On Auditing Elections When Precincts Have Different Sizes

Javed A. Aslam
College of Computer and Information Science
Northeastern University
Boston, MA 02115
jaa@ccs.neu.edu

Raluca A. Popa and Ronald L. Rivest
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139
{ralucap,rivest}@mit.edu

Draft of January 17, 2008

Abstract

We address the problem of auditing an election when precincts may have different sizes.

Prior work in this field has emphasized the simpler case when all precincts have the same size. Using auditing methods developed for use with equal-sized precincts can, however, be inefficient or result in a loss of statistical confidence when applied in to elections with variable-sized precincts.

We survey, evaluate, and compare a variety of approaches to the variable-sized precinct auditing problem, including the SAFE method [11] based on theory developed for equal-sized precincts. We introduce new methods such as the negative-exponential method (NEGEXP) that select precincts independently for auditing with predetermined probabilities, and the "PPEBWR" method that uses a sequence of rounds to select precincts with replacement according to some predetermined probability distribution that may depend on error bounds for each precinct (hence the name PPEBWR: probability proportional to error bounds, with replacement), where the error bounds may depend on the sizes of the precincts, or on how the votes were cast in each precinct.

We give experimental results showing that NEGEXP and PPEBWR can *dramatically* reduce (by a factor or two or three) the cost of auditing compared to methods such as SAFE that depend on the use of uniform sampling. Sampling so that larger precincts are audited with appropriately larger probability can yield large reductions in expected number of votes counted in an audit.

We also examine the optimal auditing strategy, which is nicely representable as a linear programming problem but only really computable for small elections (fewer than a dozen precincts).

We conclude with some recommendations for practice.

1 Introduction

Post-election audits are an essential tool for ensuring the integrity of election outcomes. Such audits can detect, with high probability, both errors due to machine misprogramming and errors due to malicious manipulation of electronic vote totals. Since such audits are based on statistical samples, they can be performed quite efficiently and economically. This paper explores auditing approaches that achieve improved efficiency (sometimes by a factor of two or three, measured in terms of number of votes counted) over previous methods for selecting a sample of precincts to audit.

Suppose we have an election with n precincts, $P_1, \ldots, P_n$.

Let v_i denote the number of voters who voted in precinct P_i; we call v_i the "size" of precinct P_i. Let the total number of such voters be $V = \sum_i v_i$. Assume without loss of generality that $v_1 \geq v_2 \geq \cdots \geq v_n$.

Precinct sizes can vary dramatically in practice. Figure 3 show a graph of precinct sizes for the 2006 Minnesota governor's race. There were 4123 precincts, with an average size of 535 votes. The largest precinct had 4110 votes, while ten precincts had no votes at all. Methods for auditing elections must, if they are to be efficient and effective, take such precinct size variations into account.

Suppose further that in precinct P_i we have both electronic records and paper records for each voter. The electronic records are easy to tally.

For the purposes of this paper, the paper records are used only as a source of authoritative information when the electronic records are audited. They may be considered more authoritative since the voters may have verified them directly. In practice, more care is needed, since the electronic records could reasonably be judged as more authoritative in situations where the paper records were obviously damaged or lost and the electronic records appear undamaged.

Auditing is desirable since a malicious party, the "adversary," may have manipulated some of the electronic tallies so that a favored candidate appears to have won the election. It is also possible that a simple software bug caused the electronic tallies to be inaccurate. However, we focus on detecting malicious adversarial behavior, as that is the more challenging task.

A precinct can be "audited" by re-counting by hand the paper records of that precinct to confirm that they match the electronic totals for that precinct. We ignore here the important fact that hand-counting may be inaccurate, and assume that any discrepancies are due to fraud on the part of the adversary. In practice, the discrepancy might have to be larger than some prespecified threshold to trigger a conclusion of fraud in that precinct.

See the overviews [9, 13, 7] for information about current election auditing procedures. In this paper we ignore many of the complexities of real elections; these complexities are addressed in other papers. We do so in order to focus on our central issue: *how to select a sample of precincts to audit when the precincts have different sizes.*

See Neff [12], Cordero et al. [6], Saltman [16], Dopp et al. [8], and Aslam et al. [2], for additional discussion of the mathematics of auditing, and additional references to the literature.

1.1 Outline

We begin with an overview of the auditor's general approach in Section 2. In Section 3 we review the adversary's objectives and capabilities. Section 4 then reviews the auditor's strategy.

Some known results for auditing when all precincts have equal size are discussed in Section 5.

We next review in Section 6 the "SAFE" method, which deals with variable-sized precincts using the mathematics developed for equal-sized preincts, by first deriving a lower bound on the number of precincts that must have been corrupted, if the election outcome was changed.

Section 7 introduces *basic* auditing methods, where each precinct is chosen *independently* according to a precomputed probability distribution.

A particular instance of the general basic auditing method is next introduced in Section 8; this method is called the "negative-exponential" (NEGEXP) auditing method.

We then turn our attention to auditing procedures where the precincts are *not* chosen independently. Section 9 introduces the method of sampling with probability proportional to error bounds, with replacement (PPEBWR); a special case of this procedure is PPSWR, "sampling with probability proportional to size, with replacement."

Section 10 discusses *vote-dependent* auditing, where the probability of auditing a precinct depends on the actual vote counts for each candidate.

Section 11 gives experimental results for our methods, using data from Ohio and Minnesota.

Then Section 12 presents a method based on linear programming for determining an optimal auditing procedure. Unfortunately this approach appears to be computationally too expensive for practical use.

A quick discussion of the effects of "aggregation" (merging precincts together) is given in Section 13.

We close in Section 14 with some discussion and recommendations for practice.

Appendix A gives a proof of the "Aggregation Theorem" stated in Section 13.

2 Auditing objectives and costs

We assume for now that the election is a winner-take-all (plurality) election from a field of k candidates; Section 14 discusses how our approaches generalize for other election types.

After the election is over, the auditor randomly selects a sample of precincts in which to perform a post-election audit. In each selected precinct the paper ballots are counted by hand, and the totals so obtained compared with the electronic tallies.

We assume that the paper ballots are maintained securely and that they can be accurately counted during the post-election audit.

The auditor wishes to assure himself (and everyone else) that the level of error and/or fraud in the election is likely to be low or nonexistant, or at least insufficient to have changed the election outcome.

If the audit finds no (significant) discrepancies between the electronic and paper tallies, the auditor announces that no fraud was discovered, and the election results may be certified by the appropriate election official.

On the other hand, if significant discrepancies are discovered between the electronic and paper tallies, then additional investigations may be needed to determine the nature and extent of the problem. For example, state or federal law may then require a full recount of the paper ballots. Stark [18] discusses procedures for incrementally auditing larger and larger samples when discrepancies are found, until the desired level of significance is achieved (i.e. until the probability that an incorrect election outcome is announced is made small enough).

When planning the audit, the auditor knows the number r_{ij} of reported (electronic) votes for each candidate j in precinct i, and also the total size v_i (total number of votes cast) of each precinct P_i.

The auditor also knows the reported *margin of victory*, denoted $M^{(r)}$ of the winning candidate over the runner-up—this is the difference between the number of votes reported for the apparently victorious candidate and the number of votes reported for the runner-up. More auditing is generally appropriate when the margin of victory is smaller (see, e.g., Norden et al. [13]).

2.1 Auditing objective

We assume that level of auditing effort should be chosen to achieve a pre-specified level of confidence in the election outcome. This is, we believe, the correct approach. It is also the efficient approach. Naive methods that audit, say, a fixed fraction of precincts tend to waste taxpayer dollars when the margin of victory is large, and tend to provide poor confidence in the election outcome when the margin of victory is small. See McCarthy et al. [11] for discussion of this point.

We thus assume that the auditor desires to test at a certain *significance level* α that the declared election result is correct—that is, that error or fraud is unlikely to have affected the election outcome. Without doing a complete recount, one can't be absolutely sure, but a well-chosen audit can reduce the likelihood that significant fraud or error has gone undetected. Choosing a significance level of $\alpha = 0.01$, $\alpha = 0.05$, or $\alpha = 0.10$ means that the chance that error large enough to have changed the election outcome will go undetected is respectively one in one hundred, one in twenty, or one in ten.*

We let c denote the "confidence level" of the audit, where

$$c = 1 - \alpha \ . \tag{1}$$

Thus, a test at significance level $\alpha = 5\%$ provides a confidence level of $c = 95\%$ (that error significant enough to have changed the election outcome will be detected in the audit).

If we follow Stark [18] in adopting as our null hypothesis that "the (electronic) election outcome is incorrect", then the significance level α of a particular auditing method is an upper bound on the probability that the null hypothesis will be rejected (i.e. the electronic election outcome will be accepted) when in fact the null hypothesis is true (the electronic election outcome is wrong).

2.2 Choosing a sample

As a function of the precinct sizes, the reported votes cast for each candidate, and thus the reported margin of victory, the auditor will determine how to randomly select an appropriately-sized sample of the precincts to be audited.

In this paper we explore three methods by which the auditor chooses a sample:

- **[BASIC]** The auditor determines a probability *for each precinct* that it will be audited, based on the size of the precinct and on the overal margin of victory, and then *independently* selects each precinct with the specified probability. These are the *basic* auditing strategies discussed in Sections 7–8.

- **[WITH REPLACEMENT]** The auditor determines a probability for each precinct that it will be selected during a round, and does t rounds of drawing *with replacement* to determine the precincts to be audited. Because this is "sampling with replacement," a precinct will be placed back into the collection of precincts after it drawn; it thus may be drawn more than once. A precinct will be audited if it is drawn at least once. An example of this approach is the PPEBWR (sampling with probability proportional to error bounds with replacement) method of Section 9. A special case of the PPEBWR method is the PPSWR method: sampling "with probability proportional to size, with replacement."

- **[OPTIMAL]** The auditor determines a probability *for each subset* of precincts specifying the probability that that subset will be audited. This yields the optimal auditing strategy presented in Section 12.

*Our methods can also be adapted to handle requirements such as auditing a certain number of precincts, or recounting a certain number of votes; see Appendix B.

2.3 Auditing cost

When all precincts have the same size, it is reasonable to measure the cost of performing an audit in terms of the *number of precincts* audited.

However, when precincts have a variety of sizes, the number of precincts audited is not a good measure of auditing cost. Rather, the *number of votes counted* appears to be the best measure of auditing cost. Each vote takes a certain amount of time to examine and recount. The auditing cost is most reasonably measured in person-hours, which will be proportional to the number of votes recounted.

3 Adversarial Objectives

We assume the adversary wishes to corrupt enough of the electronic tallies so that his favored candidate wins the most votes according to the reported electronic tallies. Without loss of generality, we'll let candidate 1 be the adversary's favored candidate.

The adversary tries to do his manipulations in such a way as to minimize the chance that his changes to the electronic tallies will be caught during the post-election audit.

Let a_{ij} denote the *actual* number of (paper) votes for candidate j in precinct i, and let r_{ij} denote the *reported* number of (electronic) votes for candidate j in precinct i.

With no adversarial manipulation, we will have

$$r_{ij} = a_{ij}$$

for all i and j. We ignore in this paper small explainable discrepancies that can be handled by slight modifications to the procedures discussed here.

We assume that "reconciliation" is performed when the election is over, confirming that the number of votes recorded electronically is equal to the number of votes recorded on paper; an adversary would presumably not try to make these totals differ, but only shift the electronic tallies to favor his candidate at the expense of other candidates. We thus have for all i:

$$\sum_j a_{ij} = \sum_j r_{ij} = v_i \ ;$$

the total number of paper votes cast in precinct i is equal to the number of electronic votes cast in precinct i; this number is v_i, the "size" of precinct i.

Let A_j denote the total actual number of votes for candidate j:

$$A_j = \sum_i a_{ij} \ ,$$

and let R_j denote the total number of votes reported for candidate j:

$$R_j = \sum_i r_{ij} \ .$$

The adversary's favored candidate, candidate 1, will be the winner of the electronic report totals if

$$R_1 > \max(R_2, R_3, \ldots, R_k) \ .$$

We assume for now that the election is really between candidate 1 and candidate 2, so that the adversary's objective is to ensure that candidate 1 is reported to win the election and that candidate 2 is not. There may be other candidates in the race, but for the moment we'll assume that they are minor candidates. It is also convenient to consider "invalid" and "undervote" to be such "minor candidates" when doing the tallying.

The adversary can manipulate the election in favor of his candidate by shifting the electronic tallies from one candidate to another. He might move votes from some candidate to candidate 1. Or move votes from candidate 2 to some other candidate. These manipulations can change the election outcome, and yield a false "margin of victory." The margin of victory plays a key role in our analysis.

Let $M^{(a)}$ denote the "actual margin of victory" (in votes) of candidate 1 over candidate 2:

$$M^{(a)} = A_1 - A_2 \ .$$

Let $M = M^{(r)}$ denote the "reported margin of victory" (in votes) for candidate 1 over candidate 2:

$$M = M^{(r)} = R_1 - R_2 \ .$$

Note that $M = M^{(r)}$ will be known to the auditor at the beginning of the audit, but that $M^{(a)}$ will not.

The adversary may be in a situation initially where $M^{(a)} < 0$ (i.e. $A_1 < A_2$); that is, his favored candidate, candidate 1, has lost to candidate 2. The adversary must, if he hopes to change the election outcome, manipulate the (electronic) votes so that $M^{(r)} > 0$ (i.e. so that $R_1 > R_2$) and do so in a way that he hopes will go undetected.

The "error" e_i^* in favor of candidate 1 introduced in the margin of victory computation in precinct i by the adversary's manipulations is (in votes):

$$e_i^* = (r_{i1} - r_{i2}) - (a_{i1} - a_{i2}) \ ;$$

Here $(r_{i1} - r_{i2})$ is the reported margin of victory for candidate 1, while $(a_{i1} - a_{i2})$ is his actual margin of victory, so their difference is the amount of error introduced by the adversary in the margin of victory.

An upper bound on the amount by which the adversary can improve the margin of victory in favor of his candidate in precinct 1 is:

$$e_i^* < 2a_{i2} + \sum_{j>2} a_{ij} = v_i - a_{i1} + a_{i2} \ . \qquad (2)$$

Each vote moved from candidate 2 to candidate 1 improves the margin by 2 votes, and each vote moved from candidate j ($j > 2$) to candidate 1 improves the margin by 1 vote.

Let E^* denote the total error (in votes, from all precincts) introduced in the margin of victory computation by the adversary:

$$E^* = \sum_i e_i^* \ .$$

Clearly,

$$M^{(r)} = M^{(a)} + E^* \ . \qquad (3)$$

That is, the reported margin of victory is equal to the actual margin of victory, plus the error introduced by the adversary.

The adversary has to introduce enough error E^* so that the reported margin of victory $M^{(r)}$ becomes positive, even though the initial (actual) margin of victory $M^{(a)}$ is negative. Thus, the amount of error introduced satisfies both of the inequalities:

$$E^* > -M^{(a)} \text{ , and} \qquad (4)$$

$$E^* > M^{(r)} \qquad (5)$$

The second inequality is of most interest to the auditor, since at the beginning of the audit the auditor knows $M^{(r)}$ but not $M^{(a)}$. For convenience, we shall use

$$M = M^{(r)}$$

in the sequel, and let m denote the fraction of votes represented by the margin of victory:

$$m = M/V$$

(recall that V denotes the total number of votes cast: $V = \sum_i v_i$).

We assume here that the adversary wishes to change the election outcome while minimizing the *probability of detection*—that is, while minimizing the chance that one or more of the precincts chosen to be audited will be one that has been corrupted.

If the post-election audit fails to find any error, the adversary's candidate might be declared the winner, while in fact some other candidate (e.g. candidate 2) actually should have won.

The adversary might not be willing to corrupt all available votes in a precinct; this would generate too much suspicion. Dopp and Stenger [8] suggest that the adversary might not dare to flip more than a fraction $s = 0.20$ of the votes in a precinct. The value s is also denoted WPM in the literature, and called the *Within-Precinct-Miscount*.

The presentation here depends heavily on the use of such upper bounds on e_i^*. We use e_i to denote such an upper bound on e_i^*. Following Dopp and Stenger, we would have as an upper bound e_i for e_i^*:

$$e_i = 2sv_i \ . \quad (6)$$

We call this the "*Linear Error Bound Assumption*". The factor of 2 occurs since we assume that the adversary is able to switch sv_i votes from candidate 2 to candidate 1.

We may also presume that the adversary knows the general form of the auditing method. Indeed, the auditing method may be mandated by law, or described in public documents. While the adversary may not know which specific precincts will be chosen for auditing, because they are determined by rolls of the dice or other random means, the adversary is assumed to know the method by which those precincts will be chosen, and thus to know the probability that any particular precinct will be chosen for auditing.

As a minor observation, we note that the the adversary can be assumed to use a *deterministic* method, and to pick the precincts to be corrupted in a way that is a deterministic function of actual vote totals, the vote-shift fraction s, and the public details of the auditing method. For example, if it is known that the auditing method will pick precincts uniformly at random, then the adversary may do best by corrupting a few of the largest precincts only, in order to be able to achieve his goal while corrupting as few precincts as possible. A dynamic programming algorithm (see Rivest [14]) gives the general solution to this problem of picking the precincts to be corrupted, both for the case that precincts are picked uniformly at random by the auditor, and for the case that the auditing probabilities are non-uniform.

We let Q denote the set of corrupted precincts, and let b denote the number $|Q|$ of corrupted precincts.

4 Auditing Method

4.1 Types of audits

There are many different ways to perform an audit; see Norden et al. [13] for discussion. We focus in this paper on how the sample is selected; an auditing method is one of following five types:

A *fixed audit* determines the amount of auditing to do by fiat—e.g., it selects a fixed number of precincts (or votes) to be counted (or perhaps a fixed percentage, instead of a fixed number). It does not pay attention to the precinct sizes, the reported margin of victory, or the reported vote counts. Fixed audits are simple to understand, but are frequently very costly or statistically weak.

If an audit is not a fixed audit, it is an *adjustable audit*—the size of the audit is adjustable according to various parameters of the election. There are four types of adjustable audits, in order of increasing utilization of available parameter information.

The first type of adjustable audit is a *margin-dependent audit*. Here the selection of precincts to be audited depends only on the reported margin of victory M. An election that is a landslide (with a very large margin of victory) results in smaller audit sample sizes than an election that is close.

The second type of adjustable audit is a *size-dependent audit*. Here the selection of precincts to be audited depends not only on the reported margin of victory M but also on the precinct sizes $\{v_i\}$. A size-dependent audit audits larger precincts with higher probability and audits small precincts with smaller probability. This reflects the fact that the larger precincts are "juicier targets" for the adversary. Overall, the total amount of auditing work performed may easily be less than for an audit that does not take precinct sizes into account.

The third type of adjustable audit is a *vote-dependent audit*. Here the selection of precincts to be audited depends not only on the reported margin of victory M and the precinct sizes $\{v_i\}$, but also on the reported vote counts $\{r_{ij}\}$. A vote-dependent audit can reflect the intuition that if precinct A reports more votes for candidate 1 (the reported winner) than precinct B reports, then precinct A should perhaps be audited with higher probability, since it may have experienced a larger amount of fraud. See Section 10; also see Calandrino et al. [4].

The fourth type of adjustable audit is a *history-dependent audit*. Here the selection of precincts to be audited depends not only on the reported margin of victory M, the precinct sizes $\{v_i\}$, and the reported vote counts $\{r_{ij}\}$, but also on records of similar data for previous elections. A precinct whose reported vote counts differ greatly from previous similar elections becomes more likely to be audited.

We assume precincts are selected for auditing according to some adjustable audit procedure of one of the above types, not including history-dependent audits.

We ignore other very interesting approaches for choosing sample of precincts to audit, such as letting the runner-up choose some of them [1].

In this paper we consider what we call an *error-bound-dependent* audit. In such an audit the audit computes for each precinct P_i an *error bound* e_i on the error (change in margin of victory) that the adversary could have made in that precinct.

An error-dependent-dependent audit is a special case of a size-dependent audit, if the error bound for precinct P_i depends only the size v_i of the precinct, as in the Linear Error Bound Assumption of equation (6) where the error bound is simply proportional to the precinct size.

The linear error bound assumption leads, for example, to sampling strategies of the form "probability proportional size," as we shall see, since our "probability proportional to error bound" strategy becomes "probability proportional to size" when "error bound is proportional to size."

However, the error-dependent audit could be a special case of a vote-dependent audit, if the error bound e_i depends on the votes cast in precinct P_i. We explore this possibility in Section 10.

In any case, it is useful to formally "decouple" the error bound from the precinct size.

We let

$$E = \sum_i e_i$$

denote the sum of these error bounds; this is an upper bound on the total amount of error the adversary could have introduced into the margin of victory.

4.2 High-level structure of an audit

The post-election audit involves the following steps. We assume that the type of audit involved has been predetermined (e.g. by state law).

1. Determine the relevant parameters of the election (margin of victory M, precinct sizes $\{v_i\}$, reported vote counts $\{r_{ij}\}$, and error bounds $\{e_i\}$).

2. Select a sample $\mathcal{S}$ of precints to be audited.

3. Count by hand all the paper ballots for every precinct in the sample. If precinct i is audited, then the actual vote counts a_{ij} and the errors e_i^* become known to the auditor. If $e_i^* = 0$ then precinct i is deemed to be *good* (i.e. uncorrrupted); otherwise (if $e_i^* > 0$) precinct i is detected as being *bad* (i.e. corrupted).

4. If no errors are found in any of the precincts audited, announce that candidate 1 (the reported winner of the electronic totals) is the winner of the election. Otherwise, trigger some enlarged examination (escalate the audit).

We don't discuss triggers and escalation in this paper, although such discussion is very important and needs to be included in any complete treatment of post-election auditing (see Stark [18]). We also don't consider stratified sampling (e.g. choosing at least one precinct per county), or "challenge" sampling by runners-up or losers.

4.3 Selecting a sample

How should the auditor select precincts to audit?

The auditor wishes to maximize the *probability of detection*: the probability that the auditor audits at least one bad precinct (with nonzero error e_i^*), if there is sufficient error to have changed the election outcome.

The auditor's method should be *randomized*, as is usual in game theory; this unpredictability prevents the adversary from knowing in advance which precincts will be audited.

A restatement of a typical auditing problem into combinatorial terms may be helpful:

> Suppose there are n boxes; each is labelled with two integers: e_i and v_i. The adversary has placed blue marbles in some of the boxes, so that at most e_i blue marbles are in box i, for $1 \le i \le n$. The auditor wishes to draw a sample of precincts to open so that:
>
> - If the adversary has placed a total of M or more marbles in the boxes, then the auditor has a chance of at least $1 - \alpha$ of finding at least one blue marble in one of the sampled boxes, where α is a prespecified significance level parameter.
> - The expected value of the sum of the v_i's for the sampled boxes is minimized.

We first review auditing procedures to use when all precincts have the same size. We then proceed to discuss the case of interest in this paper, when precincts have a variety of sizes.

5 Equal-sized precincts

This section briefly reviews the situation when all n of the precincts have the same size v (so $V = nv$). We adopt the Linear Error Bound Assumption ($e_i \le 2sv_i$) of equation (6) in this section.

Let b denote the number of precincts that have been corrupted.

Since an adversary who changed the election outcome must have introduced sufficient error,

$$2bsv \geq M \ ,$$

so that (see Dopp et al. [8])

$$b = M/2sv$$

is the minimum number of precincts the adversary could have corrupted.

When all precincts have the same size, the auditor should pick an appropriate number u of distinct precincts uniformly at random to audit. See Neff [12], Saltman [16], or Aslam et al. [2] for discussion and procedures for calculating appropriate audit sample sizes.

The probability of detecting at least one corrupted precinct in a sample of size u is

$$1 - \frac{\binom{n-b}{u}}{\binom{n}{u}} \ .$$

By choosing u so that

$$u \geq (n - (b-1)/2)(1 - \alpha^{1/b}) \qquad (7)$$

one has a test at significance level α (or equivalently, at "confidence level" $c = 1 - \alpha$): the probability is at least $c = 1 - \alpha$ that at least one corrupted precinct will be detected, if there are at least b corrupted precincts (See Aslam et al. [2].)

Rivest [15] suggests that equation (7) can be crudely, but usefully, approximated by the following "Rule of Thumb":

$$u \geq 1/m \ ;$$

one over the (fractional) margin of victory $m = M/V$. For equal-sized precincts, and assuming $s = 0.20$, this gives remarkably good results, corresponding to a confidence level of at least $c = 92\%$; such formula can provide useful "back-of-the-envelope" guidance for sample sizes when all precincts have approximately the same size.

6 The SAFE auditing method

The "SAFE" auditing method by McCarthy et al. [11] is perhaps the best-known approach to auditing elections; it adapts the standard approach for handling equal-sized precincts, discussed above, to handle variable-sized precincts.

In 2006 Stanislevic [17] presented a conservative way of handling precincts of different sizes; this approach was also developed independently by Dopp et al. [8]. This method is the basis for the SAFE auditing procedure.

The main idea is to assume that the adversary corrupts the larger precincts first. This enables one to derive a lower bound on the number b_{min} of precincts that must have been corrupted if the election outcome was changed. The auditor can then use b_{min} in an approach for auditing that samples precincts uniformly. More precisely, the auditor knows that the adversary, if he changed the election outcome, must have corrupted at least b_{min} precincts, where b_{min} is the least integer such that

$$2s \sum_{1 \leq i \leq b_{min}} v_i \geq M \ .$$

(Recall our assumption that $v_1 \geq v_2 \geq \cdots v_n$.)

Then the auditor draws a sample of size u precincts uniformly, where u satisfies equation (7); this guarantees that the probability is at least $1-\alpha$ that a precinct with nonzero error will appear in the sample, if the adversary has introduced enough error to have changed the election outcome.

7 Basic auditing methods

In this section we review "basic" auditing methods.

If the auditor adopts what we call a "basic" method, then each precinct is audited *independently* with a probability determined by the auditor. While this represents some restriction on the flexibility of the auditor, a large class of interesting auditing procedures are basic auditing procedures. We try restricting attention to "basic" methods in an effort to make some of the math easier; although we shall see in Section 9 that the math is actually fairly simple for some non-basic methods.

We thus assume in this section that the auditor will audit each precinct P_i independently with some probability p_i, where each p_i satisfies $0 \leq p_i \leq 1$. Thus, the auditor's auditing method is completely determined by the vector $\mathbf{p} = (p_1, p_2, \ldots, p_n)$.

The probabilities p_i have a sum equal to the expected number of precincts audited. (Note that they do not normally sum to 1.)

The expected workload for the auditor (in terms of the expected number of *votes* to be counted) is

$$v(\mathbf{p}) = \sum_i p_i v_i \ . \qquad (8)$$

We assume that vectors $\mathbf{p} = (p_1, p_2, \ldots, p_n)$, $\mathbf{v} = (v_1, v_2, \ldots, v_n)$, and $\mathbf{e} = (e_1, e_2, \ldots, e_n)$, are public knowledge and known to everyone, including the adversary.

With a basic auditing procedure, the chance that a precinct is audited is independent of the error introduced into that precinct by the adversary. Thus, we can assume that the adversary makes the maximum change possible in each corrupted precinct: $e_i^* = e_i$. This helps the adversary reduce the number of precincts corrupted and thus reduces the chance of him being caught during an audit.

A basic auditing method is not difficult to implement in practice in an open an transparent way:

- A table is printed giving for each precinct P_i its corresponding probability p_i of being audited.
- For each precinct p_i, four ten-sided dice are rolled to give a four-digit decimal number $x_i = 0.d_1d_2d_3d_4$. Here d_j is the digit from the j-th dice roll. If $x_i < p_i$, then precinct P_i is audited; otherwise it is not. The probability table and a video-tape of the dice-rolling are made publicly available. See Cordero et al. [6] for more discussion on the effective use of dice.

One very nice aspect of basic auditing methods is that we can easily compute the *exact* significance level for $\mathbf{p}$. Given $\mathbf{p}$, one can use a dynamic programming algorithm to compute the probability of detecting an adversary who changes the margin by M votes or more. This algorithm, and applications of it to heuristically compute *optimal* basic auditing strategies, are given by Rivest [14].

8 Negative-exponential auditing method (NEGEXP)

This Section presents the "negative exponential" auditing method NEGEXP, which appears to have near-optimal efficiency. While it is easy to use in practice, the PPEBWR method of the next Section may nonetheless be a slightly better practical choice. We present NEGEXP anyway, since it is quite simple and elegant.

The "negative-exponential" auditing method (NEGEXP)† is a heuristic basic auditing method. Intuitively, the probability that a precinct is audited is a one minus a negative exponential function of the error bound for a precinct. See Figure 1.

The "value" to the adversary of corrupting precinct i is assumed to be e_i, the known upper bound on the amount of error (in the margin of victory) that can be introduced in precinct i. In a typical situation e_i might be proportional to v_i; this is the Linear Error Bound Assumption.

Intuitively, the auditor wants to make the adversary's risk of detection grow with the "value" a precinct has to the adversary; this motivates the adversary to leave untouched those precincts with large error bounds. The adversary thus ends up having to corrupt a larger number of smaller precincts, which increases his chance of being caught in a random sample.

The motivation for the NEGEXP method is the following strategy for the auditor: determine auditing probabilities so that the chance of auditing at least one of a corrupted set of precincts depends only on the total error bound of that set of precincts. For example, the adversary will then be indifferent between corrupting a single precinct with error bound $e_\ell = (e_i + e_j)$ or corrupting two precincts with respective error bounds e_i and e_j. The chance of being caught on P_ℓ or being caught on at least one of P_i and P_j should be the same.

†An earlier note [14] by one of the authors called this method the "logistic method." That seems a misnomer, so we have adopted the more accurate term "negative exponential method" instead.

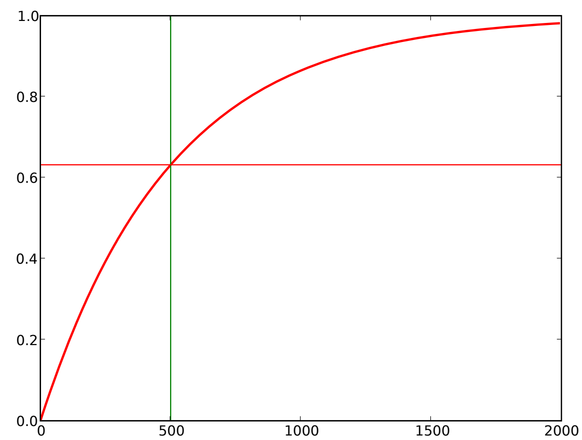


Figure 1: The negative exponential function $p_i = 1 - \exp(-e_i/w)$ for $w = 500$. The horizontal axis is the error bound e_i; the vertical axis gives the probability $p_i =$ of being audited. Here w is a positive parameter that may be set arbitrarily to achieve a given overall confidence level for the audit. Precincts with error bounds larger than w have at least a 63% chance of being audited.

This implies that the auditor should *not* audit each P_i with probability $q_i = 1 - p_i$, where

$$q_i = \exp(-e_i/w), \tag{9}$$

and where w is some fixed constant. Thus

$$q_\ell = q_i q_j$$

as desired if $e_\ell = e_i + e_j$.

This is the same as saying that q_i^{1/e_i} is constant.

Our NEGEXP auditing method thus yields

$$p_i = 1 - \exp(-e_i/w)\ ; \tag{10}$$

see Figure 1. The name "negative exponential" refers to the negative exponential appearing in this formula.

With the NEGEXP method, as the error bound e_i increases, the probability of auditing P_i increases, starting off at 0 for $e_i = 0$ and increasing as e_i increases, and levelling off approaching 1 asymptotically for large e_i. The chance of auditing P_i passes $(1 - 1/e) \approx 63\%$ as e_i exceeds w.

The value w can be thought of as approximating a "threshold" value: precincts with e_i larger than w have a fairly high probability of being audited, while those smaller than w have a smaller chance of being audited.

As w decreases, the auditing gets more stringent: more precincts are likely to be audited. As w increases, auditing becomes less stringent: fewer precincts are likely to be audited.

An auditor may choose to use the NEGEXP auditing method of equation (10), and choose w to achieve an audit with a given significance level.

The design of the NEGEXP auditing method makes this easy, since an important and very convenient property of the NEGEXP audit is that for any set Q of precincts that the adversary may choose to corrupt satisfying

$$\sum_{i \in Q} e_i \geq M \ ,$$

the chance of detection is at least

$$1 - \prod_{i \in Q} \exp(-e_i/w) \geq 1 - \exp(-M/w) \ . \tag{11}$$

This holds no matter what approach the adversary uses.

In particular, we note that if the adversary can not find a set of precincts whose error bounds total exactly M votes, then he will choose a set of precincts with total error bound M' where M' is somewhat larger than M, but then the detection probability also becomes larger, since

$$1 - \exp(-M'/w) > 1 - \exp(-M/w) \ .$$

How can an auditor audit enough to achieve a given significance level?

The relationship of equation (11) gives a very nice way for the auditor to choose w: by choosing

$$w = \frac{M}{-\ln(\alpha)} \tag{12}$$

the auditor achieves a test with significance at least α: there is probability at least $1 - \alpha$ in catching error at least M, no matter what the adversary's approach is. For example, by choosing $w \approx M/3$, the auditor tests at signficance level 5% for margin-shift error of size M or greater.

If we use equation (12) to determine w, then we have

$$p_i = 1 - \alpha^{e_i/M} \ . \tag{13}$$

With the Linear Error Bound Assumption, this becomes

$$p_i = 1 - \alpha^{2sv_i/M} \ . \tag{14}$$

If the auditor's goal is to achieve a given expected number of precincts audited or a given expected number of votes counted, he can use any of several standard packages for root-finding to find a value of w that meets the given constraints.[‡]

In any case, it is easy to print out a table of the probabilities p_i for each precinct, so that one can utilize a suitable dice-based protocol for actually picking the precincts to be audited.

We note that As the amount M of corruption being sought becomes smaller and smaller, the NEGEXP method may devolve into an approach of picking precincts uniformly. This will happen if the auditor (as seems reasonable) performs an initial "trimming" of the e_i's to ensure that none are larger than M. This operation makes all of the e_i's equal to each other and to M in the limit as M decreases, giving the uniform auditing method.

We also note that if $e_i = \text{const} * v_i$,

$$p_i = 1 - \exp(-e_i/w) \approx e_i/w \approx \text{const} * v_i/w$$

when e_i is small relative to w, so that the NEGEXP method can be viewed as an approximation to a method whereby precincts are selected with probability proportional to their size ("pps"). The next Section proves a relationship between NEGEXP and the PPSWR efficiencies.

This completes our description of the NEGEXP auditing method. Some experimental results can be found in Section 11.

In the next Section we describe a different method, which turns out to be nearly identical (but slightly better) in efficiency to the NEGEXP method, and which may be slightly easier to work with as well.

9 Sampling with Probability Proportional to Error Bound, with Replacement (PPEBWR)

In this section we present the "PPEBWR" (*sampling with probability proportional to error bound, with replacement*) auditing strategy. The PPEBWR method is simple to implement. We show that it does at least as well as the NEGEXP method. Indeed, we believe that PPEBWR is an excellent method in many respects, and recommend its use in practice.

Consider auditing an election with non-uniform error bounds $\mathbf{e} = (e_1, e_2, \ldots, e_n)$ where $E = \sum_i e_i$.

Let M be the (minimum) level of error one wishes to detect; M is the margin of victory.

Consider the following sampling-with-replacement procedure. Form a sampling distribution $\mathbf{p}$ over the precincts by normalizing the precinct error-bound vector $\mathbf{e}$, i.e.,

$$\mathbf{p} = (e_1/E, e_2/E, \ldots, e_n/E), \tag{15}$$

[‡]In our experiments, we used the routine `brentq` from the Python library `scipy.optimize`.

and draw t samples at random with replacement according to $\mathbf{p}$. Eliminate duplicates, and audit the set of precincts obtained.

It is easy to use dice to select the precincts to be audited in a public and transparent manner. The probabilities $p_i = e_i/E$ of equation (15) can be computed, and then their cumulative values computed:

$$\hat{p_i} = \sum_{1 \le j \le i} p_j$$

and printed out. For each of t rounds, four decimal dice are rolled, and the four digits d_1, d_2, d_3, and d_4 combined to yield a four-digit decimal number $x = 0.d_1d_2d_3d_4$. Then P_i is marked for auditing if

$$\hat{p}_{i-1} \le x < \hat{p}_i \ .$$

The printed tables and a videotape of the dice-rolling are made publicly available. This approach only requires rolling t random numbers, whereas the basic methods of Sections 7–8 require rolling n random numbers.

When the Linear Error Bound Assumption holds, the PPEBWR method performs *sampling with probability proportional to size* within each round. We call the overall method *sampling with probability proportional to size, with replacement*, or "PPSWR".

The use of sampling with probability proportional to size (PPS) is well-known in the statistics and survey-sampling literature as a means of reducing the variance of estimates. It was developed by by Hansen and Hurwitz [10]; Cochran [5, Ch. 9A] provides an overview. However, the current paper is one of the first to propose and analyze the use of PPS in an adversarial setting. (Stark [18, Sec. 4.2.1] also provides some discussion of the use of PPS sampling for election audits in a manner that is similar to ours, but for a different purpose (use in stratified sampling).)

We introduce notation to distinguish the per-round selection probabilities (denoted by p_i) from the overall selection probabilities (denoted by π_i); these are related via the number t of selection rounds and the equation:

$$\pi_i = 1 - (1 - p_i)^t \ ; \tag{16}$$

precinct i is audited only if it is not missed during each of the t selection rounds.

It is worth noting that while the per-round probabilities p_i are proportional to size, the overall probabilities π_i are generally not. To see this, note for example that as t gets large the overall probability of selection of each precinct approaches 1. Actually, the overall probabilities π_i turn out to be nearly identical (but slightly less) than those computed by the NEGEXP method, as we shall see.

We now see how to adjust the number t of rounds to give a desired significance level α for the audit.

Any set of precincts whose total error bound is at least M will have probability weight at least M/E. The probability that these precincts all escape detection is at most

$$(1 - M/E)^t .$$

We want this to be at most α for some desired significance level α, and solving

$$(1 - M/E)^t \le \alpha$$

for t, we obtain that

$$t \ge \frac{\ln(\alpha)}{\ln(1 - M/E)}$$

is sufficient. Let

$$t_* = \frac{\ln(\alpha)}{\ln(1 - M/E)}$$

be the minimum sample size.

We now show that the probability π_i with which any given precinct P_i is audited is approximately the negative-exponential audit probability. Using the fact that

$$t_* = \frac{\ln(\alpha)}{\ln(1 - M/E)} \approx (E/M) \cdot (-\ln(\alpha))$$

we obtain

$$\begin{aligned} \pi_i &= 1 - (1 - e_i/E)^{t_*} \\ &\approx 1 - (1 - e_i/E)^{(E/M)\cdot(-\ln(\alpha))} \\ &= 1 - \left[(1 - e_i/E)^E\right]^{(-\ln(\alpha))/M} \\ &\approx 1 - \left[e^{-e_i}\right]^{(-\ln(\alpha))/M} \\ &= 1 - e^{-e_i/w} \end{aligned}$$

where $w = M/(-\ln(\alpha))$.

Careful Analysis: A more careful analysis shows that PPEBWR is somewhat more efficient than NEGEXP, in that the precincts are all audited with somewhat smaller probabilities (as compared to the NEGEXP audit) while still achieving the desired confidence. This is possible since the PPEBWR strategy is not a *basic* random audit: the precincts are not audited independently, and if one precinct is "missed" in the audit, this fact increases the probability that other precincts will be audited.

We now show that using a with-replacement sample size of

$$t_* = \frac{\ln(\alpha)}{\ln(1 - M/E)} = \log_{(1-M/E)}(\alpha)$$

yields auditing probabilities somewhat smaller than the NEGEXP audit. Using the fact that for all $x \ge 1$,

$(1-1/x)^x$ increases monotonically, approaching $1/e$ from below, we proceed as follows:[§]

$$\begin{aligned}
1-\pi_i &= (1-e_i/E)^{t_*} \\
&= \Big(1-e_i/E\Big)^{\log_{(1-M/E)}\alpha} \\
&= \left[\left(1-\frac{1}{E/e_i}\right)^{\frac{E}{e_i}}\right]^{\frac{e_i}{E}\cdot\log_{(1-M/E)}\alpha} \\
&\geq \left[\left(1-\frac{1}{E/M}\right)^{\frac{E}{M}}\right]^{\frac{e_i}{E}\cdot\log_{(1-M/E)}\alpha} \\
&= \Big(1-M/E\Big)^{\frac{e_i}{M}\cdot\log_{(1-M/E)}\alpha} \\
&= \alpha^{e_i/M} \\
&= e^{-e_i/w}
\end{aligned}$$

where $w = M/(-\ln(\alpha))$. Thus,

$$\pi_i \leq 1 - e^{-e_i/w}.$$

Note that for all $e_i < M$, this inequality is strict.

Conjecture 1 *The result just proven also works the other way: the cost (in terms of votes counted) of the* PPEBWR *strategy is not more than the cost of the* NEGEXP *strategy, times* $1 + o(1)$*.*

The costs of the PPEBWR strategy are easy to compute.

$$\sum_i \pi_i$$

and the expected number of votes audited is

$$\sum_i v_i\pi_i \ ,$$

where

$$\pi_i = 1-(1-e_i/E)^{t_*} \ .$$

Conjecture 2 *The* PPEBWR *strategy can be proven to have efficiency nearly equal to that of the optimal auditing strategy for the Linear Error Bound Assumption (perhaps under some mild restrictions on the error bounds).*

10 Vote-dependent auditing

In this section we drop the assumption that error bounds are proportional to precinct size. That is, we drop the assumption

$$e_i = 2sv_i \ .$$

[§]We assume that $e_i \leq M$ for all i; precincts whose error bounds are greater than M can be effectively "capped" at M with the sampling vector re-normalized as appropriate.

How else can the auditor obtain a bound on the error?

Instead of having a size-dependent audit, he may have a vote-dependent audit.

He can use the reported vote totals to help, since

$$e_i^* \leq e_i$$

if

$$e_i = 2r_{i1} + \sum_{j>2} r_{ij} = v_i + r_{i1} - r_{i2} \ ;$$

here we are measuring the margin of victory between candidate 1 and candidate 2.

If we aren't sure who the "runner-up" should be, we could take the maximum bound over any such "runner-up," giving us:

$$e_i = v_i + r_{i1} - \min_j r_{ij} \ .$$

These bounds e_i are going to be larger than those obtained via a within-shift bound $2sv_i$ in most cases; giving worse results. However, in a two-candidate race if a precinct votes almost entirely for the electronic runner-up, the new bound may be smaller (indeed, if $r_{i1} = 0$, no auditing of that precinct may be needed).

Stark [18, Section 3.1] introduces the nifty notion of "pooling" several obviously losing candidates to create an obviously losing "pseudo-candidate" to reduce the error bounds in his approach; this can also be applied here to good effect.

11 Experimental Results

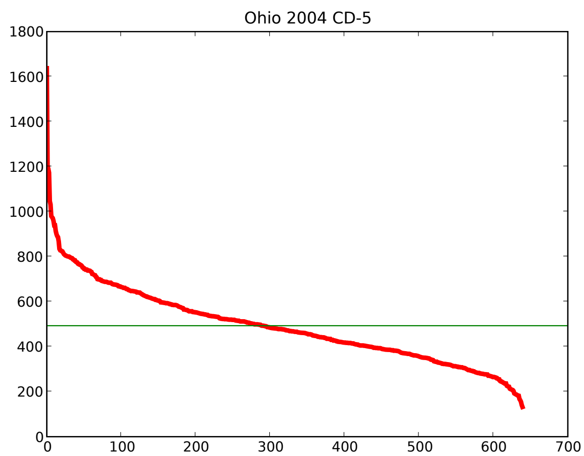
We illustrate and compare the previously described methods for handling variable-sized precincts using data from Ohio and Minnesota. These results illustrate that taking precinct size into account (e.g. by using NEGEXP or PPSWR) can result in *dramatic* reductions in auditing cost, compared to methods (such as SAFE) that do not.

11.1 Ohio 2004 CD-5

Mark Lindeman kindly supplied a dataset of precinct vote counts (sizes) for the Ohio congressional district 5 race (OH-05) in 2004. A total of $V = 315540$ votes were cast in 640 precincts, whose sizes ranged from 1637 (largest) to 132 (smallest), a difference by a factor of more than 12. See Figure 2.

Let us assume an election with a margin of victory of $m = 1\%$ (i.e. $M = 0.01V = 3155$). Assume that the adversary will change at most $s = 20\%$ of the votes in a precinct. Assume we wish to audit to a confidence level of 92% ($\alpha = 0.08$).

If the precincts were equal-sized, the Rule of Thumb [15] would suggest auditing $1/m = 100$ precincts.



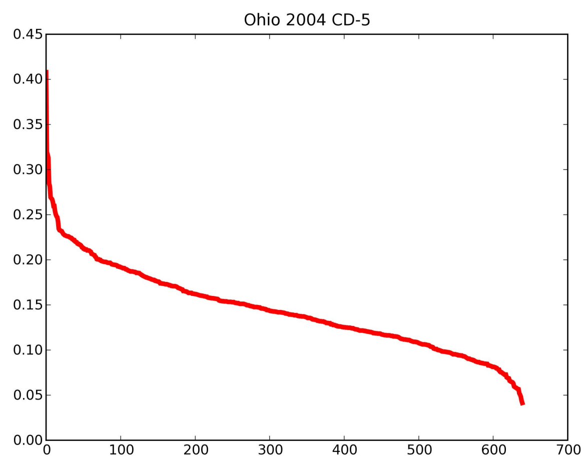


Figure 2: The upper graph shows the distribution of 640 precinct sizes for Ohio 2004 Congressional District 5. A total of 315,540 votes were cast. The maximum precinct size was 1637, the average was 493, and the minimum was 132. The lower graph shows the probability distribution for picking precincts in this example, using the NEGEXP method.

The more accurate APR formula (7) suggests auditing 93 precincts (here $b = M/2sv = 16$ precincts). The expected workload would be 45852 votes counted. But the precincts are quite far from being equal-sized.

If we sample 93 precincts uniformly (using the APR recommendation inappropriately here, since the precincts are variable-sized), we now only achieve a 67% confidence of detecting at least one corrupted precinct, when the adversary has changed enough votes to change the election outcome. The reason is that all of the corruption can fit in the 7 largest precincts now.

The SAFE auditing method [11] would determine that $b_{min} = 7$ (reduced from $b = 16$ for the uniform case, since now the adversary need only corrupt the 7 largest precincts to change the election outcome). Using a uniform sampling procedure to have at least a 92% chance of picking one of those 7 precincts (or any corrupted precinct) requires a sample size of 193 precincts (chosen uniformly), and an expected workload of 95,155 votes to recount.

With the NEGEXP method, larger precincts are sampled with greater probability. The adversary is thus forced to disperse his corruption more broadly, and thus needs to use more precincts, which makes detecting the corruption easier for the auditor. The NEGEXP method computes $w = -M/\ln(\alpha) = 1249$, and audits a precinct of size v_i with probability $p_i = 1 - \exp(-0.4v_i/w)$. The largest precinct is audited with probability 0.408, while the smallest is audited with probability 0.041. The expected number of precincts selected for auditing is only 92.6, and the expected workload is only 50,937 votes counted. (It is perhaps surprising how close this is to the 45,852 that would have be counted on the average if the precincts had had equal sizes!)

It is interesting that when we tried our iterative improvement method to try to find a better basic auditing strategy, we did not find any improvement; it may be that for this example the NEGEXP method is basic-optimal or extremely close to it.

The PPEBWR method gave results almost identical to those for the NEGEXP method. The expected number of distinct precincts sampled was 91.6, and the expected workload was 50402 votes counted. The PPEBWR method sampled each precinct with a probability that was within 0.0031 of the corresponding probability for the NEGEXP method.

We see that for this example the NEGEXP method (or the PPEBWR method) is approximately *twice* as efficient (in terms of votes counted) as the SAFE method, for the same confidence level.

The python program for the experiments in this paper is available at `http://people.csail.mit.edu/rivest/pps/varsize.py`; The datasets are also available, at `http://people.csail.mit.edu/rivest/pps/oh5votesonly.txt` (Ohio) and `http://people.csail.mit.edu/rivest/pps/MN_Gov_2006-2.csv`

(Minnesota).

11.2 Minnesota 2006 Governor's Race

Data for the 2006 Minnesota Governor's race were kindly provided by Mark Halvorson and Vittorio Addona.

There were 4123 precincts with a total of 2,202,937 votes for governor. Nine of those precincts had no votes, so there really were only 4114 precincts to work with. The margin of victory was $m = 0.0096$ (just under 1%). See Figure 3.

The Rule of Thumb suggests auditing 105 precincts. If we choose 105 precincts to audit uniformly at random, we would have an expected workload of 56,225 votes to count.

The more accurate formula (7) for the equal-sized precinct case suggests auditing 103 precincts (here $b = M/2sv = 99$ precincts), with an expected workload of 55,154 votes. But the precincts are not equal-sized here.

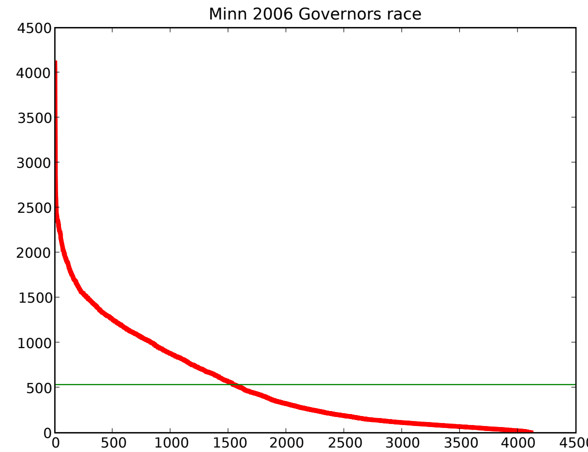
The SAFE method would work as follows. The minimum number of precincts that could hold all of the discrepancy needed to change the election outcome) was only $b_{min} = 19$ precincts. Using uniform sampling and a confidence level of 92%, one needs to sample 511 precincts to be sure (at the 92% level) of seeing at least one of 19 discrepant precincts. The expected number of votes to be audited is 273,627 votes.

The NEGEXP method for a confidence level 92% audits a precinct of size v_i with probability $p_i = 1 - \exp(-2sv_i/w)$, where $w = 8373$. The largest precinct is audited with probability 0.1783 and the smallest with probability 0.00005. The expected number of precincts audited is 102.53 and the expected number of votes counted is 111,743.

Again, the PPEBWR gives results practically identical to those of the NEGEXP method. A number $t = 104$ precincts are drawn (with replacement), with probability proportional to size (and error bound). The largest precinct audited with (overall, not per-round) probability 0.1765 and the smallest with probability 0.00005. The expected number of precincts audited is 101.39 and the expected number of votes counted is 110,507.

It appears that the SAFE method is actually *very* inefficient in this example as well, by a factor of about 2.47 in the number of votes that need to be counted, compared to the NEGEXP method (or the PPEBWR method) for the same confidence level.

The SAFE method may often be a poor choice when there are variable-precinct-sizes, particularly when there are a few very large precincts. One really needs a method that is tuned to variable-sized precincts by using variable auditing probabilities, rather than a method that uses uniform sampling probabilities.



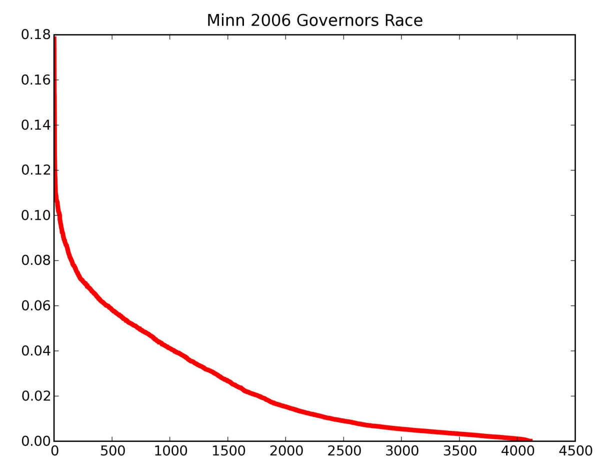


Figure 3: The upper graph shows the precinct size distribution for the Minnesota 2006 Governor's race. The 4123 precinct sizes ranged from 4110 (largest) down to 1 (smallest). There were five precincts having only 1 vote each. The average (nonzero) precinct size was 2202937/4114 = 535.47 votes. The lower graph shows the probability distribution for picking precincts in this example, using the NEGEXP method.

12 Optimal auditing method

In this section, we present the optimal auditing method. This method is theoretically interesting, but not very useful in practice, since computing the optimal auditing method seems to require time and space exponential in the number n of precincts.

The *optimal* auditing method can be represented as a probability distribution assigning a probability p_S to each *subset* S of precincts. Since there are 2^n such subsets, representing these probabilities explicitly takes space exponential in n. The probability assigned to a given subset is the probability that the auditor will choose that subset of precincts to be audited.

The optimal auditing strategy can be found with linear programming, if the number n of precincts is not too large (say a dozen at most).

The linear programming formulation requires that for each subset B of total error bound M or more votes, the sum of the probability of S, taken over all subsets S with nonnegative intersection with B, is at least $1-\alpha$.

$$(\forall B)\left[\left(\sum_{i\in B} e_i \geq M\right) \Longrightarrow \left(\sum_{S:S\cap B\neq\phi} p_S \geq 1-\alpha\right)\right]$$

In addition to these constraints, we have a constraint each p_S is nonnegative, and that

$$\sum_S p_S = 1\ .$$

Finally, the objective function to be minimized is the expected number of votes to be recounted:

$$\sum_S p_S \sum_{i\in S} v_i$$

For example, suppose we have $n = 3$ precincts A, B, C with sizes $\mathbf{v} = (60, 40, 20)$ and error bounds $\mathbf{e} = (30, 20, 10)$, an adversarial corruption target of $M = 30$ votes, and a target significance level of $\alpha = 5\%$. Then an optimal auditing strategy, when the auditor is charged on a per-vote-recounted basis, is:

$$\begin{aligned} p_\phi &= 0.013746 \\ p_A &= 0.036253 \\ p_C &= 0.036253 \\ p_{AC} &= 0.913746 \end{aligned}$$

Here ϕ denotes the empty subset; subsets not shown have zero auditing probability. The expected cost of this optimal auditing strategy is 76 votes recounted. (The above strategy also optimizes (at 1.9) the expected number of votes recounted; however, it is not always the case that the same probability distribution optimizes for both precincts counted and votes counted: a small counterexample occurs for $\mathbf{v} = \mathbf{e} = (20, 20, 10, 10)$ and $M = 30$.)

This approach is the "gold standard" for auditing with variable-sized precincts, in the sense that it definitely provides the most efficient procedure in terms of the stated optimization criterion.

However, as noted, it may yield an auditing strategy with as many as 2^n potential actions (subsets to be audited) for the auditor, and so is not efficient enough for real use, except for very small elections.

We note that it is easy to refine this approach to handle the following variations:

- An optimization criterion that is some linear combination of precincts counted and ballots counted.
- A requirement that *exactly* (or at least, or at most) a certain number of precincts be audited.

Question 1 *What is the complexity of computing an optimal strategy? While it appears to be exponential, perhaps the structure of this problem would admit a polynomial-time (or pseudo-polynomial-time) solution.*

13 Aggregation

In this section we examine the effects of *aggregation*: the merging together of two precincts.

When two precincts are merged, the resulting precinct has a size equal to the sum of the sizes of the original precincts. The total number of precincts is reduced by one.

We can view a collection of precincts of various sizes as the result of a sequence of such pairwise mergers, starting from an initial situation where each voter was in his/her own "precinct".

Theorem 1 *If one voting instance (set of precinct sizes, adversarial target, and auditors target signficance level) is the result of a merger of two precincts in an original voting instance, then the expected number of precincts that need to be audited is not increased.*

Proof: See Appendix B. ■

It is easy to devise examples where the the expected number of ballots counted in an optimal audit may increase, however.

14 Discussion and recommendations for practice

14.1 Recommendations for practice

We recommend the use of the PPEBWR method for use in an audit. It gives the most efficient audit, for a given confidence level, of the audit methods studied here (other than the optimal method, which is too inefficient for practical use).

Figure 4 summarizes the PPEBWR audit procedure recommended for use.

If the error bounds are computed using only the Linear Error Bound Assumption, so that $e_i = 2sv_i$, then the probability of picking precinct P_i is just v_i/V, so that we are picking with "probability proportional to size"—this is then the PPSWR procedure.

When the Linear Error Bound Assumption is used, one is assuming that errors larger than $2sv_i$ in a precinct will be noticed and caught "by other means"; one should ensure that this indeed happens. (Letting runners-up pick precincts to audit could be such a mechanism.)

Other considerations may result in interesting and reasonable modifications. Letting runners-up pick precincts to audit is probably helpful, although these precincts should then be ignored during the PPEBWR portion of the audit. The same suggestion holds for when a fixed number of precincts chosen randomly per county. These variations deserve more study.

The "escalation" procedure for enlarging the audit when significant discrepancies are found is (intentionally) left rather unspecified here. We recommend reading Stark [18] for guidance. At one extreme, one can perform a full recount of all votes cast. More reasonably, one can utilize a staged procedure, where the error budget α is allocated among the stages; only if enough new discrepancies are discovered in one stage does auditing proceed to the next.

14.2 Discussion

If the election is not a plurality (winner-take-all), little changes except that the notion of a "margin of victory" needs to be appropriately modified, so that the notion of a "candidate" is replaced by that of an "election outcome". (Elaboration omitted here.)

Our auditing problem is closely related to the classic notion of an "inspection game", with an "inspector" (the auditor) and an "inspectee" (the adversary). Inspection games fit within the standard framework of game theory. With optimal play, both auditor and adversary use randomized strategies. See Avenhaus et al. [3] for discussion. In our case, the adversary may choose to use a deterministic strategy¶, so inspection games have a bit more generality than we need.

It would be preferable in general, rather than having to deal with precincts of widely differing sizes, if one could somehow group the records for the larger precincts into "bins" for "pseudo-precincts" of some smaller standard size. (One can do this for say paper absentee ballots, by dividing the paper ballots into nominal standard precinct-sized batches before scanning them.) It is harder to do this if you have DRE's with wide disparities between the number of voters voting on each such machine. See Neff [12] and Wand [19] for further discussion.

It might good for legislation to to mandate an upper bound on the signficance level that must be achieved during the audit.

¶The adversary may use a deterministic strategy, since the auditor is presumed to be following a predetermined strategy that will not be influenced by other factors, such as a determination of the best counter-strategy for a possible attacker.

14.3 Open Problems

How hard is it to an compute optimal basic strategy? (See Rivest [14] for a heuristic approach.)

How hard it it to an compute optimal strategy?

How hard to compute optimal strategy for PPEBWR (i.e. optimize t and $\mathbf{p}$ to minimize audit cost while keeping detection probability for fraud at least $1-\alpha$? (Can this be done numerically?)

The aggregation theorem refers to the expected number of precincts audited. Extend the analysis to cover the expected number of votes counted. (Does it always increase?)

Show similar results to the aggregation theorem for the NEGEXP or PPEBWR strategies, instead of for the optimal strategy.

15 Conclusions

We have presented algorithms for computing the optimal auditing method, an optimal basic auditing method, a powerful "negative-exponential" (NEGEXP) method, and a "sampling with probability proportional to size, with replacement" (PPEBWR) method. In practice, the PPEBWR method seems the most efficient, usable, and promising for further development.

Acknowledgments

Thanks to Mark Lindeman for helpful discussions and the Ohio dataset of precinct sizes, and to Christos Papadimitriou for referring me to the literature on "inspection games." Thanks also to Kathy Dopp, Silvio Micali, and Howard Stanislevic for constructive suggestions. Thanks to Mark Halvorson and Vittorio Addona for supplying the Minnesota data set.

References

[1] Andrew W. Appel. Effective audit policy for voter-verified paper ballots, September 2007. Presented at 2007 Annual Meeting of American Political Science Association (Chicago, 9/1/2007). http://www.cs.princeton.edu/~appel/papers/appel-audits.pdf.

Using the recommended PPEBWR *audit procedure for variable-sized precincts.*

1. **[Gather data]** Determine n, the number of precincts, and v_i, the number of votes cast in precinct i, for $1 \leq i \leq n$. Let r_{ij} denote the number of (electronic) votes reported cast for candidate j in precinct i.

2. **[Tally electronic votes]** Let R_j denote the total number of (electronic) votes reported cast for candidate j. Let j_1 denote the candidate with the largest reported vote count, and let j_2 denote the runner-up. Determine M, the overall margin of victory in the electronic tallies:

$$M = R_{j_1} - R_{j_2} \ .$$

3. **[Choose audit parameters]** Choose a value for s, the assumed maximum within-precinct-miscount (e.g. $s = 0.20$). Choose a value for α, the significance level desired for this audit (e.g. $\alpha = 0.05$).

4. **[Compute error bounds]** For $1 \leq i \leq n$: Determine the error bound for precinct i:

$$e_i = \min(2sv_i, M, v_i + r_{ij_1} - \min_j r_{ij})$$

(It is OK just to use the first term, so that $e_i = 2sv_i$.) Also compute the total error bound:

$$E = \sum_{1 \leq i \leq n} e_i \ .$$

(Check that $M < E$; if not, you need to pick a larger value for s.)

5. **[Determine per-round selection probabilities]** Determine the per-round selection probability for each precinct:

$$p_i = e_i / E \quad \text{for } 1 \leq i \leq n \ .$$

Also determine their cumulative probabilities $\hat{p}_i$: $\hat{p}_0 = 0$ and

$$\hat{p}_i = \sum_{1 \leq j \leq i} p_j \quad \text{for } 1 \leq i \leq n \ .$$

6. **[Determine number of selection rounds]** Determine the number t of selection rounds:

$$t = \left\lceil \frac{\ln(\alpha)}{\ln(1 - M/E)} \right\rceil$$

7. **[Select precincts to be audited]** For each of t rounds, pick a precinct P_i to be audited, where P_i is picked with probability p_i, as follows:

   (a) Roll four decimal dice to obtain four decimal digits d_1, d_2, d_3, d_4; combine them to obtain a fraction $x = 0.d_1d_2d_3d_4$ (so that $0 \leq x < 1$).

   (b) Determine the unique i such that

   $$\hat{p}_{i-1} \leq x < \hat{p}_i \ .$$

   (c) Mark P_i for auditing. (If it was already so marked, it stays so marked.)

8. **[Audit selected precincts]** For each precinct P_i marked for auditing in the preceding step, hand-count its paper ballots to determine the actual number a_{ij} of (paper) votes for each candidate j.

9. **[Terminate or escalate]** If no significant discrepancies are discovered, terminate the audit and announce that no significant discrepancies were discovered. Otherwise, escalate the audit.

Figure 4: Auditing with the recommended PPEBWR method.

[2] Javed Aslam, Raluca Popa, and Ronald L. Rivest. On estimating the size and confidence of a statistical audit. In *Proceedings EVT'07*, 2007. http://www.usenix.org/events/evt07/tech/full_papers/aslam/aslam.pdf.

[3] Rudolf Avenhaus, Bernhard Von Stengel, and Shmuel Zamir. Inspection games. In R. J. Aumann and S. Hart, editors, *Handbook of Game Theory*, volume III. Elsevier, January 30 1998. http://citeseer.ist.psu.edu/212144.html.

[4] Joseph A. Calandrino, J. Alex Halderman, and Edward W. Felten. Machine-assisted election auditing. In *Proc. EVT'07*, 2007. http://www.usenix.org/event/evt07/tech/full_papers/calandrino/calandrino.pdf.

[5] William G. Cochran. *Sampling Techniques (3rd ed.)*. Wiley, 1977.

[6] Arel Cordero, David Wagner, and David Dill. The role of dice in election audits — extended abstract, June 16 2006. IAVoSS Workshop on Trustworthy Elections (WOTE 2006). http://www.cs.berkeley.edu/~daw/papers/dice-wote06.pdf.

[7] Kathy Dopp. History of confidence election auditing development (1975 to 2007) and overview of election auditing fundamentals, 2007. http://electionarchive.org/ucvAnalysis/US/paper-audits/History-of-Election-Auditing-Development.pdf.

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

[9] ElectionLine.org. Case study: Auditing the vote, March 2007. http://electionline.org/Portals/1/Publications/EB17.pdf.

[10] M. H. Hansen and W. N. Hurwitz. On the theory of sampling from finite populations. *Ann. Math. Stat.*, 14:333–362, 1943.

[11] John McCarthy, Howard Stanislevic, Mark Lindeman, Arlene Ash, Vittorio Addona, and Mary Batcher. Percentage-based versus SAFE vote tabulation auditing: a graphic comparison, October 2007. http://www.verifiedvotingfoundation.org/auditcomparison.

[12] C. Andrew Neff. Election confidence—a comparison of methodologies and their relative effectiveness at achieving it (revision 6), December 17, 2003. http://www.votehere.com/papers/ElectionConfidence.pdf.

[13] Lawrence Norden, Aaron Burstein, Joseph Lorenzo Hall, and Margaret Chen. Post-election audits: Restoring trust in elections, 2007. Brennan Center for Justice at New York University [School of Law and Samuelson Law, Technology and Public Policy Clinic at UC Berkeley School of Law], http://www.brennancenter.org/dynamic/subpages/download_file_50089.pdf.

[14] Ronald L. Rivest. On auditing elections when precincts have different sizes, 2007. http://people.csail.mit.edu/rivest/OnAuditingElectionsWhenPrecinctsHaveDifferentSizes.pdf .

[15] Ronald L. Rivest. A simple rule of thumb for election audit size determination, 2007. http://people.csail.mit.edu/rivest/Rivest-ASimpleRuleOfThumbForElectionAuditSizeDetermination.pdf.

[16] Roy G. Saltman. Effective use of computing technology in vote-tallying. Technical Report NBSIR 75–687, National Bureau of Standards (Information Technology Division), March 1975. http://csrc.nist.gov/publications/nistpubs/NBS_SP_500-30.pdf.

[17] Howard Stanislevic. Random auditing of e-voting systems: How much is enough?, revision August 16, 2006. http://www.votetrustusa.org/pdfs/VTTF/EVEPAuditing.pdf.

[18] Philip B. Stark. Conservative statistical post-election audits, Nov 15 2007. http://statistics.berkeley.edu/~stark/Preprints/conservativeElectionAudits07.pdf.

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

Appendix A. Aggregation Theorem Proof

Proof:

Let $\mathbf{I}_{orig}$ and $\mathbf{I}_{new}$ be the original and new voting instances, respectively. Let a and b be two precincts from $\mathbf{I}_{orig}$ that are merged in $\mathbf{I}_{new}$ to become (ab). Now, let $\mathcal{N}$ be the set of sets of precincts from $\mathbf{I}_{orig}$ that contain 'n'either a or b and $\mathcal{M}$ be the set of sets of precincts from $\mathbf{I}_{new}$ that contain the 'm'erged precinct (ab). The proof proceeds by finding a new set of probabilities for $\mathbf{I}_{new}$(denoted by p_i^*) that satisfy the constraints in the optimal strategy definition, yet provide a smaller expected number of precincts to audit. This would be sufficient to prove our theorem because the new minimum expected number of precincts to audit can only be smaller than the one determined by our chosen probabilities. Define

$$p_S^* = \begin{cases} p_S, & \text{if } S \in \mathcal{N} \\ p_{T\cup\{a\}} + p_{T\cup\{b\}} + p_{T\cup\{a,b\}}, & \text{otherwise} \end{cases}$$

where $T = S - \{(ab)\}$. Basically, the probabilities of the sets that contained at least a and b accumulate on the new sets that contain the merged precinct (ab). We now show that these probabilities satisfy all the conditions of the optimal auditing procedure.

$$\begin{aligned} \sum_S p_S^* &= \sum_{S\in\mathcal{N}} p_S^* + p^*_{S\cup\{(ab)\}} \\ &= \sum_{S\in\mathcal{N}} p_S + p_{S\cup\{a\}} + p_{S\cup\{b\}} + p_{S\cup\{a,b\}} \\ &= \sum_{S\in\mathcal{N}} p_S = 1 \end{aligned}$$

Thus, the probabilities still add up to one. Let B be a set of precincts in $\mathbf{I}_{new}$ such that

$$\sum_{i\in B} v_i \geq C.$$

Then, if S is a set of precincts in $\mathbf{I}_{new}$ such that $S\cap B \neq \phi$ and $T = S - \{(ab)\}$ we have

$$\begin{aligned} \sum_S p_S^* &= \sum_{S:S\in\mathcal{N}} p_S^* + \sum_{S:S\in\mathcal{M}} p_S^* \\ &= \sum_{S:S\in\mathcal{N}} p_S + \sum_{S:S\in\mathcal{M}} p_{T\cup\{a\}} + p_{T\cup\{b\}} + p_{T\cup\{a,b\}} \end{aligned}$$

Let S' and B' be two sets in $\mathbf{I}_{orig}$ such that $B' = B - \{(ab)\} \cup \{a,b\}$ and $S' \cap B' \neq 0$. Then the size (number of votes) of B' is equal to the one of B and thus at least M. Since the constraints in the optimal procedure are satisfied by the parameters of $\mathbf{I}_{orig}$, we have that $\sum_{S'} p_{S'} \geq 1-\alpha$. However, we can see that this is equal to the last sum above, which shows that our new probabilities satisfy the second constraint. Finally, we need to show that the expected number of precincts to audit does not increase with this new set of probabilities:

$$\begin{aligned} \sum_S p_S^* |S| &= \sum_{S:S\in\mathcal{N}} p_S^* |S| + p^*_{S\cup\{(ab)\}}(|S|+1) \\ &= \sum_{S:S\in\mathcal{N}} p_S |S| + (p_{S\cup\{a\}} + p_{S\cup\{b\}})(|S|+1) \\ &+ \ p_{S\cup\{a,b\}}(|S|+1) \\ &\leq \sum_{S:S\in\mathcal{N}} p_S |S| + (p_{S\cup\{a\}} + p_{S\cup\{b\}})(|S|+1) \\ &+ \ p_{S\cup\{a,b\}}(|S|+2) \end{aligned}$$

We can see that the last summation equals the expected number of precincts to audit in the original instance, which completes our proof.

Notation

n	number of precincts
P_i	precinct i
v_i	votes cast in precinct i
V	total number of votes cast
α	significance goal (e.g. $\alpha = 0.05$)
c	confidence level $= 1-\alpha$ (e.g. $c = 0.95$)
k	number of candidates in the races (including "invalid" and "undervote" as candidates)
r_{ij}	reported votes cast in P_i for candidate j
R_j	total number actual votes cast for candidate j
a_{ij}	actual votes cast in P_i for candidate j
A_j	total number actual votes cast for candidate j
$M^{(a)}$	actual margin of victory
$M^{(r)}$	reported margin of victory (votes)
M	shorthand for $M^{(r)}$
m	margin of victory (fraction of votes)
e_i^*	error in margin from precinct i
E^*	total error
e_i	bound on e_i^*
E	total error bound
t	sample size (sampling with replacement)
u	sample size (sampling without replacement)
$\ln(x)$	natural logarithm of x
$\lceil x \rceil$	x rounded up to the next integer