

For simplicity, we will consider ballot-level audits; an audit is *ballot-level* if it can sample an individual ballot from the set of all n ballots. Ballot-level audits have much greater statistical power than precinct-based audits. We restrict our attention to audits that draw ballots uniformly at random (with replacement) from the list of ballots used in the election.[8] The auditor samples K numbers $\{i_1, i_2, \ldots, i_K\}$ uniformly from $[n]$ and examines the ballots $A = \{(\mathbf{x}_{i_j}, \mathbf{y}_{i_j}) : j \in [K]\}$. We assume the auditor can determine $\mathbf{x}_i$ and $\mathbf{y}_i$ for each sampled ballot. The auditor then computes a test statistic $T(A)$ and compares it to a threshold to decide if (1) more ballots should be drawn to continue the audit, (2) the election outcome is certified, or (3) all remaining ballots are counted by hand at which point the true outcome is known.

[8] This not without loss of generality — if more information is known about the reported margins, more targeted sampling can be more efficient (Stark, 2009b).

Definition 1. A auditing procedure is risk-limiting with risk level α if

$$\mathbb{P}(\text{election certified} \mid w_r \neq w_a) < \alpha.$$

An audit works by estimating amount of *ballot error* in an election, and the margin of victory is the total ballot error necessary to change the outcome of the election. The way in which ballot errors are defined varies according to the particular voting system. We discuss the definition of ballot error for scored and ranked systems separately.

Definition 2. Let $\varepsilon(\mathbf{x},\mathbf{y})$ measure the error of a ballot. The margin of a set of reported ballots $\{\mathbf{y}_i : i \in [n]\}$ is the minimum number of ballot errors necessary changes the outcome of the election

$$m = \min_{\{\mathbf{x}_i\}:\Omega(\{\mathbf{x}_i\})\neq\Omega(\{\mathbf{y}_i\})} \sum_{i=1}^{n} \varepsilon(\mathbf{x}_i,\mathbf{y}_i). \tag{1}$$

4 Scored systems

Some methods proposed for auditing elections based on plurality voting can be easily extended to single-winner elections in which voters's preferences can be interpreted as scores given to each candidate. These systems can be audited efficiently using the methods of Stark (2010) or Checkoway et al. (2010). We illustrate our ideas by extending the method of Stark to scored systems.

4.1 Errors for scored systems

In a scored voting system, we can write the true value of ballot i as a vector of scores $\mathbf{x}_i = (x_i(1), x_i(2), \ldots, x_i(k)) \in [0,R]^k$, where $x_i(j)$ is the score that voter i gives to candidate j and R is the maximum score that a voter can assign to a candidate. Ballot i is counted as $\mathbf{y}_i \in [0,R]^k$. For example, in plurality elections, ballots have values of the form $(1,0,0,\ldots,0)$, $(0,1,0,\ldots,0)$, and so on. The true and reported outcomes are

$$\mathbf{P} = \sum_{i=1}^{n} \mathbf{x}_i, \qquad \mathbf{Q} = \sum_{i=1}^{n} \mathbf{y}_i. \tag{2}$$

The reported winner and runner-up are

$$w_r = \text{argmax}_{j\in[k]}\{Q(j)\}, \tag{3}$$

$$l_r = \text{argmax}_{j\in[k]}\{Q(j) : j \neq w_r\}, \tag{4}$$

whereas the actual winner and runner-up are

$$w_a = \text{argmax}_j\{P(j)\}, \tag{5}$$

$$l_a = \text{argmax}_j\{P(j) : j \neq w_a\}, \tag{6}$$

We will write w for w_r since the auditor only knows w_r.

Definition 3. For scored systems, the error of ballot i is

$$\varepsilon(\mathbf{x}_i, \mathbf{y}_i) = \sum_{j=1}^{k} |x_i(j) - y_i(j)|. \tag{7}$$

Because scored systems are relatively simple, calculating the margin can be done directly from the reported outcome.

Definition 4. Given an election of n reported ballots $\{\mathbf{y}_i : i \in [n]\}$ with reported outcome $\mathbf{Q} = (Q(1), Q(2), \ldots, Q(k))$, and reported winner w_r and runner-up l_r, The pairwise margin between i and j is

$$m_{ij} = |Q(i) - Q(j)| \tag{8}$$

and the margin of the election is

$$m = Q(w_r) - Q(l_r). \tag{9}$$

Our definition of the margin is measured in actual scores, not fractions or percentages of the number of ballots cast. According to our definition, the margin is the lowest level of ballot error necessary to change the outcome of the election. In scored systems, the effect of individual ballots can be different.

4.2 Auditing scored systems

We now show how to generalize the method in Stark (2010) using our definition of margins. Suppose the audit has drawn K ballots uniformly from the set of n ballots. Let $(\mathbf{X}_t, \mathbf{Y}_t)$ denote the tth ballot in the sample (that is, $(\mathbf{X}_t, \mathbf{Y}_t) = (\mathbf{x}_i, \mathbf{y}_i)$ for some i drawn uniformly from $[n]$). The relative overstatement of the tth ballot between the winner w and another candidate j is

$$e_t(w, j) = \frac{\Big(Y_t(w) - Y_t(j)\Big) - \Big(X_t(w) - X_t(j)\Big)}{m_{wj}}. \tag{10}$$

For example, in elections where votes are in $\{0,1\}^k$, this is 0 when there is no error, positive (either $1/m_{wj}$ or $2/m_{wj}$) when there is an error that, when corrected,

decreases the margin, and negative (either $-1/m_{wj}$ or $-2/m_{wj}$) when there is an error that, when corrected, increases the margin.

For the tth audited ballot, the worst case relative overstatement is

$$\hat{e}_t = \max_{j \neq w} \frac{\Big(Y_t(w) - Y_t(j)\Big) - \Big(X_t(w) - X_t(j)\Big)}{m_{wj}}. \tag{11}$$

A ballot for which $\mathbf{X}_t = \mathbf{Y}_t$ has $\hat{e}_t = 0$. We have $\hat{e}_t > 0$ whenever correcting the error in the tth ballot causes any of the margins $\{m_{wj}\}$ to decrease, and not just the margin $m_{wl} = m$ in (9). For example, in an approval election, if the tth audited ballot contains votes for two candidates w and c but the CVR only counted the vote for w, correcting this error would cause m_{wc} to decrease and therefore $\hat{e}_t > 0$, even if $c \neq l$.

To use these overstatements $\hat{e}_t$ we can apply the same martingale arguments used by Stark (2009d) to compute the Kaplan-Markov P-value. The test procedure consists of sampling ballots and computing the test statistic

$$T(K) = \prod_{t=1}^{K} \frac{1 - (m/n)/(2R\gamma)}{1 - \hat{e}_t m/(2R\gamma)}. \tag{12}$$

If we choose to certify when $T(K) < \alpha$, then this procedure is a risk-limiting audit with risk level α in the sense of Definition 1 (see Stark (2009d) and Kaplan (1987)). Otherwise, more ballots can be sampled and the audit continues or all remaining ballots can be counted by hand, thus ending the audit. The parameter $\gamma > 1$ effectively shrinks the margin m (or, equivalently, inflates the error) which helps make the test statistic more robust. Experiments with plurality elections show $\gamma = 1.01$ to $\gamma = 1.1$ work well, but choosing γ may depend on the particular non-plurality system.

This analysis is loose for a number of reasons. Firstly, the sum of the overstatements $\hat{e}_t$ over t is an upper bound on the aggregate relative overstatement of the K audited ballots. The bound uses a the worst-case upper bound on the relative error $\hat{e}_t \leq 2R/m$ rather than considering the error bound ballot-by-ballot. Refining the analysis to take into account the statistics of the actual sampled ballots could yield a more efficient test. Similarly, non-uniform sampling of ballots (say according to their CVRs) could yield a more statistically efficient audit.

This approach to auditing was proposed by Stark (2010) for multi-winner plurality contests. It is easy to apply this generalized method to approval, range voting, and Borda counts.

Approval voting. In approval voting, each voter can decide to approve or disapprove of each candidate. Therefore the ballots are $\mathbf{x}_i \in \{0,1\}^k$ and thus $R = 1$. The auditing method was originally designed to work for the setting where voters could approve of up to c candidates and there were c winners, so this is a simple extension for approving of up to k candidates with 1 winner.

Range voting. In range voting users can assign a score to each candidate. These scores are typically integers, say from 0 to 10. The winner is the candidate who garners the maximum sum score from the voters. For a range voting system, the scores will be in $[0,R]$ where R is the maximum allowed score. Note that for range voting the upper bound of $2R/m$ on $\hat{e}_t$ may be significantly more conservative than for approval voting, especially if many voters do not have polarized views about all of the candidates. This, in turn, may increase the number of ballots required, lowering the efficiency of the audit.

Borda count. The Borda count is thought of as a voting system where users rank candidates. This is true in that users submit their preferences in terms of a ranked list. However, the Borda count converts this ranked list into a numerical score for each candidate, and hence can be audited by the same mechanism as other scored systems. On a ballot for an election to be tabulated by a Borda count, voters rank candidates in order of preference. In an election with k candidates, the Borda count assigns $k-s+1$ points to the sth highest ranked candidate. Thus the top-ranked candidate for the voter gets k points, the second-ranked candidate gets $k-1$ points, and so on. Voters need not rank all candidates; an unranked candidate gets 0 points. By setting $R=k$, it is clear that the Borda count is a special case of range voting.

5 Ranked systems

Unlike scored systems, ranked systems do not share a common framework for tabulation. However, the two methods we discuss in this section, Condorcet and IRV, perform simple arithmetic operations and comparisons on the ballots in order to compute the outcome of the election.

5.1 Errors for ranked systems

We again consider an election with k candidates and n ballots cast. For a set $A \subseteq [k]$, let $\Pi(A)$ denote the set of all ordered subsets of A. That is, $\Pi(A)$ contains all ranked lists of elements of A. In a ranked-choice election with k candidates, a ballot $\mathbf{x}_i = (x_i(1), x_i(2), \ldots)$ for voter i is an element of $\Pi([k])$, where $x_i(j)$ is the jth ranked candidate of voter i. The elements of $\Pi([k])$ are called *ballot signatures*. A special case of a ballot signature is a *blank ballot*, which is denoted by the empty list $()$. The ith ballot is reported as $\mathbf{y}_i$ which may differ from its true value $\mathbf{x}_i$. The election systems we discuss in this section all operate

on the counts of the election. For a ballot signature $S \in \Pi(A)$ define the count of a set S as

$$N(S) = \sum_{i=1}^{n} \mathbf{1}(\mathbf{y}_i = S). \tag{13}$$

That is, $N(S)$ is the number of ballots reported as having signature S.

We differ from Cary (2011) and Magrino, Rivest, Shen, and Wagner (2011), who define the margin of an IRV election as the number of ballots that must be changed in order to change the outcome. Instead, we define a ballot errors for IRV as follows.

Definition 5. Suppose a ballot with signature $\mathbf{x}$ is instead recorded as $\mathbf{y}$. For an IRV election, the error for this ballot is

$$\varepsilon(\mathbf{x},\mathbf{y}) = \begin{cases} 0 & \mathbf{x} = \mathbf{y} \\ 1 & \mathbf{x} \neq \mathbf{y} \text{ and } \mathbf{x} = (\,) \text{ or } \mathbf{y} = (\,) \\ 2 & \mathbf{x} \neq \mathbf{y} \text{ and } \mathbf{x}, \mathbf{y} \neq (\,) \end{cases} \tag{14}$$

where $(\,)$ is the blank ballot signature.

Thus a ballot that is correctly recorded has 0 errors and a ballot that is incorrectly recorded has 1 or 2 errors depending on the actual or reported signature being blank or not. The intuition behind this definition is that a blank ballot neither helps nor hurts any candidate whereas a ranked ballot helps or hurts some of the candidates and changing from one that helps a candidate to one that hurts the candidate is worse than changing to or from one that helps no one.

The special case of an IRV election with two candidates is equivalent to a plurality election. In this case, the margin produced by ε in Definition 5 agrees with the margin produced in the equivalent plurality election using ε in Definition 3.[9] In contrast, merely counting the number of incorrect ballots gives a margin that is almost, but not quite, half the margin in the plurality case.

5.2 Condorcet methods

To tabulate a Condorcet election, the counts are converted into pairwise preferences

$$C(i,j) = \sum_{S \in \Pi([k])} N(S) \cdot \mathbf{1}(i \text{ precedes } j \text{ in } S). \tag{15}$$

[9] To use ε in Definition 3, the ballots must first be converted from ordered lists to pairs of scores: $(1,2),(1) \mapsto (1,0)$; $(2,1),(2) \mapsto (0,1)$; and $(\,) \mapsto (0,0)$ which is to say that only the top-ranked candidate on the ballot gets a score of 1.

That is, $C(i,j)$ is the number of ballots in which i is ranked higher than j. If there exists a candidate $w \in [k]$ such that $C(w,j) > C(j,w)$ for all $j \neq w$, then candidate w is called the *Condorcet winner*. The *Condorcet graph* has vertices which are the candidates and a directed edge from i to j with weight $C(i,j)$, for each pair (i,j).

Since there are many different Condorcet methods, it is difficult to give a unified description of the margin of victory; however, there is a simple lower bound. If there is a reported Condorcet winner w, then there are $k-1$ pairwise plurality contests involving w. The margin of the Condorcet winner m_{CW} is the minimum of the margins of the plurality elections

$$m_{CW} = \min_{j \neq w} \left(C(w,j) - C(j,w) \right). \tag{16}$$

If there is no reported Condorcet winner, then define $m_{CW} = 0$.

If the reported Condorcet winner w_r is not the actual winner, then in particular, w_r is not the Condorcet winner. Therefore, there must have been at least m_{CW} ballot errors as defined for plurality voting and so m_{CW} is a lower bound on the margin m. This insight leads to our method of auditing Condorcet elections.

If there is a reported Condorcet winner w_r, then we can audit each edge connecting w_r to the other candidates in the Condorcet graph by considering a plurality election between the two candidates. Verifying that w_r is the Condorcet winner will then verify the election outcome. We can treat this as a collection of $k-1$ races with two candidates each: w_r and some $j \neq w_r$. One way to audit these is to use Stark's simultaneous auditing method (Stark, 2010). Note that for this audit we do not need consider the $\binom{k}{2} - k + 1$ other pairwise contests between $j,k \neq w_r$.

If there is no Condorcet winner, then each of the candidates is defeated in at least one pairwise election. In this case, we need to consider the particular Condorcet completion method used to determine the winner. There is a veritable menagerie of Condorcet completion methods proposed in the literature. To illustrate how auditing applies, we restrict our discussion to a few examples for which auditing is simple to describe.

Two-method systems. A two-method system elects the Condorcet winner, if one exists. If there is no Condorcet winner, than a completely different method of tabulating the ballots is used. One possible completion method to use when there is no Condorcet winner, first described by Black (1958), uses the Borda count to decide the winner. Fishburn improves on this by restricting the Borda counts to the Smith set — the smallest set of candidates such that each beats all candidates outside the set (Fishburn, 1977, Function C_1).

Auditing a two-method system involves auditing each method — the same ballots can be used for each audit. If the reported counts indicate a Condorcet winner

we can audit at risk level α using the method described above. If the reported counts indicate that there is no Condorcet winner we first audit ballots to assure that no Condorcet winner exists at risk level α_1 by simultaneously auditing k pairwise elections, one in which each candidate was reported to have lost. We need only be sure that each candidate really lost at least one pairwise election, so, for candidate j, we can choose to audit the election in which j was reported to have lost by the largest margin. This choice of elections to audit reduces the expected number of ballots to be examined by hand.

After simultaneously auditing the k pairwise elections to ensure that there is no Condorcet winner, we can audit the particular completion method (e.g., Borda count) at risk level α_2. We pick α_1 and α_2 such that

$$1 - (1 - \alpha_1)(1 - \alpha_2) \leq \alpha. \tag{17}$$

This guarantees (by the union bound) that the overall risk is no more than the target, α.

One-method systems. A one-method system is a single procedure that elects the Condorcet winner when one exists, and selects a different candidate otherwise. In the latter case, different one-method systems may elect different candidates for the same set of cast ballots. If there is a reported Condorcet winner, the election can be audited using either the general method above or by auditing the specific method used. If there is no reported Condorcet winner, then the specific method must be used.

The Nanson method (Nanson, 1882) and the related Baldwin method (Baldwin, 1926) work in rounds with one or more candidates eliminated each round, similar to instant-runoff voting, except that Borda counts determine who is eliminated. The auditing procedure is very similar to IRV (Section 5.3). The Schulze method — the most commonly used Condorcet method — is more complicated. Developing a risk limiting audit for the Schulze method is an open problem. However, most organizations which use the Schulze method do not use physical ballots or a voter-verified paper audit trail (VVPAT), so the auditing framework used here may not be applicable.

5.3 Instant-runoff voting

In an IRV election, voters also express their preferences as an ordered subset of the candidates. The counting proceeds in rounds. In each round, the candidates with the fewest top-choice votes are eliminated. Eliminating a candidate effectively removes the candidate from all ballots in which she was ranked, causing later ranked candidates to move up one spot. A candidate who is not eliminated is called a

continuing candidate. A ballot is considered *exhausted* when all of the candidates it ranks have been eliminated. The elimination stops when one candidate has a majority of top-choice votes on the nonexhausted ballots.

There are several methods for choosing the candidates to eliminate. The simplest is to eliminate the candidate with the fewest top-choice votes. This is the base IRV elimination rule. In San Francisco municipal, ranked choice voting (RCV) elections, multiple candidates can be eliminated in a single round.[10] We refer to this as the SF RCV elimination rule. In both cases, the sum of the top-choice votes for candidates chosen to be eliminated is less than the number of top-choice votes for every candidate who is not eliminated (except in the case of a tie). That is, if E is an *elimination set* — a set of candidates to be eliminated — then

$$\sum_{i=1}^{n} \mathbf{1}\Big(y_i(1) \in E\Big) < \min_{c \notin E} \sum_{i=1}^{n} \mathbf{1}\Big(y_i(1) = c\Big), \tag{18}$$

where $y_i(1)$ is the top, noneliminated choice on ballot i. We will focus on these rules, which are are provided for completeness in Algorithm 4 of Appendix A. Both rules produce the same winner, but the SF RCV rule is more efficient.

Tabulating the outcome of an IRV election produces a list $\mathcal{E}=(E_1, E_2, \ldots, E_M)$ of sets of eliminated candidates in the order in which they were eliminated. The set E_r is the set of candidates eliminated in the rth round. Under the base IRV rules, E_r is always a single candidate for $r < M$, whereas in the SF RCV rule, E_r may contain many candidates. In either case, once one candidate has a majority, the final elimination set E_M may contain multiple candidates.

Auditing the elimination order. A simple approach is to audit the elimination order $\mathcal{E}$ to verify that the set of candidates eliminated in each round is correct. In this auditing scheme, each elimination decision is treated as a plurality contest between the lowest-ranked continuing candidate and the elimination set. If any elimination selection is a result of a tie breaker[11] then a complete hand count is necessary.

Otherwise, each round of the algorithm leads to a plurality election to be audited. For each round r: (1) eliminate and distribute the votes for candidates eliminated in previous rounds, namely $E_1 \cup E_2 \cup \cdots \cup E_{r-1}$; (2) aggregate the candidates

[10] S.F., CAL., CHARTER art. XIII, § 13.102(e) (Mar. 2002), "If the total number of votes of the two or more candidates credited with the lowest number of votes is less than the number of votes credited to the candidate with the next highest number of votes, those candidates with the lowest number of votes shall be eliminated simultaneously and their votes transferred to the next-ranked continuing candidate on each ballot in a single counting operation."

[11] For example, with the base IRV elimination rule, if the two candidates with the fewest number of top-choice votes in a round have the same number of votes, then the candidate to be eliminated may be chosen by some other mechanism such as a coin flip.

who are to be eliminated in round r, namely those in E_r, into a "super candidate"; and (3) audit a $(k'-1)$-winner plurality election with k' candidates consisting of the super candidate and the $k'-1$ continuing candidates. The audit in step (3) is to ensure that the super candidate lost. This procedure results in M plurality-elections to audit. The M plurality-elections can be audited simultaneously using Stark's method (Stark, 2010). Each ballot can cause 0, 1, or 2 errors for each of the M plurality-elections; however, due to the nature of the *diluted margin* in Stark's method, we take the maximum of the errors caused in any race as the error contributed by the ballot. Note that this ignores the correlations in the races and hence may result in a loss of statistical efficiency; more careful modeling could produce a more efficient audit.

Auditing the elimination order requires only the round-by-round tallies and not the full information $N(S)$, but in general may require that too many ballots be hand-counted. This is because candidates who are eliminated early often constitute a very small fraction of the total ballots. For example, in the 2010 Oakland Mayoral election, three candidates each received less than 1% of the votes. This led to a small margin of 83 votes in round 3 out of a total of 122,264 ballots cast in the election. Small pairwise margins for candidates eliminated early-on in the counting requires large sample sizes to detect an error in the elimination order. If instead of the base elimination rule, the SF RCV rule is used, then 8 of 11 candidates are eliminated in the first round and the smallest margin used for the audit is 1,627, or 1.33% of the cast votes. We will return to this example in Section 6.2.

Auditing by error detection. An alternative approach to building a risk-limiting audit is to attempt *error detection*. That is, the auditor can sample K ballots and compare each paper ballot to its cast vote record (CVR). If the number of ballot errors exceeds a specified threshold, then a manual count of the entire election is required. This approach treats all erroneous ballots as if they decreased the margin, which is wasteful. Indeed, there may be elections in which the total amount of ballot error is quite large but for which the reported and actual winners of the election are the same.

Suppose that the margin is m. The effect of auditing by error detection is to audit a fictitious plurality contest between two candidates whose margin is m. Therefore any method for auditing plurality contests may be adapted for the purposes of error detection. Such an audit can be performed using any of the standard methods (Johnson, 2004, Neff, 2003, Saltman, 1975) by treating all errors as being as bad as possible, or via the methods of Checkoway et al. (2010) or Stark (2010) by distinguishing between ballots with 1 or 2 errors (a ballot added or removed has 1 error; a ballot changed has 2 errors). If fewer erroneous CVRs are found than the

threshold, the auditor certifies the winner of the election. We choose the threshold so that the sample-size is risk-limiting.

This approach to risk-limiting audits requires computing the margin of an IRV election, which is a topic of recent interest (Cary, 2011, Magrino et al., 2011). Once the margin or a lower bound on it is known, then we can set the threshold to guarantee a risk-limiting audit. Recent work by Magrino et al. (2011) calculates the number of ballots that need to be changed to change the winner of the election. This exact calculation can be computationally very expensive, even with clever heuristics.

The difficulty of auditing IRV. Because the IRV elimination rules are somewhat complicated, it is unclear what a random sample of the ballots tells us about what the rule would produce on the true ballots. In elections where voters must rank all candidates, the number of ordered subsets S (or the size of $N(\cdot)$) is potentially greater than $k!$, and the empirical distribution of the ballots in the audit will not be a good approximation to $N(\cdot)$. The approaches described above convert the IRV election into plurality contests to take advantage of the rich literature on auditing, but these conversions are less statistically efficient because they audit sufficient conditions (the elimination order remaining the same) or make conservative assumptions (every error decreases the margin). In the next section we examine ways of estimating the true margin of an IRV election and give empirical results for real elections.

6 The margin of an IRV election

In this section we investigate the problem of computing the margin for an IRV election. We first describe some real IRV elections and their features. We then show a lower bound on the margin based on picking elimination sets in each round in such a way as to maximize the difference in votes between the "super candidate" described in Section 5.3 and the continuing candidate with the fewest votes. In order to evaluate how good this lower bound is, we develop an algorithm that constructs a set of ballot errors that can alter the winner of an IRV election. This gives an upper bound which is often close to the lower bound in real elections. Our bounds are fast to compute, and when possible we compare our bounds to the exact margins.

Appendix B contains several toy examples showing unintuitive features of IRV elections related to auditing: a few ballot errors for losing candidates can change the outcome, and even when IRV elects the Condorcet winner, the IRV margin can be significantly smaller than the Condorcet margin lower bound.

6.1 Margins for real elections

We purchased CVR data from OpenSTV[12] for six different elections that were conducted using ranked-choice ballots. The three 2002 Dáil Éireann elections — Dublin North, Dublin West, and Meath — are multiple winner STV elections with many candidates which we include to test the speed of our algorithms, not because they are necessarily representative of IRV elections. The others — Burlington mayoral and Takoma Park city council — are IRV elections. Mike LaBonte provided the official 2009 Aspen election data acquired through an open records request. The Aspen elections used rules similar to IRV but with different first-round rules. Similar to the STV elections, we use them merely as an example of ranked ballots. CVR data for an additional 29 San Francisco Bay Area and Pierce County RCV elections were collected from the corresponding municipalities' websites. A summary of the data is given in Table 1.

The three Dáil Éireann, two Burlington mayoral, two Aspen, and Takoma Park City Council special elections allowed voters to provide a complete ranking of all of the candidates on the ballot. The last five additionally allowed write-in candidates although in the case of the Burlington and Aspen elections, the write-in took the place of one of the candidates in the ranking. All of the California and Washington elections used ballots where voters pick their top three candidates, including write-ins.

One common feature of all of these elections is that they involve a relatively small number of ballots compared to state and national elections. As an extreme example, only 204 people voted in the election for the Takoma Park City Council. In such cases a full hand count is easy, and would certainly be risk limiting. After tabulating the results from these elections we were surprised to note that they share a more interesting common feature: The winner according to the IRV count was also a Condorcet winner for the election in every case except for the 2009 Burlington mayoral election. In Appendix B.2 we show that the IRV margin may be smaller than the Condorcet margin lower bound, even when both methods elect the same candidate.

Computing the IRV margins exactly can be a computationally difficult task for real elections that contain large numbers of candidates or allow voters to rank many candidates on the ballot (Magrino et al., 2011). Therefore, in the rest of this section we present lower and upper bounds on the margin and examine the bounds for the 34 elections. We implement a slightly modified version of the exact margin calculation from Magrino et al. (2011) which takes the difference in margin definitions into account and additionally uses knowledge of the margin upper bound

[12] `http://www.openstv.org`

Table 1: Election data.

Election	Candidates	Ranks	Ballots	Condorcet winner
2002 Dáil Éireann, Dublin North*	12	12	43,942	✓
2002 Dáil Éireann, Dublin West*	9	9	29,988	✓
2002 Dáil Éireann, Meath*	14	14	64,081	✓
2006 Burlington mayor	6‡	5	9,865	✓
2007 San Francisco mayor	18	3	149,465	✓
2007 Takoma Park city council special, ward 5	4‡	4	204	✓
2008 Pierce County assessor	7‡	3	312,771	✓
2008 Pierce County council, dist. 2	4‡	3	43,661	✓
2008 Pierce County executive	5‡	3	312,771	✓
2009 Aspen city council†	11§	9	2,544	✓
2009 Aspen mayor†	5‡	4	2,544	✓
2009 Burlington mayor	6‡	5	8,984	
2009 Pierce County auditor	4‡	3	159,987	✓
2010 Berkeley auditor	2‡	3	45,986	✓
2010 Berkeley city council, dist. 1	5‡	3	6,426	✓
2010 Berkeley city council, dist. 4	5‡	3	5,708	✓
2010 Berkeley city council, dist. 7	4‡	3	4,862	✓
2010 Berkeley city council, dist. 8	4‡	3	5,333	✓
2010 Oakland auditor	3‡	3	122,268	✓
2010 Oakland city council, dist. 2	3‡	3	15,243	✓
2010 Oakland city council, dist. 4	8‡	3	23,884	✓
2010 Oakland city council, dist. 6	4‡	3	14,040	✓
2010 Oakland mayor	11‡	3	122,268	✓
2010 Oakland school board director, dist. 2	2‡	3	15,243	✓
2010 Oakland school board director, dist. 4	3‡	3	23,884	✓
2010 Oakland school board director, dist. 6	2‡	3	14,040	✓
2010 San Francisco board of supervisors, dist. 2	7‡	3	28,911	✓
2010 San Francisco board of supervisors, dist. 6	15‡	3	25,057	✓
2010 San Francisco board of supervisors, dist. 8	5‡	3	38,551	✓
2010 San Francisco board of supervisors, dist. 10	22‡	3	20,550	✓
2010 San Leandro city council, dist. 1	4‡	3	23,494	✓
2010 San Leandro city council, dist. 3	2‡	3	23,494	✓
2010 San Leandro city council, dist. 5	3‡	3	23,494	✓
2010 San Leandro mayor	7‡	3	23,494	✓
2011 San Francisco district attorney	6‡	3	197,242	✓
2011 San Francisco mayor	25‡	3	197,242	✓
2011 San Francisco sheriff	5‡	3	197,242	✓

* Multiseat STV elections that have been treated as IRV.
† IRV-like rules that have been treated as IRV.
‡ Includes a single combined write-in candidate.
§ Includes two combined write-in candidates.
The Ranks column denotes how many candidates a voter was allowed to rank on the ballot.
There is a ✓ in the Condorcet winner column if the IRV procedure elects the Condorcet winner.

Table 2: Margin bounds from real elections using ranked-choice ballots.

Election	Lower bound	%	Exact margin	%	Upper bound	%	Condorcet bound	%
2002 Dáil Éireann, Dublin North	203	0.5	—	—	5,670	12.9	2,723	6.2
2002 Dáil Éireann, Dublin West	8	0.0	1,443	4.8	1,444	4.8	1,443	4.8
2002 Dáil Éireann, Meath	129	0.2	—	—	8,256	12.9	6,197	9.7
2006 Burlington mayor	482	4.9	776	7.9	778	7.9	776	7.9
2007 San Francisco mayor	68,060	45.5	—	—	101,676	68.0	101,674	68.0
2007 Takoma Park city council special, ward 5	36	17.6	36	17.6	38	18.6	36	17.6
2008 Pierce County assessor	276	0.1	2,221	0.7	7,300	2.3	2,221	0.7
2008 Pierce County council, dist. 2	4,014	9.2	4,014	9.2	4,016	9.2	4,014	9.2
2008 Pierce County executive	4,054	1.3	4,054	1.3	4,056	1.3	4,054	1.3
2009 Aspen city council	8	0.3	70	2.8	324	12.7	70	2.8
2009 Aspen mayor	177	7.0	177	7.0	178	7.0	125	4.9
2009 Burlington mayor	253	2.8	253	2.8	254	2.8	588	6.5
2009 Pierce County auditor	16,792	10.5	16,792	10.5	16,794	10.5	16,792	10.5
2010 Berkeley auditor	30,711	66.8	30,711	66.8	30,712	66.8	30,711	66.8
2010 Berkeley city council, dist. 1	1,770	27.5	2,348	36.5	2,350	36.6	2,348	36.5
2010 Berkeley city council, dist. 4	777	13.6	1,033	18.1	1,034	18.1	1,033	18.1
2010 Berkeley city council, dist. 7	728	15.0	728	15.0	730	15.0	728	15.0
2010 Berkeley city council, dist. 8	1,011	19.0	1,756	32.9	1,758	33.0	1,756	32.9
2010 Oakland auditor	33,045	27.0	34,162	27.9	34,164	27.9	34,162	27.9
2010 Oakland city council, dist. 2	4,349	28.5	4,349	28.5	4,350	28.5	4,349	28.5
2010 Oakland city council, dist. 4	131	0.5	4,657	19.5	4,658	19.5	4,657	19.5
2010 Oakland city council, dist. 6	3,653	26.0	5,205	37.1	5,206	37.1	5,205	37.1
2010 Oakland mayor	2,025	1.7	2,025	1.7	2,026	1.7	2,025	1.7
2010 Oakland school board director, dist. 2	9,660	63.4	9,660	63.4	9,662	63.4	9,660	63.4
2010 Oakland school board director, dist. 4	7,089	29.7	7,239	30.3	7,240	30.3	7,239	30.3
2010 Oakland school board director, dist. 6	9,651	68.7	9,651	68.7	9,652	68.7	9,651	68.7
2010 San Francisco board of supervisors, dist. 2	258	0.9	258	0.9	260	0.9	258	0.9
2010 San Francisco board of supervisors, dist. 6	1	0.0	—	—	1,338	5.3	1,337	5.3
2010 San Francisco board of supervisors, dist. 8	3,552	9.2	3,552	9.2	3,554	9.2	3,552	9.2
2010 San Francisco board of supervisors, dist. 10	2	0.0	—	—	444	2.2	286	1.4
2010 San Leandro city council, dist. 1	6,262	26.7	6,262	26.7	6,264	26.7	6,262	26.7
2010 San Leandro city council, dist. 3	16,675	71.0	16,675	71.0	16,676	71.0	16,675	71.0
2010 San Leandro city council, dist. 5	1,484	6.3	1,484	6.3	1,486	6.3	1,484	6.3
2010 San Leandro mayor	232	1.0	232	1.0	234	1.0	232	1.0
2011 San Francisco district attorney	3,357	1.7	16,646	8.4	39,490	20.0	16,646	8.4
2011 San Francisco mayor	19	0.0	—	—	27,298	13.8	22,971	11.6
2011 San Francisco sheriff	1,137	0.6	11,455	5.8	11,456	5.8	11,455	5.8

Exact margins of — mean that we were unable to compute the margin within 24 hours.

Algorithm 1: Lower bound on IRV margin

inputs: candidates A, ballots B
outputs: lower bound on margin lb

`Enqueue(`∞, (A,B)`)`
while true do
  $p, (A,B) \leftarrow$ `Dequeue()`
  if $|A| = 1$ **then**
    $lb \leftarrow p$
    return
  $ESets \leftarrow$ `ValidElimSets(`A,B`)`
  foreach $E \in ESets$ **do**
    $A' \leftarrow A \setminus E$
    $p' \leftarrow \min\left\{p, \min_{c \in A'} \sum_{i=1}^{n} \mathbf{1}\left(y_i(1) = c\right) - \sum_{i=1}^{n} \mathbf{1}\left(y_i(1) \in E\right)\right\}$
    $B' \leftarrow$ `EliminateCandidates(`B, E`)`
    `Enqueue(`p', (A',B')`)`

to speed up the calculation.[13] Where we are able to compute the exact margin in under 24 hours, we compare the bounds with the margin.

6.2 Lower bounds on the margin

One obvious lower bound on the margin in an IRV election is the difference in votes between the two candidates with the fewest top-choice votes in each round. Certainly if the number of ballots which are modified is not enough to change any elimination decision, then the outcome must be correct. However, it is trivial to show that this lower bound is arbitrarily bad by considering an election in which there are two candidates who each receive exactly one vote.

The example of the 2010 Oakland Mayoral race mentioned in Section 5.3 shows that different choices of elimination sets can lead to different margins in the constructed plurality elections used for the audit. Instead of auditing the elimination order by considering the actual elimination sets dictated by the election rules as described in Section 5.3, one can choose different, valid elimination sets satisfying (18) for each round in such a way as to maximize the margin of the constructed plurality elections. This also provides a lower bound on the IRV margin that can be used to provide a conservative estimate for auditing by error detection. In the

[13] All of our code is available at `https://www.cs.jhu.edu/~s/elections/irv.html`.

2010 Oakland Mayoral election, if the 7 lowest candidates are eliminated in the first round instead of the 8 chosen by the SF RCV rule, then the smallest margin used in the audit rises to 2,025 from 1,627.

The idea to pick the best elimination sets to use comes directly from David Cary's IRV lower bound computation in simultaneous, independent work (Cary, 2011). For concreteness, we describe the lower bound algorithm and its correctness below but refer the interested reader to Cary's work for a more complete treatment as well as an alternative implementation.

The obvious lower bound is the number of ballot errors necessary to change the order that candidates are eliminated. However, by definition of the elimination set, in any round, any valid choice of elimination set can be chosen and those candidates eliminated without changing the ultimate winner of the election (this is the basis of the SF RCV elimination rule). This is a relaxation on the order in which candidates must be eliminated to ensure the correct outcome. In the Oakland Mayoral election, since the 8 candidates with the fewest top-choice votes in the first round can be eliminated without changing the winner, it is immaterial in which order those candidates are eliminated so long as they are eliminated before any others.

Any sequence of valid elimination sets satisfying (18) thus gives a lower bound on the margin: if $\mathcal{E} = (E_1, E_2, \ldots, E_M)$ is a valid sequence of elimination sets, then

$$lb_{\mathcal{E}} = \min_{E \in \mathcal{E}} \left(\min_{c \notin E} \sum_{i=1}^{n} \mathbf{1}\Big(y_i(1) = c\Big) - \sum_{i=1}^{n} \mathbf{1}\Big(y_i(1) \in E\Big) \right) \tag{19}$$

is a lower bound. If each $E_r \in \mathcal{E}$ consists of a single candidate, then $lb_{\mathcal{E}}$ is the obvious lower bound. Algorithm 5 in Appendix A calculates the valid elimination sets of a single stage of the algorithm.

Since each valid $lb_{\mathcal{E}}$ is a lower bound, we can take the maximum over all valid $\mathcal{E}$ to arrive at the bound

$$lb = \max_{\text{valid } \mathcal{E}} lb_{\mathcal{E}}. \tag{20}$$

The bound lb can be efficiently computed by using a priority queue[14] whose elements are pairs of sets of candidates and sets of ballots. The queue initially contains the pair (A, B) — where A is the set of candidates in the election and B is the set of ballots — with an infinite priority as its only element. The main loop removes the pair of candidates and ballots with the highest priority p. If all but one of the

[14] A *priority queue* is an abstract data type which is conceptually a set of elements each of which has an associated priority. Common implementations of priority queues support fast insertion of a new element with arbitrary priority and fast removal of the element with the highest priority.

candidates have been eliminated, then the priority is returned as the lower bound. Otherwise, all valid elimination sets are computed. For each valid elimination set, a copy of the ballots is constructed, the candidates in the set are eliminated, and the new ballots are placed into the queue with priority p' where p' is the minimum of p and the difference in votes between the sum of top-choice votes for candidates in the elimination set and the top-choice votes for the continuing candidate with the fewest votes. This procedure is given in Algorithm 1. Since we are using a priority queue, once we reach a set of ballots for which all candidates but one have been eliminated, every other sequence of elimination sets leads to a lower bound that is no better.

Note that the sequence of elimination sets used to construct the lower bound in Algorithm 1 is the optimal set to use when performing an audit of an IRV election by considering plurality elections for each round as described in Section 5.3. It is trivial to modify Algorithm 1 to return the sequence of eliminations used.

The weakness in the lower bound is that it considers the slimmest margin in any elimination decision. However, the margin between two candidates in a given round can be quite low but the candidates together have too many votes to be grouped into an elimination set. For example, in the 2010 Oakland City Council, District 4 election, in round 5 — using the base IRV elimination rule — Melanie Helby had 3,017 top-choice votes and Daniel Afford had 2,886. Neither can be eliminated in an earlier round using a larger elimination set and they cannot be eliminated together leading to a lower bound of 133 which is about 0.5% of the total number of ballots cast in the election. By contrast, the exact margin is 5,657 or about 19.5% of the ballots cast.

6.3 Algorithmic upper bound on the margin

In this section we develop an algorithm that takes a set of CVRs and constructs a set of ballot errors that changes the winner of the IRV election. This gives an upper bound on the margin of the election. This upper bound, which is efficiently computable, is useful to bound how far the lower bound described in the previous section is from the exact value if the exact value is not known.

Our method, Algorithm 2, is based on calculating an upper bound on the margin for each possible alternative winner of the election. For a given alternative winner j, we calculate a sufficient number of errors required to make j the winner of the election. Because our algorithm is "greedy" in a sense, the total number of errors we calculate may be quite a bit larger than the minimum number of material errors needed to change the outcome.

The algorithm proceeds as follows: for each alternative j, we tabulate the IRV election round by round until j is eliminated. Let E be the elimination set such

Algorithm 2: Greedy upper bound on IRV margin

inputs: candidates A, ballots $B = \{\mathbf{y}_i \in \Pi(A)\}_{i=1}^n$
outputs: upper bound on margin $\hat{m}$
$\langle$ *underscore _ means the corresponding output of* IRV *is not used* $\rangle$

$\mathsf{Winner}, \mathsf{ElimOrder}, _ \leftarrow$ IRV $(A,\ |A|,\ B)$
foreach $j \in A \setminus \{\mathsf{Winner}\}$ **do**
  $(\mathsf{Winner}', A', B', \mathsf{ElimOrder}') \leftarrow (\mathsf{Winner}, A, B, \mathsf{ElimOrder})$
  $e_j = 0$
  while $\mathsf{Winner}' = \mathsf{Winner}$ **do**
    $l \leftarrow$ index of E in $\mathsf{ElimOrder}'$ such that $j \in E$
    $_, _, B' \leftarrow$ IRV $(A',\ l-1,\ B')$ $\langle$ *tabulate up to when j is eliminated* $\rangle$
    foreach $E \in \mathsf{ElimOrder}'(1 : l-1)$ **do**
      $A' \leftarrow A' \setminus E$ $\langle$ *remove eliminated candidates* $\rangle$
    $k \leftarrow \operatorname{argmin}_{c \in A' \setminus \mathsf{ElimOrder}'(l)} \sum_{\mathbf{y}' \in B'} \mathbf{1}(y'(1) = c)$
    $s \leftarrow \sum_{\mathbf{y}' \in B'} \mathbf{1}(y'(1) = k)$
    $\mu \leftarrow s - \sum_{c \in \mathsf{ElimOrder}'(l)} \sum_{\mathbf{y}' \in B'} \mathbf{1}(y'(1) = c)$
    if $|\mathsf{ElimOrder}'(l)| > 1$ **then**
      $\mu \leftarrow s - 1$
    $B', e \leftarrow$ ModifyMargin $(B',\ \mu,\ j,\ k,\ \mathsf{ElimOrder}'(l : end),\ \mathsf{Winner})$
    $e_j \leftarrow e_j + e$
    $\mathsf{Winner}', \mathsf{ElimOrder}', _ \leftarrow$ IRV $(A',\ |A'|,\ B')$
$\hat{m} \leftarrow 2 \min_j \{e_j\}$

that $j \in E$; we can calculate elimination sets using the SF RCV rule or the base rule. At the point during tabulation when the candidates in E are eliminated, let k be the candidate with the fewest top choice votes such that $k \notin E$. Let t be the total number of top-choice votes for candidates in E and s be the total number of top-choice votes for k and let $\mu = s - t$.

Now, the goal is to change enough votes so that j will not be in the elimination set. There are two cases. If E is a singleton ($E = \{j\}$), then the only way to prevent j from being eliminated is for j to receive more top-choice votes than k, so the margin of this decision is μ and shifting $\lfloor \mu/2 \rfloor + 1$ votes from k to j will give j enough votes to not be eliminated in this round. If E has more than one candidate, then due to the strict inequality in (18), shifting $\lfloor (s-1)/2 \rfloor + 1$ votes from k to j is enough to remove j from the elimination set. To see this, first note that this case arises only in the SF RCV rule. If B' is the set of partially tabulated ballots, then from (18),

$$\sum_{\mathbf{y}' \in B'} \mathbf{1}(\mathbf{y}' \in E \cup \{k\}) > \min_{c \notin E \cup \{k\}} \sum_{\mathbf{y}' \in B'} \mathbf{1}(\mathbf{y}' = c),$$

Algorithm 3: `ModifyMargin` – choosing errors to decrease the margin

inputs: ballots $B = \left\{\mathbf{y}_i \in \Pi(A)\right\}_{i=1}^{n}$, margin m, recipient j, victim k, ElimOrder, Winner

outputs: modified $\{\mathbf{y}_i\}$ such that the margin between j and k has decreased by more than m, number of ballots changed c

$c \leftarrow \lfloor m/2 \rfloor + 1$

$\sigma \leftarrow \Big(\mathsf{ElimOrder}(2:\text{end}) \parallel \mathsf{Winner} \parallel \mathsf{ElimOrder}(1) \setminus \{j\} \parallel j\Big)$ $\langle$ σ *is an ordered list* $\rangle$

$\sigma \leftarrow (k) \parallel \sigma \setminus \{k\}$ $\langle$ *Move k to beginning of* σ $\rangle$

Sort $\{\mathbf{y}_i\}$ lexicographically according to σ, where longer matches appear before shorter matches; e.g.,

$\Big(\sigma(1), \sigma(2), \sigma(3)\Big)$ precedes $\Big(\sigma(1), \sigma(2)\Big)$

Change the first c of the sorted $\mathbf{y}_i$ into votes with j as the only choice

so the new elimination set after shifting ballots is a strict subset of $E \cup \{k\}$. Furthermore, $s-1$ is an upper bound on the difference in top-choice votes between j and any other candidate in $E \cup \{k\}$, so by adding $\lfloor (s-1)/2 \rfloor + 1$ votes to j from k, candidate j will become the winner of a plurality election between the candidates in $E \cup \{k\}$ and therefore will not be in the new elimination set.

Once we have manipulated the ballots to prevent j from being eliminated, we tabulate the IRV election with the modified votes and find the next point where j appears in an elimination set. We repeat the margin modification process until a different candidate than w_r is elected. Note that it need not be j who is elected, since it is sufficient that *any* candidate other than the reported winner win.

The greedy part of the algorithm comes from how we choose the ballots for k that are switched to ballots for j. The selection heuristic is given in Algorithm 3. The method changes ballots of the form $(k, \ldots)$ to ballots with first-choice equal to j. We can write the elimination order $\mathcal{E} = (E_1, E_2, \ldots, E_R)$. The intuition is that for j to win, she must defeat the other candidates in each of the sets $E_1, E_2, E_3, \ldots, E_R$ and finally w_r. The heuristic is to preferentially change ballots closest to the elimination order of the election. This corresponds to lexicographically ordering the cast vote records as a function of $\mathcal{E}$ (see the definition of σ in Algorithm 3).

There are other heuristics possible for selecting which ballots to shift. For example, since j must eventually defeat the reported winner w_r, it may be better to change preferentially ballots of the form $(k, w_r, \ldots)$, thereby greedily reducing the margin between j and w_r. Another set of heuristics can be derived by looking at the Condorcet graph of the ballots in round r and greedily ordering the ballots to be

changed by the Condorcet margin. Since any heuristic generating a set of errors that alter the outcome of the election is a valid upper bound on the margin, we could take the minimum of margins generated by Algorithm 2 with each ballot ordering.

6.4 Revisiting margins for real elections

Table 2 shows the results of margin calculations for the 32 elections in Table 1. We show four margin calculations: the lower bound of Section 6.2, exact margins (when possible), the upper bound of 6.3, and the margin lower bound corresponding to treating the election as a Condorcet election.

In some elections, the lower bound produces a margin which is less than 0.5%, which is the threshold for a recount in many jurisdictions.[15] Because the three 2002 Dáil Éireann elections were for multiple-winner STV elections, the small values for the lower bound may not be representative. However, for two of the San Francisco Board of Supervisors elections, the lower bound produced a margin that is essentially zero whereas the upper bound is a significant fraction of the number of ballots cast.

For those elections where the exact margin was computable in 24 hours, our upper bound is frequently within two ballots of the exact margin. The difference is because our upper bound produces a set of errors that cause a different candidate to win outright whereas the exact calculation produces a tie. The tightness of our upper bound suggests that the heuristic that we used is a good one. For most of these elections, the exact margin is just the margin in the last round and in nearly every case is the Condorcet margin lower bound. This suggests that these elections had at most two viable candidates and in the end, the election came down to the difference in their popularity. In such cases, we suspect that the winner of an IRV election will be the Condorcet winner and that the margin will be the Condorcet margin lower bound.

Many of the elections studied have margins which are quite large, which suggests that these elections should be efficiently auditable. However, these are not, in many cases, hotly contested elections, e.g., the 2010 Oakland School Board Director for District 6 was the only candidate on the ballot. For larger or more competitive elections, it is unclear if the IRV margin would still be so large.

7 Conclusions

Alternatives to plurality voting like instant-runoff voting, if deployed, should be accompanied by risk-limiting audits. In this paper, we have initiated the study of

[15] Cf., ALA. CODE §17-16-20 (2010) or FLA. STAT. §102.141 (2010).

risk-limiting audit procedures for these alternative single-winner election systems. For scored systems we generalized a plurality auditing technique, and for Condorcet elections with a Condorcet winner we can audit a set of plurality elections. The method we describe is adapted from Stark (2010) and contains many simplifications which could have an impact on the statistical efficiency of this auditing procedure.

For instant-runoff voting, the question appears to be more complicated, and we propose two methods for auditing IRV elections. Auditing the elimination order constructs a set of multi-winner plurality elections based on the elimination sets used to eliminate candidates in each round and audits each of those. Auditing by error detection involves making an initial calculation or bound on the margin of the election and then assumes all ballot errors increase the margin. Except for auditing by error detection — which is essentially the original ballot-level method for auditing plurality election — knowledge of the margin for an IRV election is not sufficient for auditing.

An important question is whether the effects seen in the relatively small elections for which we have data will be present in larger state or national elections. Analyzing data from such elections (or polling data) could be quite valuable. As with plurality elections, there is a dearth of information on the nature and distribution of real ballot errors. Such data could be used to optimize the statistical efficiency of an auditing procedure.

We do not believe that our proposed auditing procedures are the last word on risk-limiting audits of alternative election systems; we hope that future work will provide simpler and more efficient audits. As things stand now, however, some voting systems (e.g., the Borda count and range voting) appear to be substantially easier to audit than others (e.g., IRV). We believe that ease and efficiency of auditing should be criteria when adopting a voting system.

Acknowledgments

We thank Eric Rescorla and Joseph Lorenzo Hall for numerous helpful discussions. Philip Stark for suggesting a universal definition of margin of victory, Mike LaBonte for providing Aspen election data and Michael Byrne, David Cary, Stephen Goggin, Thomas R. Magrino, Ronald L. Rivest, Emily Shen, and David Wagner for sharing advance copies of their papers. The manuscript improved greatly from the careful reading and comments of the reviewers. We thank the editor, Justin H. Gross, for his extreme patience during our revision process.

This material is based upon work supported by National Science Foundation under grants CNS-0831532 and CNS-0963702 and by the MURI program under AFOSR Grant No. FA9550-08-1-0352.

Algorithm 4: `IRV` – tabulating IRV results

inputs: candidates A, rounds ρ, ballots $B = \{\mathbf{y}_i \in \Pi(A)\}_{i=1}^n$
outputs: Winner, ElimOrder, modified $\{\mathbf{y}_i\}$

Winner $\leftarrow 0$
ElimOrder $\leftarrow ()$
$r \leftarrow 0$
while $r < \rho$ **do**
  $r \leftarrow r+1$
  foreach $c \in A$ **do**
    $Q(c) \leftarrow \sum_{i=1}^n \mathbf{1}\big(\mathbf{y}_i(1) = c\big)$
  if $\max_c\Big\{Q(c)\Big\} > \frac{1}{2}\sum_c Q(c)$ **then**
    Winner $\leftarrow \operatorname{argmax}_c\Big\{Q(c)\Big\}$
    Append $A \setminus \{\text{Winner}\}$ to ElimOrder
    break
  else
    $E \leftarrow$ `EliminationSet(`A, $\mathbf{y}_i$`)`
    Append E to ElimOrder
    $A \leftarrow A \setminus E$
    foreach $\mathbf{y}_i$ **do**
      $\mathbf{y}_i \leftarrow \mathbf{y}_i \setminus E$

A IRV tabulation algorithm

One way to perform IRV tabulation is described in Algorithm 4. It takes as input the set of candidates A, the maximum number of rounds of the algorithm to perform ρ, and the set of ballots $\{\mathbf{y}_i \in \Pi(A)\}_{i=1}^n$. It iteratively eliminates the candidates with the fewest top choice votes — $\mathbf{y}_i(1)$ is the top, continuing candidate on ballot i — by removing the candidates from every ballot on which they appear. As output, it produces the winner, if any, after ρ rounds, the set of candidates eliminated in each round, and the modified set of ballots after candidates have been eliminated.

As discussed in Section 5.3, there several rules for choosing which candidates to eliminate in each round. By abstracting the choice of the elimination set, all varieties of IRV can be described at once. The function `EliminationSet(`A, B`)` takes the set of candidates A and the set of ballots B and returns the set of candidates to be eliminated next. For example, using the base IRV elimination rule, `EliminationSet(`A, B`)` returns a singleton set consisting of the candidate with the fewest top-choice votes. The SF RCV elimination rule returns largest set of

Algorithm 5: `ValidElimSets` – return valid elimination sets

inputs: candidates A, ballots $B = \{\mathbf{y}_i \in \Pi(A)\}_{i=1}^n$
outputs: ElimSets

ElimSets $\leftarrow ()$
$E \leftarrow ()$
$r \leftarrow 0$
$s \leftarrow \sum_{c \in A} \sum_{i=1}^n \mathbf{1}(y_i(1) = c)$
while $r < s - r$ **do**
    $d \leftarrow \text{argmin}_{c \in A} \sum_{i=1}^n \mathbf{1}(y_i(1) = c)$
    $r \leftarrow r + \sum_{i=1}^n \mathbf{1}(y_i(1) = d)$
    if $r < s - r$ **then**
        Append d to E
        Append E to ElimSets

candidates E such that the sum of the top-choice votes for all candidates in E is less than the number of top-choice votes for all of the candidates not in E, S.F., CAL., CHARTER art. XIII, § 13.102(e) (2002).

Rather than iterating over each ballot every time, one can pick smarter representations such as keeping track of how many ballots with each particular candidate ranking exist or using tree data structure in which paths from the root to a node correspond to candidate rankings (O'Neill, 2006). Using a tree, eliminating a candidate involves recursively removing nodes corresponding to that candidate and merging their children.

The function `ValidElimSets(`A, B`)` in Algorithm 5 takes a set of candidates A and a set of ballots B and returns all sets of candidates E that satisfy (18). This function is used in the construction of the IRV lower bound in Section 6.2 as well as auditing the elimination order in Section 5.3. Under the base IRV elimination rule, `EliminationSet(`A, B`)` returns the smallest element of `ValidElimSets(`A, B`)` whereas under the SF RCV rule, it returns the largest.

B Examples for IRV margins

In this appendix we give some toy examples of IRV elections that illustrate two points. First, a small number of errors can dramatically change the outcome of an IRV election, even when the final round margin between the final two candidates is quite large. Secondly, the IRV margin can be smaller than the Condorcet margin lower bound, even when IRV elects the Condorcet winner.

Table 3: Unmodified six candidate, 1000 ballot IRV election.

Candidate	Round 1	Round 2	Round 3	Round 4	Final
Ulric	199: U	199: U	199: U	199: U	199: U
				180: ~~W~~ U	180: ~~W~~ U
Velma	200: V	200: V	200: V	200: V	200: V
			170: ~~X~~ V	170: ~~X~~ V	170: ~~X~~ V
					126: ~~Y~~ ~~X~~ V
Wilard	180: W U	180: W U	180: W U	—	—
Xavier	170: X V	170: X V	—	—	—
Yolanda	126: Y X V	126: Y X V	126: Y ~~X~~ V	126: Y ~~X~~ V	—
		125: ~~Z~~ Y	125: ~~Z~~ Y	125: ~~Z~~ Y	
Zoë	125: Z Y	—	—	—	—

Table 4: IRV election in Table 3 with a single Y X V ballot changed to Z Y.

Candidate	Round 1	Round 2	Round 3	Round 4	Final
Ulric	199: U	199: U	199: U	199: U	199: U
				180: ~~W~~ U	180: ~~W~~ U
Velma	200: V	200: V	200: V	200: V	—
Wilard	180: W U	180: W U	180: W U	—	—
Xavier	170: X V	170: X V	170: X V	170: X V	170: X ~~V~~
		125: ~~Y~~ X V	125: ~~Y~~ X V	125: ~~Y~~ X V	125: ~~Y~~ X ~~V~~
Yolanda	125: Y X V	—	—	—	—
Zoë	126: Z Y	126: Z ~~Y~~	—	—	—

B.1 IRV can be sensitive to small errors

IRV is sensitive to errors, in the following sense: Switching even a single vote from one losing candidate to another (or fabricating a vote for a losing candidate) may be enough to change the winner of an election.[16] We illustrate this via a simple example. Consider the six candidate, 1000 ballot election in Table 3. Zoë has the fewest votes of any candidate. She is eliminated in the first round and ultimately Velma wins with 496 votes. Ulric comes in second with 379 votes. Naïvely, one

[16] I. D. Hill describes a slightly different example of instability in a real Single Transferable Vote election — the multiseat analogue of instant-runoff voting. Hill points out that a change in a single ballot's 15th choice (out of 23) would result in a different winner. In this case, it was the difference between voting for one of the (eventual) winners and the closest runner up rather than between two losers (Hill, 2004).

Table 5: IRV election where the IRV margin is smaller than the Condorcet margin.

Candidate	Round 1	Final
Xavier	6: X Y Z	6: X Y ~~Z~~
	5: X Z Y	5: X ~~Z~~ Y
		10: ~~Z~~ X Y
Yolanda	10: Y X Z	10: Y X ~~Z~~
	5: Y Z X	5: Y ~~Z~~ X
Zoë	10: Z X Y	—

might say that Velma won with a margin greater than 10% (either 11.7% or about 13.4% depending on whether the denominator is 1000 or $379 + 496 = 876$).

If an adversary is able to arrange for a single Y X V ballot to be counted as a Z Y ballot, then we get the election in Table 4. Here, the small error cascades through the rest of the rounds and Ulric, who previously came in second, is the winner with 379 votes. The correct winner, Velma, does not even make it to the final round. Instead, Xavier, who was previously eliminated in the second round makes it all the way to the final round to lose with 295 votes. Again, naïvely, the margin appears to be quite large (either 8.4% or about 12.5%). This example shows that intuition about margin calculations in plurality elections may not be applicable to IRV elections.

B.2 Margins for Condorcet versus IRV

The margin for IRV may be smaller than the Condorcet margin. Consider an election between Xavier, Yolanda, and Zoë. Only 36 ballots were cast in this election, and the results are summarized in Table 5.

Under IRV, in the first round Xavier gets 11 votes, Yolanda 15 votes, and Zoë 10 votes, so Zoë is eliminated. However, the supporters of Zoë break unanimously for Xavier over Yolanda, so in the final round Xavier defeats Yolanda 21 votes to 15 and Xavier is the IRV winner. The simple lower bound for the margin of this election is one vote, the gap between Xavier and Zoë in the first round. Note that Xavier is also the Condorcet winner of this election — voters prefer Xavier to both Yolanda and Zoë by 21 to 15. The Condorcet margin is therefore six votes. Further, voters also prefer Yolanda to Zoë 21 to 15 so the minimum difference in preference between candidates is also six. However, the IRV margin really is two since one ballot shifted from Xavier to Zoë will cause Xavier to be eliminated in the first round and Yolanda to win.

References

Arrow, K. J. (1951): *Social Choice and Individual Values*, Yale University Press, second edition.

Aslam, J. A., R. A. Popa, and R. L. Rivest (2007): "On estimating the size and confidence of a statistical audit," in R. Martinez and D. Wagner, eds., *Proceedings of EVT 2007*, USENIX and ACCURATE, 8–8.

Aslam, J. A., R. A. Popa, and R. L. Rivest (2008): "On auditing elections when precincts have different sizes," in D. Dill and T. Kohno, eds., *Proceedings of EVT 2008*, USENIX and ACCURATE, 1–1.

Australian Electoral Commission (2011): "Counting the votes – house of representatives," URL `http://www.aec.gov.au/Voting/counting/vid_hor.htm`.

Baldwin, J. M. (1926): "The technique of the Nanson preferential majority system of election," *Proceedings of the Royal Society of Victoria*, n.s. 39, 42–52.

Bartholdi III, J. J. and J. B. Orlin (1991): "Single transferable vote resists strategic voting," *Social Choice and Welfare*, 8, 341–354.

Bartholdi III, J. J., C. A. Tovey, and M. A. Trick (1989): "The computational difficulty of manipulating an election," *Social Choice and Welfare*, 6, 227–241.

Bartholdi III, J. J., C. A. Tovey, and M. A. Trick (1992): "How hard is it to control an election," *Mathematical Computer Modelling*, 16, 27–40, URL `http://dx.doi.org/10.1016/0895-7177(92)90085-Y`.

Benaloh, J., D. Jones, E. L. Lazarus, M. Lindeman, and P. B. Stark (2011): "SOBA: Secrecy-preserving observable ballot-level audit," in H. Shacham and V. Teague, eds., *Proceedings of EVT/WOTE 2011*, USENIX, ACCURATE, and IAVoSS.

Betzler, N., R. Niedermeier, and G. J. Woeginger (2011): "Unweighted coalitional manipulation under the Borda rule is NP-hard," in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, 55–60, URL `http://ijcai.org/papers11/Papers/IJCAI11-021.pdf`.

Black, D. (1958): *The Theory of Committees and Elections*, Cambridge University Press.

Brams, S. J. and P. C. Fishburn (1983): *Approval Voting*, Cambridge, MA: Birkhäuser Boston.

Brams, S. J. and P. C. Fishburn (2005): "Going from theory to practice: the mixed success of approval voting," *Social Choice and Welfare*, 25, 457–474, URL `http://dx.doi.org/10.1007/s00355-005-0013-y`.

Brelsford, E., P. Faliszewski, E. Hemaspaandra, H. Schnoor, and I. Schnoor (2008): "Approximability of manipulating elections," in D. Fox and C. P. Gomes, eds., *AAAI*, AAAI Press, 44–49.

Calandrino, J. A., J. A. Halderman, and E. W. Felten (2007): "Machine-assisted

election auditing," in R. Martinez and D. Wagner, eds., *Proceedings of EVT 2007*, USENIX and ACCURATE.

Cary, D. (2011): "Estimating the margin of victory for instant-runoff voting," in H. Shacham and V. Teague, eds., *Proceedings of EVT/WOTE 2011*, USENIX, ACCURATE, and IAVoSS.

Checkoway, S., A. Sarwate, and H. Shacham (2010): "Single-ballot risk-limiting audits using convex optimization," in D. Jones, J.-J. Quisquater, and E. Rescorla, eds., *Proceedings of EVT/WOTE 2010*, USENIX, ACCURATE, and IAVoSS.

Conitzer, V., T. Sandholm, and J. Lang (2007): "When are elections with few candidates hard to manipulate?" *Journal of the ACM*, 54, URL http://dx.doi.org/10.1145/1236457.1236461.

Consortium for Elections and Political Process Strengthening (2011): "ElectionGuide," URL http://www.electionguide.org/.

Davies, J., G. Katsirelos, N. Narodytska, and T. Walsh (2011): "Complexity of and algorithms for Borda manipulation," in *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*.

Dopp, K. (2008): "History of confidence election auditing development (1975 to 2008) & overview of election auditing fundamentals," URL http://electionmathematics.org/ucvAnalysis/US/paper-audits/History-of-Election-Auditing-Development.pdf, online.

Doron, G. and R. Kronick (1977): "Single transferable vote: An example of a perverse social choice functioned," *American Journal of Political Science*, 21, 303–311, URL http://www.jstor.org/stable/2110496.

Dummett, M. (1984): *Voting Procedures*, Clarendon: Oxford.

Faliszewski, P., E. Hemaspaandra, and L. A. Hemaspaandra (2010): "Using complexity to protect elections," *Communications of the ACM*, 53, 74–82, URL http://dx.doi.org/10.1145/1839676.1839696.

Faliszewski, P., E. Hemaspaandra, and L. A. Hemaspaandra (2011): "Multimode control attacks on elections," *Journal of Artificial Intelligence Research*, 40, 305–351, URL http://dx.doi.org/10.1613/jair.3136.

Farvaque, E., H. Jayet, and L. Ragot (2009): "A "winner" under any voting rule ? an experiment on the single transferable vote," Documents de travail du Centre d'Economie de la Sorbonne 09067, Université Panthéon-Sorbonne (Paris 1), Centre d'Economie de la Sorbonne.

Fishburn, P. C. (1977): "Condorcet social choice functions," *SIAM Journal on Applied Mathematics*, 33, 469–489, URL http://dx.doi.org/10.1137/0133030.

Friedgut, E., G. Kalai, and N. Nisan (2008): "Elections can be manipulated often," in *FOCS*, IEEE Computer Society, 243–249, URL http://dx.doi.org/10.

1109/FOCS.2008.87.

Gardenfors, P. (1976): "Manipulation of social choice functions," *Journal of Economic Theory*, 13, 217–228.

Gibbard, A. (1973): "Manipulation of voting schemes: A general result," *Econometrica*, 41, 587–601.

Higgins, M. J., R. L. Rivest, and P. B. Stark (2011): "Sharper p-values for stratified election audits," *Statistics, Politics, and Policy*, 2, URL http://dx.doi.org/10.2202/2151-7509.1031.

Hill, I. D. (2004): "An odd feature in a real STV election," *Voting Matters*, 18, 9.

Isaksson, M., G. Kindler, and E. Mossel (2010): "The geometry of manipulation: A quantitative proof of the gibbard-satterthwaite theorem," in *FOCS*, IEEE Computer Society, 319–328, URL http://dx.doi.org/.

Johnson, K. C. (2004): "Election certification by statistical audit of voter-verified paper ballots," URL http://dx.doi.org/10.2139/ssrn.640943, sSRN Working Paper.

Kaplan, H. M. (1987): "A method of one-sided nonparametric inference for the mean of a nonnegative population," *The American Statistician*, 41, 157–158, URL http://dx.doi.org/10.1080/00031305.1987.10475470.

Lindeman, M. and P. B. Stark (2012): "A gentle introduction to risk-limiting audits," *IEEE Security and Privacy*, URL http://dx.doi.org/10.1109/MSP.2012.56.

Magrino, T. R., R. L. Rivest, E. Shen, and D. Wagner (2011): "Computing the margin of victory in IRV elections," in H. Shacham and V. Teague, eds., *Proceedings of EVT/WOTE 2011*, USENIX, ACCURATE, and IAVoSS.

Nanson, E. J. (1882): "Methods of election," *Transactions and Proceedings of the Royal Society of Victoria*, 18, 197–240.

Neff, C. A. (2003): "Election confidence: A comparison of methodologies and their relative effectiveness at achieving it (revision 6)," URL www.verifiedvoting.org/downloads/20031217.neff.electionconfidence.pdf.

O'Neill, J. (2006): "Fast algorithms for counting ranked ballots," *Voting Matters*, 21, 1–5.

Poundstone, W. (2008): *Gaming the Vote: Why Elections Aren't Fair (and What We Can Do about It)*, Hill and Wang.

Procaccia, A. D. and J. S. Rosenschein (2007): "Junta distributions and the average-case complexity of manipulating elections." *Journal of Artificial Intelligence Research*, 28, 157–181, URL http://dx.doi.org/10.1613/jair.2148.

Reilly, B. (2002): "Social choice in the south seas: Electoral innovation and the Borda count in the pacific island countries," *International Political Science Review*, 23, 355–72.

Roberts, S. (2010): "Minor parties see threat in ballot quirk," *The New York Times*, October 26, A24.

Saari, D. (1995): *Basic Geometry of Voting*, Berlin: Springer.

Saari, D. G. (2001): *Decisions And Elections: Explaining The Unexpected*, Cambridge University Press.

Saltman, R. G. (1975): "Effective use of computing technology in vote-tallying," Technical Report Tech. Rep. NBSIR 75–687, National Bureau of Standards (Information Technology Division), Washington, D.C., USA, URL http://csrc.nist.gov/publications/nistpubs/NBS_SP_500-30.pdf.

San Francisco Voting Systems Task Force (2011): "Recommendations on voting systems for the city and county of San Francisco (draft)," URL http://www.sfgov2.org/Modules/ShowDocument.aspx?documentid=155.

Satterthwaite, M. A. (1975): "Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting and social welfare functions," *Journal of Economic Theory*, 10, 187–217.

Schulze, M. (1997): "Condorect sub-cycle rule," The Election-Methods mailing list, URL http://lists.electorama.com/pipermail/election-methods-electorama.com/1997-October/001570.html.

Schulze, M. (2011): "A new monotonic, clone-independent, reversal symmetric, and condorcet-consistent single-winner election method," *Social Choice and Welfare*, 36, 267–303, URL http://dx.doi.org/10.1007/s00355-010-0475-4.

Stanislevic, H. (2006): "Random auditing of e-voting systems: How much is enough?" URL http://www.votetrustusa.org/pdfs/VTTF/EVEPAuditing.pdf.

Stark, P. B. (2008a): "Conservative statistical post-election audits," *Annals of Applied Statistics*, 2, 550–581, URL http://dx.doi.org/10.1214/08-AOAS161.

Stark, P. B. (2008b): "Election audits by sampling with probability proportional to an error bound: dealing with discrepancies," URL http://www.stat.berkeley.edu/~stark/Preprints/ppebwrwd08.pdf, draft.

Stark, P. B. (2008c): "A sharper discrepancy measure for post-election audits," *Annals of Applied Statistics*, 2, 982–985, URL http://dx.doi.org/10.1214/08-AOAS171.

Stark, P. B. (2009a): "Auditing a collection of races simultaneously," URL http://arxiv.org/abs/0905.1422v1, arXiV Preprint arXiv:0905.1422v1 [stat.AP].

Stark, P. B. (2009b): "CAST: Canvass audits by sampling and testing," *IEEE Transactions on Information Forensics and Security*, 4, 708–717, URL http://dx.doi.org/10.1109/TIFS.2009.2034210.

Stark, P. B. (2009c): "Efficient post-election audits of multiple contests: 2009 California tests," in *Proceedings of the 4th Annual Conference on Empirical Legal Studies*, URL http://dx.doi.org/10.2139/ssrn.1443314.

Stark, P. B. (2009d): "Risk-limiting postelection audits: Conservative *P*-values from common probability inequalities," *IEEE Transactions on Information Forensics and Security*, 4, 1005–1014, URL http://dx.doi.org/10.1109/TIFS.2009.2034190.

Stark, P. B. (2010): "Super-simple simultaneous single-ballot risk-limiting audits," in D. Jones, J.-J. Quisquater, and E. Rescorla, eds., *Proceedings of EVT/WOTE 2010*, USENIX, ACCURATE, and IAVoSS.

Sturton, C., E. Rescorla, and D. Wagner (2009): "Weight, weight, don't tell me: Using scales to select ballots for auditing," in D. Jefferson, J. L. Hall, and T. Moran, eds., *Proceedings of EVT/WOTE 2009*, USENIX, ACCURATE, and IAVoSS.

Szpiro, G. (2010): *Numbers Rule : The Vexing Mathematics Of Democracy, From Plato To The Present*, Princeton University Press.

The Heisman Memorial Trophy (2011): "Heisman Trophy Balloting," URL http://www.heisman.com/history/balloting.php.

Tideman, N. (1995): "The single transferable vote," *The Journal of Economic Perspectives*, 9, 27–38, URL http://dx.doi.org/10.1257/jep.9.1.27.

Tideman, T. N. (1987): "Independence of clones as a criterion for voting rules," *Social Choice and Welfare*, 4, 185–206.

Woodall, D. R. (1997): "Monotonicity of single-seat preferential election rules," *Discrete Applied Mathematics*, 77, 81–98, URL http://dx.doi.org/10.1016/S0166-218X(96)00100-X.