Methods of Post-election Confidence-level Auditing

Stephen Checkoway
Department of Computer Science & Engineering
University of California, San Diego
La Jolla, CA, USA
s@cs.ucsd.edu

Abstract

We give a self-contained presentation and evaluation of a number post-election audit procedures. Methods which do not provide a statistical bound on the chance of being wrong give little confidence in the reported outcome of an election.

1 Introduction

With the introduction of electronic voting machines, a new class of security concerns has arisen. Namely, how can we ensure that a voter's intent is accurately reflected by the ballot cast when the choices are stored electronically in aggregate as a collection of counters. This is the case with Direct Recording Electronic (DRE) voting machines currently in use.

Trust in the electoral process is an essential component of any democracy. One way to build trust is to have open and transparent elections. Using computers in elections is an excellent way to increase efficiency, but it comes at the cost of transparency. How can we be sure that the votes we cast are the votes recorded? One way is to have redundant records which are compared to ensure that every vote is counted as intended. Unfortunately, this is prohibitively expensive and in many cases, unnecessary. Instead, we relax the constraint that every vote be recorded correctly and give a statistical guarantee that the *outcome* of the election matches the voters' will, rather than each individual's ballot. We do this by keeping a paper record of each vote in addition to an electronic tally and performing *audits*—typically hand recounts—of individual precincts to detect miscounts. If we find no miscounts in a sufficiently large sample, we confirm the election.

1.1 Election process

The election process consists of five distinct phases: setup, voting, reporting, tabulation, and auditing [HRSW08].

Setup. During the period between elections, voting machines are stored away from polling places under the control of either the local government or the voting machine vendors. Shortly before an election, during the setup phase, the voting machines are moved to the polling places by election workers and configured for the upcoming election. At this time, hardware and software checks of the machines take place, such as printing of the zero tape[1] and ensuring that the machine works in a simulated election. After setup is complete, the voting machine is ready to begin counting votes.

Voting. On election day, voters queue at their assigned polling place and sign in. Once they sign in, they are given either a ballot or some token to use with electronic voting machines. The voter then makes her choices and *casts* her ballot—either depositing a physical ballot in a collection box or confirming her selection on a review screen. If the voter had received a token, she returns it to the election workers who prepare it for the next voter.

Reporting. After all votes have been cast, the election workers close the polls, collect the ballots or electronic voting machine totals and send the results to a central location in the county. In the case of DRE or precinct-count opscan machines, the results are transmitted to the county using either a memory card or electronically over a network. At this time, an early report of the vote totals can be reported to the candidates and the press.

Tabulation. In the tabulation phase, an official count of each race is prepared by election officials by counting centrally-counted, absentee, and provisional ballots as well as write-ins and combining with the precinct-counted ballots [HRSW08]. Official winners of each race are determined at this time.

Auditing. After electronic totals have been produced, the results are audited in an attempt to check election accuracy. Each state has its own laws regarding election auditing. Most require a flat-rate or tiered percentage based on how close the race was. For example, in

[1] The zero tape is the record of how many votes have been cast so far. Since none have been cast, the counters should all be zero or something is amiss.

California, there is a mandatory 1% manual recount requirement.

At the conclusion of the five election phases, we should know that the outcome of the election accurately reflects the will of the people with high probability.

The current trend in voting for the past several years has been a move from the more traditional paper ballots or lever-machines to electronic voting machines. The current implementation of electronic voting machines and in particular DREs is flawed. We briefly discuss a few of the more severe flaws in Section 2. In order to be able to perform post-election audits of voting machines, a paper record of the votes cast is required. A few of the challenges associated with the voter-verified paper trail emitted by DREs is discussed in Section 3. At the conclusion of the tabulation phase of an election, it is time to verify the results so that the election can be certified. There are a number of post-election audit methods that one can use. A number of methods for selecting which precincts to audit and how to perform the audit itself are the subjects of Sections 4 and 5. We conclude and discuss possible directions for future research in Section 6

2 Voting machine security

In addition to equipment malfunction and procedural deficiencies which can lead to inaccurate vote totals, the current generation of electronic voting machines contain serious security flaws. A brief list of just a few of the more serious flaws with the Premier[2] AccuVote-TS compiled from [KSRW04, FHF07, CFH+07] is given below. A complete discussion of the myriad of flaws is outside the scope of this report, but see California's "Top-To-Bottom" review [Bow07] for a more comprehensive report on the issues plaguing the electronic voting machines used in California.

The AccuVote-TS is a DRE voting machine. During the setup phase, poll workers prepare the machines for the election by inserting a memory card which contains a file providing ballot definitions for use in the election. Voters are given smartcards as tokens that allow a vote to be cast. A printer can be attached to the voting machine to provide a voter-verified paper audit trail (VVPAT) which is used in post-election audits. Voting ends when poll workers insert an *ender* card or an administrator card and enter a password. A memory card is then used to collect the vote totals and is sent to the county central office for county-wide tallying by the *Election Management System* (EMS)—a general purpose computer running a standard version of Microsoft Windows.

- The most serious issue is the potential for a voting machine virus to spread from machine to machine. This can be accomplished by infecting the memory cards used to provide the ballot definition files used in the setup phase as well as the memory cards used for reporting precinct vote totals. These memory cards can then infect the EMS. Once the EMS is compromised, all election tallies in the current and future elections are suspect. The attack works because the TS will overwrite its firmware with the contents of a file with a particular name on a memory card present at boot-time.

- During the voting phase, a voter is given a smartcard which authorizes the voter to cast one vote. After the vote is cast, the data is overwritten on the smartcard so that no more votes may be cast with that card until it is reset by election workers. A malicious voter who knows the simple layout of the card can vote multiple times by changing the smartcard such that it does not overwrite its data. Alternatively, multiple smartcards could be prepared in advance to allow voting multiple times. This can be detected by election officials during the reporting phase by ballot reconciliation where the number of ballots cast is compared to the number of voters who signed in—assuming this check is performed.

- Similar to the previous issue, a voter can construct an ender card or an administrator card to prevent more votes from being cast. For example, in a precinct known to favor candidate A, a group of malicious voters who favor candidate B could disrupt voting in the precinct by inserting ender cards into the machines.

- Vote-stealing software can be installed alongside the election software by an insider or by the firmware exploit described above. The vote-stealing software can undetectably move votes from one candidate to another and can be written in such a way that this only happens during the election phase and not during the testing done in the setup phase.

- The VVPAT is subject to several attacks. Assume the attacker prefers candidate A over candidate B. When a voter votes for A, the software behaves correctly casting a ballot for voter A. When a voter votes for B, the software prints CANCELED and quickly scrolls the spoiled ballot and prints a new one for candidate A which also scrolls quickly past the window. The voter may not even notice that this has happened.

  A second attack involves switching the provisional status of ballots. If the provisional voter votes for B, the software behaves as normal and casts a provisional ballot for B. If she instead voted for A, the ballot is cast as nonprovisional. At a later time, when a normal voter votes for B, his ballot is marked as provisional. If the first voter—the one who voted for A—is found to be ineligible, a vote will be removed from B's count instead.

[2] Premier Election Solutions, which was formerly Diebold Election Systems, Inc.

- Since the VVPAT is printed in a reel-to-reel fashion, if the order of voters is known, the secrecy of the ballots can be compromised by an insider.

The risks associated with these issues can be somewhat mitigated by following the recommendations in [HRSW08]. In any case, it is clear that measures beyond those implemented by voting machine vendors are required to ensure the accuracy of our elections.

3 Paper trails

The need for a voter-verifiable paper audit trail (VVPAT) is nearly universally recognized by election researchers and election officials. A VVPAT serves two essential functions. One function is to provide a way for the voter to verify her selection before casting her ballot. Equally important is that the VVPAT serves as the ballot of record in the event of an audit. Currently, thirty-two states have legislation or regulations requiring a voter-verified, paper ballot of record while an additional eight have had legislation proposed but not yet enacted [Kib08].[3]

With a traditional paper ballot, punch card ballot, or optical scan (opscan) ballot, the ballot itself serves as a VVPAT. DRE machines can be equipped with a printer—usually a thermal printer akin to that used to print receipts—to print a VVPAT. Other voting technology such as lever machines and DREs without a printer offer no way to provide a VVPAT. As a result, there is no way for the voter to verify that her choices were recorded as intended and an audit is limited to summing the counts for each voting machine in a precinct and comparing to the previously reported total.

In theory, VVPATs are an excellent way to both verify selections and provide a method of auditing. The reality is quite a bit different. As Section 2 shows, the VVPAT is not at all immune to fraud. Studies show that when using a DRE with a VVPAT, most voters do not know about or do not check the VVPAT before casting the ballot. Everett's studies [Eve07] show that over 60% of voters do not notice that their selections have been changed as reported by a review screen—even if the changes are so extensive as to add or remove entire races. If a voter does not notice a changed selection on the screen used to make the selection in the first place, it seems even less likely that she would notice a change in the VVPAT which are typically off to the side or even completely covered [FHF07].

Even if the voter verifies that her selection is correct, the VVPATs in use by DREs are plagued by problems. The Election Science Institute's report on Cuyahoga County, Ohio discusses a number of issues. They report that 9.7% of the VVPAT ballots were "either destroyed, blank, illegible, missing, taped together, or compromised in some way." In addition, 1.4% of the VVPATs were missing ballots [ESI06]. Worse still, studies have shown that auditing the ballots is an error prone task with 40%–60% of the participants in the studies giving an incorrect count of ballots [GB07, GBG+08].

Despite all of the flaws in VVPATs, without them, there is very little that can be done from an auditing standpoint to provide statistical guarantees of correctness [HRSW08]. As such, the paper record produced by an electronic voting machine is our main object of study. In Stark's words, "Hand counts are subject to error, but they are the gold standard" [Sta08a].

4 Confidence level election auditing

The goal of confidence level election auditing is to provide a mathematically sound upper bound on the probability that the reported outcome of an election does *not* represent voters intent. Appel [App07] suggests that it is important that the losing candidate have confidence that the reported outcome is correct and the mere presence of the auditing process should deter fraud by providing a significant chance of detection.

In the electoral process, there are many potential sources of error. The voter can vote for fewer candidates than allowed—called an undervote—or vote for too many and thus spoil her ballot—called an overvote. A software error could cause a vote for candidate A to be counted for candidate B, count votes that do not exist, not count votes that were cast, or a whole host of other errors. In addition to simple errors, human or otherwise, there is the possibility of fraud. As shown in Section 2, fraud is a very real possibility with current electronic voting machines. As fraud is the most difficult sort of error to detect, it is our main focus. What to do when evidence of fraud is detected is a political or criminal matter and as such is far outside the scope of this report.

Election auditing is typically studied from the perspective of hypothesis testing. The *null hypothesis* H_0 which we wish to reject or nullify is that the outcome is wrong—i.e., one or more of the presumptive winners are losers and vice-versa. The probability of the error of rejecting H_0 when it is true—a type I error—is the significance level α of the test used to make that determination. In other words, if a significance level α test confirms the election (rejects H_0), then either the outcome is correct or an event with probability at most α has occurred. The other possible error, accepting H_0 when it is false—a type II error—is less severe since while it should lead to further audits which take time and cost money, it cannot lead to confirming a losing candidate as a winner. The power of a test is the probabil-

[3] Verified Voting's web site lists Arkansas as having mixed requirements; however, the text of AR Code §7-5-504 was amended by H.B.360 to read, "If the machine is a direct read electronic voting machine, it shall include a voter verified paper audit trail as provided under §7-5-532," where §7-5-532 states that the VVPAT is the ballot of record in the case of a recount.

ity of correctly rejecting the null hypothesis. The p-value[4] is the smallest α such that our test rejects H_0. For a more complete discussion of hypothesis testing see [Was04].

Note that this is not the only choice of H_0. We could have chosen to make H_0 be that the outcome is correct. In that case, a type I error would be incorrectly rejecting H_0 and thus erroneously deciding that the outcome was incorrect. This error would lead to further investigation which should eventually discover that the outcome was correct—perhaps by a complete hand count. However, a level α test does not bound the probability of the more serious error of incorrectly accepting H_0—that is, incorrectly confirming the election. This is our primary concern and thus we want a level α test for the null hypothesis that the outcome is incorrect.

Define the *confidence level* of an election audit to be $c = 1 - \alpha$ where α is the significance level.[5] For a given confidence level c—e.g., $c = 95\%$ or $c = 99\%$—the question is which precincts must be audited after an election to be sure that the reported results are accurate with a confidence of c.

It is well-known that ballot-based auditing methods are more efficient than precinct-based auditing in terms of number of ballots audited [Wan04, CHF07, Dop08]. Unfortunately, the electronic voting machines currently in use do not provide electronic ballots to audit. For this reason we focus on precinct-based auditing, but see Subsection 4.12 for a brief discussion of ballot-based auditing.

Our goal is to give a statistical bound on the chance of an error occurring. To do so, we will be sampling from a probability distribution and so we need a source of randomness to generate random numbers. There are a number of options. Cordero, Wagner, and Dill [CWD06] recommend using translucent, 10-sided dice. Election officials would first publish the probability distribution from which they are going to sample. The dice would then be rolled and video taped in the presence of observers. Sampling can be performed in a simple manner. When given a probability distribution with probabilities $P_1, P_2, \ldots, P_k$, we can form the cumulative distribution $F_0, F_1, \ldots, F_k$ where $F_i = \sum_{j=1}^{i} P_j$—where $F_0 = 0$—and write the values in a table. We can sample from this distribution by rolling a die q times where the ith roll is the ith decimal digit d_i of the number $x = 0.d_1 d_2 \ldots d_q$. If $F_{j-1} \le x < F_j$, then we have sampled the jth event. The number of rolls q depends on the distribution.

A pseudo-random number generator (PRNG) is a common, easily accessible option. Knuth [Knu97] shows the dangers of implementing one's own PRNG; however, there are a number of good choices available. The use of a cryptographically secure PRNG—where the seed is chosen by combining random numbers produced by a number of different people using dice—is recommended by Calandrino, Halderman, and Felten [CHF07, CHF08]. A PRNG that is not suitable for use with cryptography should not be used for elections either.

[4] We are overloading p as both p-value and a precinct p. Which one we mean should be clear from context.

[5] Confidence as used in the election literature differs from its usage in statistics.

4.1 Notation

For the purposes of auditing, we focus on a single race or ballot issue at a time. Following the notation of Stark [Sta08a], there are N precincts and K candidates. Each voter can vote for up to f candidates. Let $v_{k,p}$ be the reported number of votes candidate k received in precinct p and $a_{k,p}$ be the actual number of votes candidate k received in precinct p as reported by an audit of p. Let $V_k = \sum_{p=1}^{N} v_{k,p}$ be the total number of reported votes for candidate k and $A_k = \sum_{p=1}^{N} a_{k,p}$ be the actual total number of votes for candidate k. The set of indices of the f winners is denoted K_w while $K_l = \{1, 2, \ldots, K\} \setminus K_w$ is the set of indices of the losing candidates. Define the margin M to be the difference in reported votes between the lowest performing candidate in K_w and the highest performing candidate in K_l:

$$M = \min_{k \in K_w} V_k - \max_{k \in K_l} V_k. \tag{4-1}$$

Define the potential margin overstatement discrepancy—think *error*—in precinct p as

$$e_p = \sum_{k \in K_w} \max\{v_{k,p} - a_{k,p}, 0\} + \sum_{k \in K_l} \max\{a_{k,p} - v_{k,p}, 0\}. \tag{4-2}$$

This is the maximum amount by which fraud in precinct p could change the margin. In addition, let u_p be an a priori upper bound on e_p. Let $E = \sum_{p=1}^{N} e_p$ be the total discrepancy. Define b_p to be the number of voting opportunities in precinct p calculated as f times the number of ballots, including undervoted and invalid ballots.

Let n denote the number of precincts to audit which will typically be a function of several parameters. The number of *bad* precincts will be denoted by b—i.e., the number of precincts where fraud has occurred. If there are N precincts, b of which are bad, then with a uniform sample without replacement of size n, the chance that all b bad precincts *escape* detection is the hypergeometric distribution

$$\begin{aligned} e(N, b, n) &= \frac{\binom{b}{0}\binom{N-b}{n}}{\binom{N}{n}} \\ &= \prod_{k=0}^{n-1} \frac{N - b - k}{N - k}, \end{aligned} \tag{4-3}$$

while the chance of *detection* is $d(N, b, n) = 1 - e(N, b, n)$.

4.2 100% confidence

Even if a confidence level of $c = 100\%$ is required, we need not always audit every precinct [MSL+07, Dop08]. After auditing some number of precincts, if the difference between the actual lowest performing candidate so far and

the actual highest performing candidate so far is greater than the number of ballots left to audit, then the auditing can stop. At this point, the true results are known without 100% auditing and the election can be certified.

4.3 Fixed percentage audits

One simple idea is to pick a fixed percentage of precincts to audit uniformly at random. For example, California mandates a fixed 1% audit for every election.[6] This has the two advantages of both being simple to understand and every ballot has the same chance of being audited—the lack of either could lead to voter disenfranchisement.

The problem with this simple strategy is that it provides no guarantees on the probability that vote fraud will be discovered. Consider an election with $N-1$ small precincts and one large precinct. If a fixed fraction F of the precincts are audited uniformly at random so $n = \lceil FN \rceil$, then the chance that the large precinct is audited is

$$d(N, 1, n) = 1 - \frac{\binom{N-1}{n}}{\binom{N}{n}} = \frac{\binom{N-1}{n-1}}{\binom{N}{n}}, \tag{4-4}$$

where the ceiling $n = \lceil FN \rceil$ is to sample an integral number of precincts that is at least the fraction F of the total number of precincts N. For example, if $N = 100$ and $F = 0.01$, then the chance that the large precinct is audited is $\binom{99}{0}/\binom{100}{1} = 1/100$. If the number of votes that can be switched from a winning candidate to a losing candidate in the large precinct without causing suspicion (see Subsection 4.4) is large enough to change the outcome of the election, then the election can be stolen by fraud in only the large precinct.

4.4 1%+7 auditing

One alternative to fixed percentage auditing is to allow a candidate or a candidate's party to select some number of precincts to audit in addition to a fixed percentage of precincts selected uniformly at random. Appel proposes a 1% fixed percentage audit and that candidates be allowed to select—and pay for—7 precincts to be audited [App07]. The argument is that in large elections, the 1% will be enough to detect fraud whereas in a smaller election, the 7 additional precincts will be sufficient.

This depends on an assumed maximum percentage of votes s that can be shifted from one candidate to another in a precinct. It is assumed that if more than s of the votes in a precinct are changed, it will be "obvious" and lead to further investigation. There is no mathematical foundation to this number but several different values have been proposed including 14% [App07], 15% [Sal75, DS06], 20% [Sta06, MSL+07, NBHC07], and 50% [Dop08]. This value has variously gone by the names "maximum level of undetectability by miscount" (MLU) [Sal75], "maximum vote shift" [DS06, NBHC07], and "within precinct miscount" (WPM) [Sta06, MSL+07].

[6] Actually, California mandates that at least 1% of the precincts in each county be audited at random plus at least one precinct in each contest missed by the sample [Sta08a].

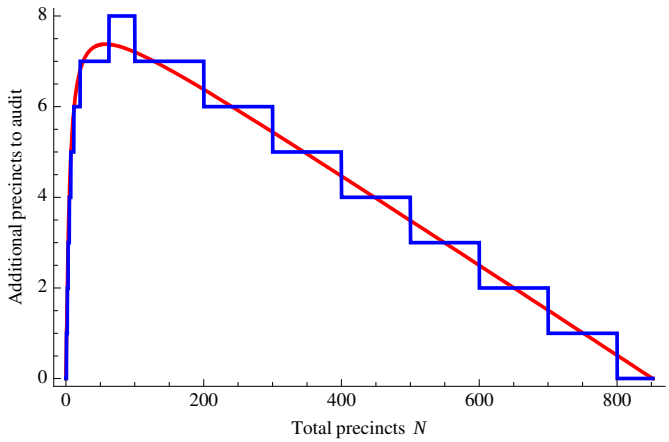

Figure 1: Plots of the number of additional precincts to audit on top of a 1% audit to ensure confidence level c. The plots are $y(N) = N\big(1 - \exp(s\log(1-c)/Nx)\big) - 0.01N$—which corresponds to auditing fractional precincts—and $y(N) = \lceil N\big(1 - \exp(s\log(1-c)/Nx)\big)\rceil - \lceil 0.01N \rceil$—which corresponds to auditing an integral number of precincts—for $0 \le N \le 852 \approx -s\log(1-c)/\big(x\log(100/99)\big)$ with $c = 0.95$, $s = 1/7$ and $x = 0.05$.

Assume that no more than a fraction s of the total votes in a precinct are changed and assume that each of the N precincts has a uniform size. If the total fraction of votes changed is x, then there are $b = Nx/s$ bad precincts. Appel uses Rivest's "Improved Rule of Three" [Riv06][7] for an upper bound on the number of precincts needing to be audited to detect fraud in b precincts with confidence c:

$$\begin{aligned} n(N, b, c) &\ge N\left(1 - \exp\left(\frac{\log(1-c)}{b}\right)\right) \\ &= N\left(1 - \exp\left(\frac{s\log(1-c)}{Nx}\right)\right). \end{aligned} \tag{4-5}$$

Setting $n(N, b, c) = 0.01N$ and solving for N we find that a 1% audit is sufficient for

$$N \ge -\frac{s\log(1-c)}{x\log(100/99)}. \tag{4-6}$$

Using the values $s = 1/7 \approx 14\%$, $x = 0.05$, and $c = 0.95$, we see that by auditing an additional 8 precincts on top of a 1% audit there is a 95% confidence of detecting a 5% fraud; see Figure 1. Note that there is only a small range of numbers of precincts for which 7 does not suffice. This range shrinks to about $60 \le N \le 100$ if $\lceil 0.01N \rceil$ is used as the number of precincts to audit for the mandatory 1%.

Unfortunately, the method of 1% + 7 suffers from several problems. The confidence bounds only apply if the 7

[7] This was proved to be an upper bound in [APR07] whereas [Riv06] merely gives numeric and heuristic arguments.

precincts are chosen uniformly at random—something unlikely to happen when chosen by a candidate. In addition, the precincts are assumed to have uniform size, something most counties do not have. Furthermore, low levels of vote shifting or miscount (below 2%) have a small chance of detection by this auditing strategy. In close elections, this is insufficient.

4.5 Auditing precincts of equal size

Subsection 4.4 demonstrates the need for statistical bounds on the number of precincts to audit. In order to guarantee a given confidence level c, we need several assumptions which we will relax in later subsections. As before, we assume N equal sized precincts and a maximum percentage s of votes that can be shifted from one candidate to another without being "obvious." The question is, how many precincts need to be audited (uniformly at random) to have confidence c.

Saltman, Dopp, and Stanislevic independently develop the basic framework given here for counting the minimum number of precincts to audit [Sal75, Dop06, Sta06]. For a given margin M, we can compute the minimum number of bad precincts b necessary to alter the outcome of the election based on the maximum vote shift s. Since shifting one vote changes the margin by up to 2, if $V = \sum_{k=1}^{K} V_k$ is the total number of votes, we have

$$b = \left\lfloor \frac{MN}{2sV} \right\rfloor. \tag{4-7}$$

Once we have b and our desired confidence level c, the question is the minimum value of n—the number of precincts to sample—such that $d(N, b, n) \geq c$, or equivalently, $e(N, b, n) \leq \alpha$ where α is the significance level.

In order to solve for n, Dopp and Stenger [DS06] give a numerical solution. Starting with $c \leq d(N, b, n)$, we expand to

$$\frac{(N-n)!}{(N-b-n)!} \leq \frac{(1-c)N!}{(N-b)!}, \tag{4-8}$$

take the natural log of both sides

$$\log(N-n)! - \log(N-b-n)! \leq \log(1-c) + \log N! - \log(N-b)! \tag{4-9}$$

By using $x! = \Gamma(x+1)$ where $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt$ and numeric implementations of $\log \Gamma(z)$, one can compute numeric solutions to Equation (4-9), and thus find the optimal number of precincts to audit n.

One drawback of this approach is that it requires a numeric calculation—for example, the MATLAB code given in [DS06]—that is not easy to perform on a hand calculator by election officials. We would like a conservative upper bound that is not too large. Rivest [Riv06] gives the "Improved Rule of Three" mentioned above which is proved correct in [APR07] and goes further. The remainder of the calculations in this subsection follow from the latter paper.

There is a nice duality between the number of bad precincts b and the number of sampled precincts n given by

$$e(N, b, n) = \frac{(N-b)!}{(N-b-n)!} \cdot \frac{(N-n)!}{N!} = e(N, n, b). \tag{4-10}$$

Combining Equations (4-3) and (4-10), we have a new way to compute the probability of escaping detection:

$$e(N, b, n) = \prod_{k=0}^{b-1} \left(1 - \frac{n}{N-k}\right). \tag{4-11}$$

For the purposes of auditing, we must have an integral number of precincts n, but Equation (4-11) is perfectly well-defined for arbitrary real values of n. This observation, combined with a very clever application of the arithmetic-geometric-harmonic mean inequalities (see Appendix A) gives the conservative upper bound

$$n(N, b, c) \geq \left(N - \frac{b-1}{2}\right) \cdot \left(1 - \exp\left(\frac{\log(1-c)}{b}\right)\right). \tag{4-12}$$

As usual, we need for $n(N, b, c)$ to be integral so we take $n(N, b, c)$ to be the ceiling of the right-hand side of Inequality (4-12). If we compute a lower bound on the optimal number of precincts to audit $n_{\mathrm{OPT}}(N, b, c)$, we see that

$$n(N, b, c) \leq n_{\mathrm{OPT}}(N, b, c) + \left\lceil -\frac{\log(1-c)}{2} \right\rceil + 1.^{8} \tag{4-13}$$

See Appendix B for the calculations of these bounds, including the upper bound in Equation (4-5).

The major advantage of the method in this subsection is that sampling $n(N, b, c)$ precincts uniformly at random is guaranteed to find at least one of the b bad precincts with probability c. The assumption that all precincts have equal size has the effect of maximizing the minimum number of bad precincts b for a given maximum vote shift s. This works well in states like New Jersey which have roughly uniform precinct sizes [App07], but is unlikely to be true in general. If individual ballots are audited instead of precincts—something not generally possible with current DREs—the total number of ballots cast must be used instead of total numbers of votes in order to handle overvotes and undervotes [Dop08].

4.6 SAFE auditing

The SAFE auditing method of [MSL+07] is a slight extension to the methods of the previous subsection by allowing variable sized precincts.

We can relax the equal sized precincts restriction to assuming that the average number of votes in precincts with fraud is the same as the average number of votes in all

[8] The calculation in [APR07] does not contain the $+1$. In that paper, the values $n(N, b, c)$ and $n_{\mathrm{OPT}}(N, b, c)$ are not constrained to be integers while computing the bound on $n(N, b, c) - n_{\mathrm{OPT}}(N, b, c)$, leading to the off by one.

precincts. In this case, we compute the number of bad precincts b as in Equation (4-7) and use Inequality (4-12) to compute the number of precincts to sample uniformly at random [Sal75, MSL+07].

If precinct sizes are not equal or we suspect that fraud will not happen only in average-sized precincts, then we can adjust the number of bad precincts by assuming that fraud only happens in the largest. This can be calculated directly assuming a maximum vote shift s by using the number of ballots cast in each precinct or by a heuristic such as

$$\tilde{b} = \left\lceil \frac{b}{\log_{10}(N/b) + 1} \right\rceil, \tag{4-14}$$

where b is computed as in Equation (4-7) [MSL+07]. As before, this value is used to calculate the number of precincts to sample uniformly at random using Inequality (4-12).

Using the SAFE auditing method, we can audit precincts of unequal sizes. Overvotes and undervotes are handled correctly as long as the number of ballots cast in a precinct are used instead of the number of votes [Dop08].

4.7 Margin overstatements, overvotes, undervotes, and Stark's "pooling rule"

Recall that the potential margin overstatement discrepancy e_p in Equation (4-2) is the maximum amount by which error in precinct p could increase the margin M, and that the total discrepancy is $E = \sum_{p=1}^{N} e_p$. As long as $E < M$, the apparent set of winners K_w must be the actual winners; see Appendix C. Furthermore, if u_p is an upper bound on e_p, then if $\sum_{p=1}^{N} u_p < M$, then the apparent winners are the actual winners. See Subsection 4.9 for choosing the upper bounds.

Stark [Sta08a] describes a way to handle overvotes, undervotes, and "obviously" losing candidates in a unified manner. To handle overvotes and undervotes, a new *pseudo-candidate* is created and credited with all of the undervotes and f times the overvotes. Doing this handles the overvote and undervote in each precinct by simply increasing the maximum discrepancy e_p. Furthermore, obviously losing candidates together with overvotes and undervotes can be grouped together—pooled—into pseudo-candidates. Each pseudo-candidate's aggregate total may not exceed the total votes for the highest performing apparent loser. Overvotes are counted as f times the number of ballots.

As we will see in Subsection 4.9, it is preferable for the candidate with the fewest votes to have as many votes as possible. This suggests that we pool the candidates into some number of pseudo-candidates such that the pseudo-candidate with the fewest aggregate votes has as many as possible. In general, this pooling is NP-hard; however, for a small number of candidates, it should be tractable [Imp08].

As an example, consider a five candidate race with two winners—so $f = 2$. Suppose the reported votes were as in Table 1a. Since there are 2 winners, both A and B are winners while C is the runner up. The margin is the difference between B and C so $M = 400$. There are several ways to pool the candidates such that the margin does not change. For example, pooling D and E into a pseudo-candidate DE and pooling the undervotes, overvotes, and writeins into UOW makes the pseudo-candidate with the fewest votes be UOW with $20 + 2 \cdot 20 + 15 = 75$ votes—where we multiplied the overvotes by $f = 2$. If we instead pool D with the overvotes and E with the undervotes and writeins, then the pseudo-candidate with the fewest votes as many as possible, see Table 1b.

Table 1: Example election

(a) The vote tallies for the candidates.

Candidate	Votes
A	1000
B	900
C	500
D	250
E	250
undervotes	20
overvotes	20
writeins	15

(b) The vote tallies after pooling into pseudo-candidates.

Pseudo-candidate	Votes
A	1000
B	900
C	500
DO	290
EUW	285

This pooling approach has the advantage that votes shifted from one candidate to another when both are part of the same pseudo-candidate cannot affect the outcome of the election. Thus they can be ignored—at least for the purpose of certifying the election. Of course any evidence of fraud is worth further investigation!

4.8 Negative-exponential auditing method

The negative-exponential auditing method (NEGEXP) of Aslam, Popa, and Rivest [APR08] does away with the sampling uniformly at random approach; as do the remainder of the auditing methods discussed in this report. The NEGEXP method is mathematically quite elegant.

We have discussed above how more fraud can be "hidden" in a larger precinct—at least under the assumption that the maximum vote shift fraction s is constant. To handle this, the SAFE auditing method of Subsection 4.6 increased the number of precincts to sample uniformly at random to account for the smaller number of bad precincts that would need to be corrupted. NEGEXP handles this in a particularly pleasing manner.

Recall that u_p is an a priori upper bound on the potential margin overstatement discrepancy e_p. That is, it is an upper bound on the amount by which fraud in precinct p can change the margin M.

Given a set of precincts S, we would like the probability of auditing at least one of the precincts in S to depend only on the bound $\sum_{p \in S} u_p$. Thus, for any precincts p, q and r, if $u_p = u_q + u_r$, the probability of sampling precinct

p is the same as the probability of sampling precincts q or r. Rephrasing, the probability of not sampling precinct p should be the same as the probability of not sampling precincts q and r. For each precinct p, let P_p be the probability of auditing precinct p. The requirement is

$$(1 - P_p) = (1 - P_q)(1 - P_r). \tag{4-15}$$

Thus, we define $P_p = 1 - \exp\big(-u_p/w(c)\big)$ where $w(c)$ depends on our desired confidence level and will be determined shortly.

The major benefit to defining our precinct sampling probabilities in this manner is that given any set of precincts S with $\sum_{p\in S} u_p \geq M$, the chance of sampling at least one precinct in S is

$$1 - \prod_{p\in S} \exp\left(\frac{-u_p}{w(c)}\right) \geq 1 - \exp\left(\frac{-M}{w(c)}\right). \tag{4-16}$$

For a given confidence level c, we want the lower bound $1 - \prod_{p\in S}\big(-u_p/w(c)\big) \geq c$. Using Inequality (4-16), we set $w(c) = -M/\log(1-c)$, and thus precinct p is audited with probability

$$P_p = 1 - (1 - c)^{u_p/M}. \tag{4-17}$$

The NEGEXP method is simple to implement in a transparent manner. Once we have the bounds u_p, the probabilities P_p can easily be calculated and each precinct is audited according to that value. By employing Stark's pooling rule, NEGEXP handles undervotes and overvotes.

One downside is that since NEGEXP samples each precinct with probability independent of the others, the probability of sampling each precinct is higher than it would be if the precincts were not sampled independently. For example, assume each precinct has the same size and the error bounds are computed as a fraction s of the total votes in a precinct as in Saltman's method of Subsection 4.5. Then precinct p is audited by NEGEXP with probability $P_p^{\text{NEGEXP}} = 1 - (1 - c)^{1/b}$ and by Saltman's method with probability

$$P_p^{\text{Saltman}} = \frac{n(N, b, c)}{N} \leq \left(1 - \frac{b-1}{2N}\right)\left(1 - (1 - c)^{1/b}\right) + \frac{1}{N}, \tag{4-18}$$

where this inequality uses the bound in Inequality (4-12) with a $+1$ replacing the ceiling. As b increases relative to N, NEGEXP samples each precinct with probability increasingly higher than Saltman's method.

Another downside is a potential for voter disenfranchisement since votes do not have an equal probability of being audited and thus smaller precincts could be perceived as being less important.

The question remains: how should we choose the a priori bounds u_p? This question is relevant to the remainder of the auditing methods discussed in this report and thus is the subject of the next subsection.

4.9 Choosing the bounds

In the auditing methods discussed so far, we have assumed a bound proportional to the size of the precinct. That is, we have assumed that no more than a fraction s of the votes in a given precinct could be changed. Thus if precinct p has b_p voting opportunities—that is, f times the number of ballots returned, where each voter can vote for f candidates—then $u_p = 2sb_p$ where the worst case is that the fraction s of the b_p votes were changed from the apparent losers to the apparent winners.

We can derive a mathematically sound upper bound u_p for each e_p [Sta08a]. Let r_p be an upper bound on the actual total vote in precinct p. That is, $\sum_{k\in K} a_{k,p} \leq r_p$. This can be f times the number of ballots or f times the number of voters registered in a precinct, for example. No matter how r_p is determined, e_p is maximized if every one of the r_p votes is for the apparent loser with the fewest reported votes in precinct p:

$$e_p \leq r_p + \sum_{k\in K_w} v_{k,p} - \min_{k\in K_l} v_{k,p}. \tag{4-19}$$

The number on the right hand side can be used for the upper bound u_p.

Note that the right-hand side of Inequality (4-19) is maximized when there is a losing candidate with zero votes. Stark's pooling rule discussed in Subsection 4.7 attempts to maximize the number of votes for the pseudo-candidate with the fewest votes. This has the effect of decreasing the u_p and thus requires fewer precincts to be sampled.

4.10 Sampling with probability proportional to error bounds with replacement

The NEGEXP method is essentially optimal for sampling precincts independently in that no attacker's strategy for committing fraud performs any better than another's (assuming that the maximum amount of fraud is committed in each precinct). As noted above, when precincts have the same size, it is more efficient to sample using Saltman's method. This is because the precincts are not sampled independently and missing a precinct in one draw increases the chance of selecting it in subsequent draws. The method of auditing by sampling with probability proportional to error bounds with replacement (PPEBWR) of Aslam, Popa, and Rivest [APR08] and Stark [Sta08b] likewise does not sample precincts independently and can consequently perform better than NEGEXP.

The main idea is to construct a probability distribution $\mathbf{P} = (P_1, P_2, \ldots, P_N)$ and then sample n precincts (with replacement) where the probability of precinct p is P_p. The unique precincts sampled are then audited. If the sum of discrepancy bounds $\sum_{p=1}^{N} u_p$ is less than M, then the outcome must be correct—see Appendix C—so we assume that $M \leq \sum_{p=1}^{N} u_p$. Once we have this, all that remains is

the choose the number of sampling rounds n such that the confidence is at least c.

Following [APR08, Sta08b, Dop08], we construct the probability distribution $\mathbf{P} = (P_1, P_2, \ldots, P_N)$ where $P_p = u_p / \sum_{k=1}^{N} u_k$. For any subset $S \subset \{1, 2, \ldots, N\}$ of the precincts, the probability of sampling a precinct in S in one round is $\sum_{p \in S} P_p$. If $\sum_{p \in S} u_p \geq M$, then

$$\Pr[\text{sampling from } S] = \frac{\sum_{p \in S} u_p}{\sum_{p=1}^{N} u_p} \geq \frac{M}{\sum_{p=1}^{N} u_p}. \quad (4\text{-}20)$$

Therefore, the chance Π of not sampling from S in n rounds is bounded above by

$$\Pi \leq \left(1 - \frac{M}{\sum_{p=1}^{N} u_p}\right)^n. \quad (4\text{-}21)$$

Since we want $\Pi \leq 1 - c$, we set

$$n = \left\lceil \frac{\log(1-c)}{\log\left(1 - M / \sum_{p=1}^{N} u_p\right)} \right\rceil. \quad (4\text{-}22)$$

Using this choice of n, we can prove that for the same confidence c, the probability of sampling precinct p is smaller using PPEBWR than using NEGEXP; see Appendix D. Since PPEBWR does not perform a simple random selection of precincts, it can perform better than Saltman's method in the case of uniformly sized precincts. A given precinct p is audited with probability $P_p^{\text{PPEBWR}} = 1 - (1 - 1/N)^{\log_{1-b/N}(1-c)}$ using PPEBWR while p is audited with probability P_p^{Saltman} as given in Inequality (4-18). As Figure 2 shows, as the number of bad precincts relative to N increases, the number of precincts audited by PPEBWR relative to Saltman's method decreases. In addition to being more efficient, PPEBWR requires fewer random numbers, n instead of N, which are required for methods that audit each precinct independently at random. As with NEGEXP, using Stark's pooling rule allows PPEBWR to handle undervotes and overvotes

4.11 Sequential auditing with PPEBWR: dealing with discrepancies

All of the post-election auditing methods discussed so far have not addressed the question of what to do when discrepancies in the vote tallies are discovered. It is clear that further investigation is required, but what should the further investigation entail? By contrast, Stark proposes a sequential auditing method that either confirms the election outcome with confidence level c or an audit of every precinct is performed [Sta08a, Sta08b, Sta08c].

The auditing method works by performing some number of rounds of sampling. For each round r, a additional number of precincts $n_r - n_{r-1}$ are sampled with replacement according to a probability distribution $\mathbf{P} = (P_1, P_2, \ldots, P_N)$. A test statistic t_r is computed for each round as is a maximum p-value for t_r. If the p-value is less than the per-round

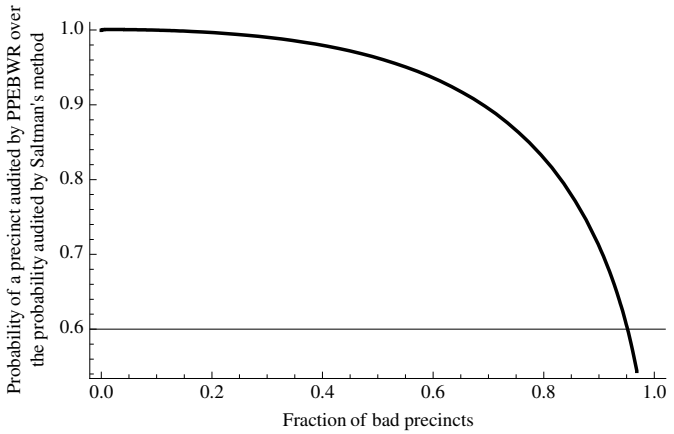

Figure 2: A plot of $P_p^{\text{PPEBWR}} / P_p^{\text{Saltman}}$ as a function of the fraction of total precincts which are bad with $N = 1000$ and $c = 0.95$.

significance level α_r, then the outcome is confirmed and the procedure stops. Otherwise a new round is begun.

For each precinct p, let w_p be a monotonically increasing function and $w_p^{-1}(t) = \sup\{q \in \mathbb{R} : w_p(q) \leq t\}$. The test statistic is the maximum of $w_p(e_p)$ for each p in our sample. That is, in round r, if we have sampled the precincts S_r, then

$$t_r = \max_{p \in S_r} w_p(e_p). \quad (4\text{-}23)$$

The functions w_p let us weight the discrepancies we discover independently in each precinct. For example, if $w_p(q) = q$, then we weight each discrepancy identically.

The p-value of round r the test is the probability that a sample of n_r precincts has $w_p(e_p) \leq t_r$ for each precinct p in the sample S_r given that $\sum_{p=1}^{N} e_p \geq M$. This depends on the probability distribution $\mathbf{P}$, which we will pick shortly.

For every real number t, define

$$T_t = \{p \in \{1, 2, \ldots, N\} : w_p(e_p) > t\}. \quad (4\text{-}24)$$

Note that for all $p \notin T_t$, $w_p^{-1}(t) \geq e_p$. Assuming that the null hypothesis holds—that is, the reported outcome is incorrect—it must be the case that $\sum_{p=1}^{N} e_p \geq M$. Then for any real t,

$$\begin{aligned}
\sum_{p \in T_t} \left(u_p - w_p^{-1}(t)\right) &= \sum_{p \in T_t} u_p + \sum_{p \notin T_t} w_p^{-1}(t) - \sum_{p=1}^{N} w_p^{-1}(t) \\
&\geq \sum_{p \in T_t} u_p + \sum_{p \notin T_t} e_p - \sum_{p=1}^{N} w_p^{-1}(t) \\
&\geq M - \sum_{p=1}^{N} w_p^{-1}(t),
\end{aligned} \quad (4\text{-}25)$$

where the first inequality comes from $w_p^{-1}(t) \geq e_p$ for $p \notin T_t$ and the second comes from $u_p \geq e_p$ and the sum of the e_p being at least M.

As an aside, using Inequality (4-25) and a given value t, we can bound the the number of draws n required to find an element in T_t with probability at least c assuming we sample each precinct p (with replacement) with probability

$$P_p = \frac{u_p - w_p^{-1}(t)}{\sum_{k=1}^{N}(u_k - w_k^{-1}(t))}. \tag{4-26}$$

We can easily check if a given p is in T_t by performing the audit and checking $w_p(e_p) > t$. If we find no such precincts, then we should reject the null hypothesis. This is an extension of the PPEBWR method of the previous section to account for discrepancy.

Instead, we want to calculate the p-value of round r. Since we do not know t_r until after we sample, we choose the w_p such that the P_p in Equation (4-26) does not depend on t. Let $w_p(q) = q/u_p$. Then $w_p^{-1}(t) = t \cdot u_p$ and thus,

$$P_p = \frac{(1-t)u_p}{\sum_{k=1}^{N}(1-t)u_k} = \frac{u_p}{\sum_{k=1}^{N} u_k}. \tag{4-27}$$

This probability is exactly the same as for PPEBWR. Now, the chance that a single sample from $\mathbf{P}$ will be in T_{t_r} is $\sum_{p\in T_{t_r}} P_p$. From Inequality (4-25), it is clear that $\sum_{p\in T_{t_r}} u_p \geq M - \sum_{p=1}^{N} w_p^{-1}(t_r)$. Therefore,

$$\sum_{p\in T_{t_r}} P_p \geq \frac{M - t\sum_{p=1}^{N} u_p}{\sum_{p=1}^{N} u_p} = \frac{M}{\sum_{p=1}^{N} u_p} - t. \tag{4-28}$$

Each round has n_r samples (with replacement) so the probability that all elements of T_{t_r} are missed is at most the p-value of the test in round r

$$p^{(r)} = \left(1 - \frac{M}{\sum_{p=1}^{N} u_p} + t\right)^{n_r}. \tag{4-29}$$

If $p^{(r)} < \alpha_r$, then we can confirm the election outcome.

All that remains is to pick the sample sizes n_r and the per-round significance levels α_r. Any way of picking the n_r works as long as $n_r - n_{r-1} \geq 1$. The per-round significance levels should be chosen such that their sum is at most $\alpha = 1-c$. For example, $\alpha_r = \alpha/2^r$ or if the sample size selection rule ensures that all precincts are sampled by round R, then $\alpha_r = \alpha/R$ suffices.

The full procedure is [Sta08b]:

1. Set $r \leftarrow 1$, $n_0 \leftarrow 0$, $S_0 \leftarrow \emptyset$.
2. Sample $n_r - n_{r-1}$ times from $\mathbf{P}$, getting the set S and set $S_r \leftarrow S_{r-1} \cup S$.
3. Audit precincts in S_r that haven't yet been audited.
4. If $S_r = \{1, 2, \ldots, N\}$, the true outcome is known so stop. Otherwise calculate $t_r = \max_{p\in S_r} w_p(e_p)$ and $p^{(r)}$ from Equation (4-29).
5. If $P^{(r)} < \alpha_r$, confirm the outcome and stop. Otherwise, set $r \leftarrow r + 1$ and go to step 2.

If the procedure ends in round r, then the probability of incorrectly confirming the outcome is less than α_r. By the union bound, if the procedure ends before a full recount, then the probability of incorrectly confirming the outcome is less than $\sum_r \alpha_r \leq \alpha = 1 - c$.

By applying Stark's sequential auditing method to the PPEBWR method of Aslam et al., we have a test that can confirm an election outcome, even in the case of discrepancies found during the audit. The question of how exactly to choose the sample sizes and the per-round significance levels remain to be answered. One could select the initial sample size using the value of n given in Equation (4-22) for PPEBWR. Alternatively, one could optimize the values to increase the power of the test for the same confidence level.

4.12 Ballot-based auditing

As mentioned above, ballot based auditing is typically not an option due to the lack of availability of electronic ballots. However, the mathematics behind the selection of audits to sample is exactly the same as presented in Subsection 4.5 where we make the modification of letting N be the number of ballots rather than the number of precincts and b being the number of bad ballots. It is important that N be the number of ballots rather than the number of votes since fraud can be hidden in spoiled and undervoted ballots which may not be counted in a vote count [Dop08].

Calandrino et al. [CHF07] suggest a method of using specialized scanning/printing machines to perform most of the work of an audit. It also allows the use of ballot based auditing even for electronic voting machines that do not maintain electronic ballots. Rather than manually auditing a precinct, the paper ballots—for example the output of the printer of a DRE—are scanned in some order and a unique serial number is printed sequentially on the ballots. At the end of the scanning, a list of votes on each ballot and the corresponding serial number is printed. If the tally of votes does not match the initially reported tally, then further investigation must happen. Otherwise, the scanning/printing machine is audited itself. The ballots must be checked for a unique serial number and then a random sample are checked against the printed results.

There are several choices of how to perform the audit. In the first, the number $n(N, b, c)$ from Subsection 4.5 is computed where N is the number of total ballots and b is the number of fraudulent ballots. Each precinct is assigned a range of numbers between 1 and N according to the number of reported ballots. Then $n(N, b, c)$ numbers between 1 and N are chosen and the corresponding $\tilde{n} \leq n(N, b, c)$ precincts are machine audited as described above. One ballot from each precinct is chosen uniformly at random from each of the $\tilde{n}$ precincts and the remainder of the $n(N, b, c) - \tilde{n}$ ballots are chosen uniformly at random from the entire pool of machine audited ballots. These ballots are then checked against the results printed by the scanner/printer as described above.

Another option is to select the precincts according to one of the methods described above. Those precincts are machine audited and the number of hand audits is based on the probability of sampling k ballots in a precinct of size m given that at least one is sampled:

$$\frac{\binom{m}{k}P^k(1-P)^{m-k}}{1-(1-P)^m}, \tag{4-30}$$

where $P = 1-(1-c)^{1/b}$ is the probability that a ballot is chosen given b bad ballots.

Since audits are useful for more than detecting fraud—e.g., for uncovering malfunctioning equipment or electoral processes—all elections should have a mandatory 1% audit. For example, in a state-wide election, the audit procedure may select fewer than 1% of the precincts. In that case, the remainder of the precincts may be audited uniformly at random to bring the total number up to 1%. In states such as California where fixed percentages are mandated, this provides a way to comply with the law while still having the confidence level that following the auditing method alone confers.

5 Performing the audit

Once the precincts to audit have been selected, there is the question of how the audit should be performed. Appel [App07] distinguishes between three types of audits. The first is an audit without hand recount. In this type of audit, the vote total printed by each voting machine in the precinct will be compared to the total contained in the memory card. The second is a hand recount of a single race in the precinct. The third is a hand recount of all races in the precinct. An additional type of audit is a machine assisted audit as suggested by Calandrino et al. [CHF07] as described in Subsection 4.12.

Wherever feasible, the printed total tape for each electronic voting machine should be compared against the memory card when the card is inserted into the EMS. Checking the value on the card as compared to a printed tape should not take a significant amount of extra time per memory card. Some electronic voting machines—such as Premier's AccuVote TS—allow aggregating the counts of multiple machines onto a single memory card. This should be avoided, both to prevent potential spread of voting machine viruses—see Section 2—and to maintain one memory card and total tape per machine. If this is not feasible because there are simply too many DREs and all of the machines at the precinct level are combined, then the tape totals should be compared to the memory card ultimately used in the EMS.

When performing an audit to check for systemic problems with voting machines or electoral processes—for example the 1% on top of whichever fraud-detection auditing process is used—all of the races on a ballot should be hand counted [App07]. To gain confidence in just a single race, only that race needs to be recounted. However, election officials may wish to recount all of the races on a ballot as an added measure of confidence—especially if the cost of recounting additional races is low as compared to the cost of counting the first. This might be the case if physically separating the ballots is required [GB07].

6 Conclusion

In order to gain confidence in an election outcome in the face of vote miscounts and adversaries who wish to corrupt the results, we use post-election auditing. Requiring absolute certainty in an election outcome is not feasible so we instead want statistical evidence that the outcome is correct. Simple auditing strategies such as fixed percentage audits do not confer any statistical guarantees, especially in smaller elections or elections with a large disparity in precinct sizes.

If precincts have the same size, then we can compute the number of precincts necessary to audit uniformly at random to achieve a particular confidence level for an assumed maximum number of votes that have been shifted from one candidate to another. This can be done exactly by numerical optimization, or we can compute a conservative upper bound on the number. If precincts have different sizes, we can lower bound the number of precincts that have miscounts and then compute the audit size as before.

We can do better if we are willing to give up auditing precincts uniformly at random. This has the potential downside of causing voters in smaller precincts to feel as if their votes count less. By considering upper bounds on the amount of vote shifting in each precinct, we can sample each precinct with probability 1 minus a negative exponential in the error bound. This makes any strategy to conceal fraud as good as any other. Rather than sampling each precinct with probability independent of the others, we can do even better by sampling (with replacement) from a probability vector where the probabilities are proportional to the error bounds.

Using sequential auditing methods, we can handle discrepancies in the audit by increasing the size of the audit until either we are convinced that the outcome is correct, or we have audited every precinct. There is more work to be done in this respect. The parameters α_r, and n_r are left unspecified. While the test is correct no matter how they are chosen so long as the sum of the α_r is at most α and $n_r - n_{r-1} \geq 1$, there may be a way to choose them to optimize the power of the test. A more powerful test would allow the election outcome to be confirmed using fewer rounds of sampling.

There are several ways in which one could potentially increase the power of the audits. Stark suggests applying ideas from sequential analysis [Sta08a] using methods from financial auditing [Sta08b]. Another way would be to combine historical knowledge of how a given precinct

voted in the past on similar issues. There is a large corpus of voter history. For example, the University of California, Berkeley's Statewide Database contains election data at the precinct level including voting, registration, and geographic data from a number of elections from 1996 onwards [SWD]. This historical data could be combined with the reported votes in a precinct to determine where evidence of miscounts or fraud is most likely.

Redistricting is a potential problem. Once the district's boundaries change, the voting history is no longer completely valid. One possible solution to this problem would be to use geographic data to identify districts in close proximity and use a weighted average. Thus, even when districts change, the general history of the region is maintained.

Another factor that would need to be accounted for is variance in voting history. In a district that consistently votes for a particular party, fraud would be more likely to be noticed than in a less consistent district. As a result, all else being equal, audits should occur with higher probability in the districts with a higher variance. For example, the NEGEXP method of Subsection 4.8 could potentially be extended to account for error bounds, result difference from voting history and variance of historical data in a way that makes every strategy for fraud based on both size and historical variance as effective as any other.

There seems to be a dearth of election research in this direction. The Brennan Center for Justice's report [NBHC07] cites *An assessment of the May 2006 Election Recount and a Proposed Permanent Recount Sample Design* by Kalsbeek and Zhang of the University of North Carolina at Chapel Hill, School of Public Health's Survey Research Unit as advocating using historical data in election audits. This report does not seem to be publicly available at this time.

Another direction for future research is gaining confidence in multiple races simultaneously. A typical election consists of a number of races, potentially comprised of statewide races as well as a number of local races. Not every precinct would vote on an identical set of races but rather on some subset of the total races. All of the audit methods discussed in this paper considered a single race at a time—with the exception of the suggested mandatory 1% audit of all races. To complicate matters further, we might desire different confidence levels in each race. For example, we might require a confidence of 99% for a statewide race whereas for a city race, we might accept a confidence of only 95%.

In the case of multiple races, auditing each race independently is likely to be inefficient since there is overhead associated with each precinct audited that is independent of the number of races being audited. Similarly, selecting each precinct to audit by using the maximum confidence required of any race in the precinct is likely to be inefficient. For example, a precinct with a number of races with a wide margin and one with a small margin could have all of its races audited with unnecessarily high probability. More research is required to audit multiple races simultaneously to achieve a given set of confidence levels without performing an excessive amount of work.

Acknowledgments

The author thanks Russell Impagliazzo for an email conversation regarding the NP-hardness of Stark's pooling rule. The author also thanks his research exam committee: Hovav Shacham, Alex Snoeren, and Stefan Savage.

References

[App07] Andrew W. Appel. Effective audit policy for voter-verified paper ballots. In *Annual meeting of the American Political Science Association*, Chicago, IL, USA, Sep 2007. http://www.cs.princeton.edu/~appel/papers/appel-audits.pdf.

[APR07] Javed A. Aslam, Raluca A. Popa, and Ronald L. Rivest. On estimating the size and confidence of a statistical audit. In *EVT'07: Proceedings of the USENIX/Accurate Electronic Voting Technology on USENIX/Accurate Electronic Voting Technology Workshop*, pages 8–8, Berkeley, CA, USA, 2007. USENIX Association. http://www.usenix.org/events/evt07/tech/full_papers/aslam/aslam.pdf.

[APR08] Javed A. Aslam, Raluca A. Popa, and Ronald L. Rivest. On auditing elections when precincts have different sizes. In *EVT'08: Proceedings of the USENIX/Accurate Electronic Voting Technology on USENIX/Accurate Electronic Voting Technology Workshop*, pages 1–1, Berkeley, CA, USA, Jul 2008. USENIX Association. http://www.usenix.org/event/evt08/tech/full_papers/aslam/aslam.pdf.

[Bow07] California Secretary of State D. Bowen. "Top-To-Bottom" Review of voting machines certified for use in California. Technical report, California Secretary of State, 2007. http://sos.ca.gov/elections/elections.vsr.htm.

[CFH+07] J. A. Calandrino, A. J. Feldman, J. A. Halderman, D. Wagner, H. Yu, and W. P. Zeller. Source code review of the Diebold voting system. Technical report, California Secretary of State, Aug 2007. http://sos.ca.gov/elections/elections.vsr.htm.

[CHF07] Joseph A. Calandrino, J. Alex Halderman, and Edward W. Felten. Machine-assisted election auditing. In *EVT'07: Proceedings of*

the USENIX/Accurate Electronic Voting Technology on USENIX/Accurate Electronic Voting Technology Workshop, pages 9–9, Berkeley, CA, USA, 2007. USENIX Association. http://www.usenix.org/events/evt07/tech/full_papers/calandrino/calandrino.pdf.

[CHF08] Joseph A. Calandrino, J. Alex Halderman, and Edward W. Felten. In defense of pseudorandom sample selection. In *EVT'08: Proceedings of the USENIX/Accurate Electronic Voting Technology on USENIX/Accurate Electronic Voting Technology Workshop*, pages 3–3, Berkeley, CA, USA, 2008. USENIX Association. http://www.usenix.org/events/evt08/tech/full_papers/calandrino/calandrino.pdf.

[CWD06] Arel Cordero, David Wagner, and David Dill. The role of dice in election audits — extended abstract. In *IAVoSS Workshop on Trustworthy Elections (WOTE 2006)*, Jun 2006. http://www.cs.berkeley.edu/~arel/vote/observable_randomness.pdf.

[Dop06] Kathy Dopp. How can independent paper audits ensure election integrity. http://electionarchive.org/ucvAnalysis/US/paper-audits/Paper_Audits.pdf, Jul 2006.

[Dop08] Kathy Dopp. History of confidence election auditing development (1975 to 2008) & overview of election auditing fundamentals. http://electionarchive.org/ucvAnalysis/US/paper-audits/History-of-Election-Auditing-Development.pdf, Mar 2008.

[DS06] Kathy Dopp and Frank Stenger. The election integrity audit. http://electionarchive.org/ucvAnalysis/US/paper-audits/ElectionIntegrityAudit.pdf, 2006.

[ESI06] Election Science Institute. *DRE Analysis for May 2006 Primary Cuyahoga County, Ohio*, Aug 2006. http://bocc.cuyahogacounty.us/GSC/pdf/esi_cuyahoga_final.pdf.

[Eve07] Sarah P. Everett. *The Usability of Electronic Voting Machines and How Votes Can Be Changed Without Detection*. PhD thesis, Rice University, Houston, Texas, USA, May 2007. http://chil.rice.edu/alumni/petersos/EverettDissertation.pdf.

[FHF07] Ariel J. Feldman, J. Alex Halderman, and Edward W. Felten. Security analysis of the Diebold AccuVote-TS voting machine. In *EVT'07: Proceedings of the USENIX/Accurate Electronic Voting Technology on USENIX/Accurate Electronic Voting Technology Workshop*, pages 2–2, Berkeley, CA, USA, 2007. USENIX Association. http://itpolicy.princeton.edu/voting/ts-paper.pdf.

[GB07] Stephen N. Goggin and Michael D. Byrne. An examination of the auditability of voter verified paper audit trail (VVPAT) ballots. In *EVT'07: Proceedings of the USENIX/Accurate Electronic Voting Technology on USENIX/Accurate Electronic Voting Technology Workshop*, pages 10–10, Berkeley, CA, USA, 2007. USENIX Association. http://www.usenix.org/events/evt07/tech/full_papers/goggin/goggin.pdf.

[GBG+08] Stephen N. Goggin, Michael D Byrne, Juan E. Gilbert, Gregory Rogers, and Jerome McClendon. Comparing the auditability of optical scan, voter verified paper audit trail (VVPAT) and video (VVVAT) ballot systems. In *EVT'08: Proceedings of the USENIX/Accurate Electronic Voting Technology on USENIX/Accurate Electronic Voting Technology Workshop*, pages 10–10, Berkeley, CA, USA, 2008. USENIX Association. http://www.usenix.org/events/evt08/tech/full_papers/goggin/goggin.pdf.

[HRSW08] J. Alex Halderman, Eric Rescorla, Hovav Shacham, and David Wagner. You go to elections with the voting system you have: Stop-gap mitigations for deployed voting systems. In *EVT'08: Proceedings of the USENIX/Accurate Electronic Voting Technology on USENIX/Accurate Electronic Voting Technology Workshop*, pages 4–4, Berkeley, CA, USA, 2008. USENIX Association. http://www.usenix.org/events/evt08/tech/full_papers/halderman/halderman.pdf.

[Imp08] Russell Impagliazzo. Personal communication, August 2008.

[Kib08] Robert Kibrick. Voter-verified paper record legislation, May 2008. http://www.verifiedvoting.org/article.php?list=type&type=13.

[Knu97] Donald E. Knuth. *The Art of Computer Programming*, volume 2. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, third edition, 1997.

[KSRW04] T. Kohno, A. Stubblefield, A.D. Rubin, and D.S. Wallach. Analysis of an electronic voting system. *Security and Privacy, 2004. Proceedings. 2004 IEEE Symposium on*, pages 27–40, May 2004. http://www.cs.washington.edu/homes/yoshi/papers/eVoting/vote.pdf.

[MSL+07] John McCarthy, Howard Stanislevic, Mark Lindeman, Arlene Ash, Vittorio Addona, and Mary Batcher. Percentage-based versus SAFE vote tabulation auditing: A graphic comparison. http://www.verifiedvoting.org/downloads/SAFE-Auditing-Nov-2-Final4.pdf, Nov 2007.

[NBHC07] Lawrence Norden, Aaron Burstein, Joseph Lorenzo Hall, and Margaret Chen. Post-election audits: Restoring trust in elections. http://brennan.3cdn.net/1c6df39b4f8e755d48_89m6vgnnt.pdf, 2007.

[Riv06] Ronald L. Rivest. On estimating the size of a statistical audit. http://theory.csail.mit.edu/~rivest/Rivest-OnEstimatingTheSizeOfAStatisticalAudit.pdf, Nov 2006.

[Sal75] R. G. Saltman. 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, March 1975. http://csrc.nist.gov/publications/nistpubs/NBS_SP_500-30.pdf.

[Sta06] Howard Stanislevic. Random auditing of e-voting systems: How much is enough? http://www.votetrustusa.org/pdfs/VTTF/EVEPAuditing.pdf, Aug 2006.

[Sta08a] Philip B. Stark. Conservative statistical post-election audits. *Ann. Appl. Stat.*, 2(2):550–581, Mar 2008. http://statistics.berkeley.edu/~stark/Preprints/conservativeElectionAudits07.pdf.

[Sta08b] Phillip B. Stark. Election audits by sampling with probability proportional to an error bound: dealing with discrepancies. Draft. http://www.stat.berkeley.edu/~stark/Preprints/ppebwrwd08.pdf, Feb 2008.

[Sta08c] Phillip B. Stark. A sharper discrepancy measure for post-election audits. *Ann. Appl. Stat.*, Apr 2008. To appear. http://statistics.berkeley.edu/~stark/Preprints/pairwise08.pdf.

[SWD] University of California, Berkeley Statewide Database. http://swdb.berkeley.edu/.

[Wan04] Jonathan Wand. Auditing an election using sampling: The impact of bin size on the probability of detecting manipulation. http://wand.stanford.edu/elections/probability.pdf, Feb 2004.

[Was04] Larry Wasserman. *All of Statistics : A Concise Course in Statistical Inference (Springer Texts in Statistics)*. Springer, September 2004.

A Arithmetic geometric harmonic means inequality

Proposition A.1 (Arithmetic-geometric-harmonic means inequality)**.** *Given any set of positive, real-valued numbers $x_1, x_2, \ldots, x_k$, the arithmetic, geometric, and harmonic means are related by*

$$\frac{1}{k}\sum_{i=1}^{k} x_i \geq \sqrt[k]{\prod_{i=1}^{k} x_i} \geq k\left(\sum_{i=1}^{k}\frac{1}{x_i}\right)^{-1}. \tag{A-1}$$

Proof. To see the first inequality, let $\mu = (1/k)\sum_{i=1}^{k} x_i$ be the arithmetic mean and $\rho = \prod_{i=1}^{k} x_i^{1/k}$ be the geometric mean. From elementary calculus, $e^x - x - 1$ has a global minimum at $x = 0$ so $e^x \geq x + 1$ for all real values x. Since each $x_i > 0$, $\mu > 0$ and thus

$$\exp\left(\frac{x_i}{\mu} - 1\right) \geq \frac{x_i}{\mu}, \tag{A-2}$$

for each i. Taking products of both sides, we have

$$\prod_{i=1}^{k}\exp\left(\frac{x_i}{\mu} - 1\right) \geq \prod_{i=1}^{k}\frac{x_i}{\mu} \tag{A-3}$$

$$\exp\left(\sum_{i=1}^{k}\frac{x_i}{\mu} - k\right) \geq \frac{\rho^k}{\mu^k} \tag{A-4}$$

Since $\sum_{i=1}^{k} x_i = k\mu$, the left-hand side of Inequality (A-4) is $e^0 = 1$ so by taking kth roots, we have $\mu \geq \rho$.

To prove the second inequality in Inequality (A-1), we employ the first. By the arithmetic-geometric means inequality, we have

$$\frac{1}{k}\sum_{i=1}^{k}\frac{1}{x_i} \geq \prod_{i=1}^{k}\frac{1}{x_i^{1/k}} \tag{A-5}$$

and thus by taking the inverse,

$$k\left(\sum_{i=1}^{k}\frac{1}{x_i}\right)^{-1} \leq \sqrt[k]{\prod_{i=1}^{k} x_i}. \tag{A-6}$$

□

B Bounds for optimal sample size without replacement

We follow [APR07] to derive the upper and lower bounds for sample size without replacement when precincts have equal size.

B.1 Weak upper bound

To compute the weaker upper bound in Equation (4-5), we use the dual equation for $e(N,b,n)$ in Equation (4-11).

$$\begin{aligned} e(N,b,n) &= \prod_{k=0}^{b-1}\left(1-\frac{n}{N-k}\right) \\ &\le \prod_{k=0}^{b}\left(1-\frac{n}{N}\right) \\ &= \left(1-\frac{n}{N}\right)^b. \end{aligned} \tag{B-1}$$

Recall that we wish to bound $e(N,b,n)$ by the significance level α. Combining that with Inequality (B-1), it suffices to bound

$$\left(1-\frac{n}{N}\right)^b \le 1-c. \tag{B-2}$$

Solving Inequality (B-2) for $n = n(N,b,c)$ and writing $(1-c)^{1/b}$ as $\exp(\log(1-c)/b)$, we get the bound in Equation (4-5):

$$n(N,b,c) \ge N\big(1-\exp(\log(1-c)/b)\big). \tag{B-3}$$

B.2 Tight upper bound

We get a tighter upper bound by using the arithmetic-geometric-harmonic means inequality in Proposition A.1 together with Equation (4-11). We first write $e(N,b,n) = (e(N,b,n)^{1/b})^b$ and then we use the arithmetic-geometric means inequality to get

$$\begin{aligned} e(N,b,n) &= \left(\prod_{k=0}^{b-1}\left(1-\frac{n}{N-k}\right)^{1/b}\right)^b \\ &\le \left(\frac{1}{b}\sum_{k=0}^{b-1}\left(1-\frac{n}{N-k}\right)\right)^b \\ &= \left(1-\frac{n}{b}\sum_{k=0}^{b-1}\frac{1}{N-k}\right)^b. \end{aligned} \tag{B-4}$$

Since we want to bound $e(N,b,n) \le 1-c$, we solve

$$\left(1-n\frac{1}{b}\sum_{k=0}^{b-1}\frac{1}{N-k}\right)^b \le 1-c \tag{B-5}$$

for n, getting

$$n \ge b\left(\sum_{k=0}^{b-1}\frac{1}{N-k}\right)^{-1}\cdot\,(1-(1-c)^{1/b}). \tag{B-6}$$

This is a conservative upper bound on the optimal number $n_{\mathrm{OPT}}(N,b,c)$ of precincts we need to audit; however, by using the arithmetic-harmonic means inequality, we have

$$\begin{aligned} b\left(\sum_{k=0}^{b-1}\frac{1}{N-k}\right)^{-1} &\le \frac{1}{b}\sum_{k=0}^{b-1}(N-k) \\ &= \frac{1}{b}\left(\frac{N(N+1)}{2}-\frac{(N-b)(N-b+1)}{2}\right) \\ &= N-\frac{b-1}{2}. \end{aligned} \tag{B-7}$$

We can now replace Inequality (B-6) with the slightly weaker inequality given in Inequality (4-12):

$$n(N,b,c) \ge \left(N-\frac{b-1}{2}\right)\big(1-\exp(\log(1-c)/b)\big). \tag{B-8}$$

This has the advantage of being much easier to calculate on a standard hand calculator.

B.3 Lower bound

The computation of the lower bound is similar to the weak upper bound. We start with Equation (4-11) to get

$$e(N,b,n) = \prod_{k=0}^{b-1}\left(1-\frac{n}{N-k}\right) \ge \left(1-\frac{n}{N-b+1}\right)^b. \tag{B-9}$$

For $e(N,b,c) > 1-c$, we must have

$$n < \big(N-(b-1)\big)\big(1-(1-c)^{1/b}\big), \tag{B-10}$$

whence we get a lower bound on $n_{\mathrm{OPT}}(N,b,c)$:

$$n_{\mathrm{OPT}} \ge \left\lceil \big(N-(b-1)\big)\big(1-(1-c)^{1/b}\big)\right\rceil, \tag{B-11}$$

where the ceiling comes since n_{OPT} must be an integer.

We can now bound how far $n(N,b,c)$ is from the optimum $n_{\mathrm{OPT}}(N,b,c)$. Since we need $n(N,b,c)$ to be integral, we let

$$n(N,b,c) = \left\lceil \left(N-\frac{b-1}{2}\right)(1-\alpha^{1/b})\right\rceil \tag{B-12}$$

where $\alpha = 1-c$ and subtract to get

$$n(N,b,c) - n_{\mathrm{OPT}}(N,b,c) \le \frac{b-1}{2}(1-\alpha^{1/b}) + 1 \tag{B-13}$$

where the $+1$ comes from $\lceil x\rceil - \lceil y\rceil \le \lceil x-y\rceil + 1$. Note that this difference is independent of N. For a fixed $0 < \alpha < 1$, let $D(b) = (b-1)(1-\alpha^{1/b})/2 + 1$ be the difference. Computing the first derivative of D, we get

$$\begin{aligned} D'(b) &= \frac{1}{2}\left(1-\alpha^{1/b}+\frac{\alpha^{1/b}(b-1)\log\alpha}{b^2}\right) \\ &= \frac{1}{2b^2}\big(b^2(1-\alpha^{1/b}) + b\alpha^{1/b}\log\alpha - \alpha^{1/b}\log\alpha\big) \end{aligned} \tag{B-14}$$

Since $-\log\alpha > 0$, we can drop the last parenthetical term and simplify to get

$$\begin{aligned} D'(b) &> \frac{1}{2b^2}\left(b^2(1-\alpha^{1/b}) + b\alpha^{1/b}\log\alpha\right) \\ &= \frac{\alpha^{1/b}}{2b}\left(b(\alpha^{-1/b}-1) + \log\alpha\right). \end{aligned} \tag{B-15}$$

Since $a^x = \exp(x\log a)$ and $e^x \geq 1+x$ for all real x (see Appendix A),

$$\exp\left(-\frac{1}{b}\log\alpha\right) \geq 1 - \frac{1}{b}\log\alpha. \tag{B-16}$$

Plugging this into Inequality (B-15), we get

$$D'(b) > \frac{\alpha^{1/b}}{2b}(-\log\alpha + \log\alpha) = 0. \tag{B-17}$$

Since the derivative of D is always positive (for $b > 0$), the difference between $n(N,b,c)$ and $n_{\text{OPT}}(N,b,c)$ is strictly increasing in b. In the limit as $b \to \infty$, we find

$$\begin{aligned} \lim_{b\to\infty} D(b) &= 1 + \lim_{b\to\infty} \frac{1-\alpha^{1/b}}{\frac{2}{b-1}} \\ &= 1 - \lim_{b\to\infty} \frac{\alpha^{1/b}\log\alpha}{\frac{2b^2}{(b-1)^2}} \\ &= 1 - \frac{\log\alpha}{2}. \end{aligned} \tag{B-18}$$

Thus,

$$n(N,b,c) - n_{\text{OPT}}(N,b,c) \leq 1 + \left\lceil -\frac{\log(1-c)}{2} \right\rceil, \tag{B-19}$$

which matches Inequality (4-13).

C Condition for certifying the election

Following [Sta08a], we show that if $E < M$, then the apparent set of winners K_w must be the actual winners. We can actually do much better than this by considering pairwise discrepancies, but the calculations are much more complex [Sta08c]. To simplify the notation, define the function $(\cdot)_+ : \mathbb{R} \to \mathbb{R}_{\geq 0}$ by $(x)_+ = \max\{x, 0\}$. In this notation, Equation (4-2) becomes

$$e_p = \sum_{k\in K_w} (v_{k,p} - a_{k,p})_+ + \sum_{k\in K_l} (a_{k,p} - v_{k,p})_+. \tag{C-1}$$

Define the net discrepancy to be

$$\mathcal{E} = \sum_{k\in K_w} (V_k - A_k)_+ + \sum_{k\in K_l} (A_k - V_k)_+. \tag{C-2}$$

For the election outcome to be correct, it must be the case that the number of votes for each of the apparent winners is greater than the number of votes for each of the apparent losers. In symbols,

$$D = \min_{k\in K_w} A_k - \max_{k\in K_l} A_k > 0. \tag{C-3}$$

We can bound this difference D from below by

$$\begin{aligned} D &\geq \left(\min_{k\in K_w} V_k - \max_{k\in K_w} (V_k - A_k)_+\right) \\ &\quad - \left(\max_{k\in K_l} V_k + \max_{k\in K_l} (A_k - V_k)_+\right) \\ &\geq \left(\min_{k\in K_w} V_k - \sum_{k\in K_w} (V_k - A_k)_+\right) \\ &\quad - \left(\max_{k\in K_l} V_k + \sum_{k\in K_l} (V_k - A_k)_+\right) \\ &= M - \mathcal{E}. \end{aligned} \tag{C-4}$$

From Inequality (C-3), if $\mathcal{E} < M$, then the apparent winners must be the actual winners. All that remains is the bound the net discrepancy $\mathcal{E}$ from above by the total discrepancy E.

For any set of real numbers S,

$$\sum_{x\in S} (x)_+ = \sum_{\substack{x\in S \\ x>0}} x = \left(\sum_{\substack{x\in S \\ x>0}} x\right)_+ \geq \left(\sum_{x\in S} x\right)_+. \tag{C-5}$$

Expanding Equation (C-2), we have

$$\begin{aligned} \mathcal{E} &= \sum_{k\in K_w} \left(\sum_{p=1}^{N} (v_{k,p} - a_{k,p})\right)_+ + \sum_{k\in K_l} \left(\sum_{p=1}^{N} (a_{k,p} - v_{k,p})\right)_+ \\ &\leq \sum_{p=1}^{N} \left(\sum_{k\in K_w} (v_{k,p} - a_{k,p})_+ + \sum_{k\in K_l} (a_{k,p} - v_{k,p})_+ \right). \end{aligned} \tag{C-6}$$

Combining with Equation (C-1) and $E = \sum_{p=1}^{N} e_p$, we get $\mathcal{E} \leq E$ and thus if $E < M$, then the election can be certified.

D Demonstrating the efficiency of PPEBWR over NEGEXP

To see that PPEBWR is more efficient than NEGEXP, it suffices to consider a single precinct p and show that the probability of auditing p is smaller with PPEBWR than NEGEXP for a given confidence level c [APR08]. That is, we need to show

$$1 - \left(1 - \frac{u_p}{U}\right)^n \leq 1 - \alpha^{u_p/M}, \tag{D-1}$$

where $U = \sum_{p=1}^{N} u_p$, $\alpha = 1 - c$, and $n \geq \log_{1-M/U}(\alpha)$ by Equation (4-22). In particular, we need to show

$$\left(1 - \frac{u_p}{U}\right)^n \geq \alpha^{u_p/M}. \tag{D-2}$$

By straight-forward algebraic manipulation, we have

$$\begin{aligned}\left(1-\frac{u_p}{U}\right)^n &\geq \left(1-\frac{u_p}{U}\right)^{\log_{1-M/U}(\alpha)} \\ &= \left(\left(1-\frac{u_p}{U}\right)^{U/u_p}\right)^{(u_p/U)\log_{1-M/U}(\alpha)} \qquad \text{(D-3)} \\ &\quad .\end{aligned}$$

If we require that $u_p \leq M$ for all precincts p—a reasonable requirement since we may cap u_p at M when deciding error bounds without loss of statistical power or confidence—then since $(1-1/x)^x$ is strictly increasing for $x \geq 1$, we may substitute $(1-M/U)^{U/M} \leq (1-u_p/U)^{U/u_p}$ to get

$$\begin{aligned}\left(1-\frac{u_p}{U}\right)^n &\geq \left(\left(1-\frac{M}{U}\right)^{U/M}\right)^{(u_p/U)\log_{1-M/U}(\alpha)} \\ &= \left(1-\frac{M}{U}\right)^{(u_p/M)\log_{1-M/U}(\alpha)} \qquad \text{(D-4)} \\ &= \alpha^{u_p/M},\end{aligned}$$

as required. Thus PPEBWR samples each precinct with probability bounded above by that of NEGEXP—for the same confidence level—so PPEBWR is more efficient.