CLIPAUDIT: A Simple Risk-Limiting Post-Election Audit

Ronald L. Rivest

MIT CSAIL

rivest@mit.edu

January 28, 2017

Abstract

We propose a simple risk-limiting audit for elections, CLIPAUDIT. To determine whether candidate A (the reported winner) actually beat candidate B in a plurality election, CLIPAUDIT draws ballots at random, without replacement, until either all cast ballots have been drawn, or until

$$a - b \geq \beta\sqrt{a+b}$$

where a is the number of ballots in the sample for the reported winner A, and b is the number of ballots in the sample for opponent B, and where β is a constant determined a priori as a function of the number n of ballots cast and the risk-limit α.

CLIPAUDIT doesn't depend on the unofficial margin (as does Bravo).

We show how to extend CLIPAUDIT to contests with multiple winners or losers, or to multiple contests.

Keywords: elections, auditing, post-election audits, risk-limiting audit.

1   Introduction and Motivation

The paper is organized as follows: Section 2 provides some terminology, notation, and orientation. Then Section 3 provides an overview of the CLIPAUDIT audit method. Section 4 discusses the computation of the key constant β.

1

Section 5 analyzes the final expected audit sample size. Some CLIPAUDIT variants are presented in Section 6. Section 7 provides more discussion of CLIPAUDIT. Finally, related work is described in Section 8.

2 Preliminaries

Outcomes, Ballots, Profiles.   We assume an election designed to produce an **outcome** from a set $\mathcal{C}$ of C alternatives (or candidates).

Each of n voters casts a single ballot. A ballot specifies a single candidate, chosen by the voter.

We denote the **profile** of cast ballots as

$$P = \{v_1, v_2, \ldots, v_n\} \,,$$

where each $v_i \in \mathcal{C}$. (One may wish to allow a cast vote v_i to represent an "overvote" or an "undervote" as well, although the election outcome will never be "overvote" or "undervote.") The profile may be viewed as a sequence or as a multiset, since profiles may contain repeated items (identical ballots).

Plurality Election.   We assume a *plurality election*: the candidate with the most votes is the winner. (There may be ties, which can be resolved according to relevant election law.)

Post-election audits.   We assume here that voters have cast votes on **paper ballots** on which their choices were recorded and verified by the voter. These paper ballots were scanned, and the electronic versions of the ballots aggregated to provide the initial or **reported outcome** for the election.

Confidence in the reported election outcome can be derived from a **post-election audit**.

The paper ballots are the "ground truth" for the election; a full and correct count of the paper ballots should give (essentially by definition) the **actual (or true) outcome** for the election.

A "compliance audit" can provide assurance that the paper trail has the necessary integrity. For details, see Benaloh et al. [1], Lindeman and Stark [7], and Stark and Wagner [15].

Statistical post-election audits Instead of using a recount, it is usually more efficient to audit using a statistical method based on hand examination of a **sample** of the paper ballots, a method first proposed by Johnson [5]. Such a **statistical (post-election) audit** typically provides statistical assurance that the reported outcome is indeed equal to the actual outcome, while examining only a relatively small sample of the paper ballots. In the presence of errors or fraud sufficient to make the reported outcome incorrect, the audit may need to examine many ballots, or even all ballots, before concluding that the reported outcome was incorrect.

Risk-limiting audits. Stark [16] has provided a refined notion of a statistical audit—that of a **risk-limiting (post-election) audit** (or **RLA**). If the reported outcome is incorrect, the RLA is guaranteed to have a large, pre-specified chance $(1 - \alpha)$ of examining all cast ballots and thereby correcting the reported outcome. In other words, if the reported outcome is incorrect, the audit will accept the reported outcome as correct with probability at most α.

Here α is an audit parameter (for example, $\alpha = 0.05$).

Lindeman and Stark have provided a "gentle introduction" to RLAs [7]. General overviews of post-election audits are available from Lindeman et al. [6], Norden et al. [9], and the Risk-Limiting Audit Working Group [2]. Stark and Wagner [15] promulgate the notion of an "evidence-based election," which includes not only a risk-limiting audit but also the larger goals of ensuring that the evidence trail has integrity.

A variety of statistical methods for providing RLAs have been developed [17, 4, 18, 19, 20, 14, 21, 3, 10, 8, 13, 23]. We also note the availability of online tools for risk-limiting audits [22]. We compare our CLIPAUDIT proposal with Bravo [8], one of the best and best-known RLA methods.

3 ClipAudit overview

CLIPAUDIT is, like most statistical post-election audits, structured as a sequential decision-making procedure.

Randomly chosen ballots are examined one at a time, until a *stopping rule* says that the audit should stop (accepting the reported outcome as correct),

or until all n ballots in the profile have been examined (thereby revealing the true outcome, which may or may not be equal to the reported outcome).

CLIPAUDIT is a *ballot-polling audit*: it looks only at the paper ballots, and does not consider the electronic version of the ballot data (cast vote record, or CVR) produced by the initial machine-scan of the paper ballots. (A *comparison audit* would do that, but CLIPAUDIT is not a comparison audit. Comparison audits are generally more efficient than ballot-polling audits, and are thus to be preferred to ballot-polling audits when CVRs are available.)

With CLIPAUDIT, ballots are sampled **without replacement**. (A variant of CLIPAUDIT can easily be defined that use sampling with replacement, but such an approach is somewhat less efficient.)

Let α be the desired risk-limit.

CLIPAUDIT should accept an incorrect reported outcome as correct with probability at most α.

If the reported outcome is incorrect, CLIPAUDIT will, with probability at least $1 - \alpha$, proceed to determine the correct outcome by examining *all* of the paper ballots, in accordance with the definition of a risk-limiting audit.

We first describe CLIPAUDIT for the simple case that there are only two candidates. Its extension to handle multiple candidates is described in Section 6.

Let the two candidates be A and B. In the ballots sampled by the audit so far, let a denote the number of votes seen for A and let b denote the number of votes seen for B.

Let $\beta = \beta(n, \alpha)$ be a constant depending on the number n of cast ballots and the desired risk limit α. To get started, the reader may imagine that $\beta = 3$ (which is about right for many typical values of n and $\alpha = 0.05$); more accurate values can be obtained from Figure 1.

CLIPAUDIT has a very simple structure:

$n\backslash\alpha$	0.010	0.020	0.050	0.100	0.200	0.500
100	2.683	2.500	2.236	2.000	1.732	1.155
300	2.887	2.694	2.425	2.145	1.877	1.343
1000	3.054	2.864	2.546	2.294	2.000	1.414
3000	3.184	3.000	2.670	2.401	2.095	1.511
10000	3.290	3.077	2.770	2.496	2.183	1.633
30000	3.357	3.144	2.828	2.556	2.240	1.715
100000	3.411	3.206	2.889	2.638	2.324	1.747
300000	3.487	3.273	2.958	2.684	2.375	1.817
1000000	3.530	3.309	3.000	2.734	2.438	1.890
3000000	3.560	3.352	3.040	2.782	2.474	1.937

Figure 1: Approximate values of $\beta(n, \alpha)$ for various values of n and α, computed using simulation; number of trials per entry is 1,000,000. The computation of these values took about 60 hours on a MacPro laptop, using a short Python3 program.

ClipAudit Procedure:

- Determine $\beta = \beta(n, \alpha)$ from Figure 1, rounding n up and α down if necessary to find a relevant table entry.

- Draw cast ballots at random without replacement, keeping track of the number a of votes seen for the reported winner A and the number b of votes seen for the opposing candidate, until either

$$a - b > \beta\sqrt{a + b} , \tag{1}$$

in which case the reported winner A is accepted as the correct outcome, or until all cast ballots are examined, in which case the correct election outcome is revealed.

---

4 Computing β

The CLIPAUDIT procedure depends on a well-defined value $\beta = \beta(n, \alpha)$, which depends upon the number n of cast votes and upon the risk limit α.

When using CLIPAUDIT to audit a particular election, one can determine an appropriate value of β use by consulting Figure 1. If the exact values of n

and α for the election audit are not available in the table, one can round n up and round α down as needed.

In this section we describe how to compute $\beta(n, \alpha)$.

4.1   Computation of β using simulation

Let n be an even integer.

Let

$$X \in \{-1, 1\}^n$$

be a randomly chosen vector **with zero sum**, and let

$$S_t = X_1 + X_2 + \cdots + X_t \tag{2}$$

for $1 \leq t \leq n$ be the associated random walk. S_t corresponds to the margin of candidate A over B within a sample of size t, as the sample size t increases, when A and B are in fact *tied* in the overall profile of size n.

Note that sampling is done from the profile "without replacement," so that at most n ballots can be drawn for the sample, and when all n are drawn, the two candidates are tied (since $S_n = 0$).

Let S'_t be the observed margin in the actual audit for the sample of size t. Assume that A is the announced winner, and that S'_t is the margin of A over B in the sample.

The stopping condition for CLIPAUDIT is

$$S'_t > \beta\sqrt{t} \tag{3}$$

where β is chosen so that a random walk

$$S = (S_1, S_2, ..., S_n) \tag{4}$$

for a tied race has chance exactly α of stopping (i.e. satisfying (3) for some t, $1 \leq t \leq n$).

CLIPAUDIT is thus a risk-limiting audit (RLA) since if Bob really won, the chance that the audit stops and accepts Alice is not greater than the chance that the audit stops and accepts Alice if there is actually a tie.

Computing β can be done in a couple of ways: simulations or approximate computations via dynamic programming. ("Approximate" because this

method assume that the sampling is with replacement, rather than without replacement.) An analytic approach might also work, but I haven't seen how to do that yet.

In any case, the simulation method is the fastest.

Computations yield, e.g.

$$\beta(n = 10000, \alpha = 0.05) \approx 2.77 \ .$$

Thus, for a 10000-vote election with announced winner Alice with $\alpha = 0.05$ the audit would stop when the margin $a - b$ for Alice over Bob exceeds

$$2.77\sqrt{a + b}$$

when the sample shows a votes for Alice and b votes for Bob.

Simulation method   We now describe the computation of $\beta(n, \alpha)$ by the simulation method in more detail. Let T denote the desired number of trials (simulation runs). We use $T = 10^6$ in our computations.

For each trial i, $1 \leq i \leq T$:

1. Generate a random vector X of length n having zero sum. Each entry in X is $+1$ or -1. (If n is odd, then let X have sum $+1$.)

2. Compute the associated random walk via equation (2).

3. Compute
$$\beta_i = \max_{1 \leq t \leq n} \frac{S_t}{\sqrt{t}} \ . \tag{5}$$

4. Return $\beta(n, \alpha)$ as that kth smallest value β_i, where $k = \lfloor (1-\alpha)T \rfloor$. Here "$k$th smallest" means the value that is larger than or equal to k values in the set—that is, the kth element if the values in the set are sorted into increasing order.

Formula for estimating $\beta(n, \alpha)$ We worked to fit a formula to the entries in Figure 1. Our best fit was to the formula:

$$\beta(n, \alpha) \approx 0.075 \ln(n) + 0.700 \operatorname{isf}(\alpha) + 0.860 \ ; \tag{6}$$

where isf(α) is the inverse survival function for the standard normal distribution—the value x such that $\Phi(x) = 1 - \alpha$, and Φ is the cdf for the standard normal distribution. Over the values shown in the table the approximation was always within 0.16, and typically less than 0.05.

To get a formula that is a good fit but always an upper bound on the desired β value, it suffices to raise the constant in the formula to 1.00:

$$\beta(n, \alpha) \leq 0.075 \ln(n) + 0.700\, \text{isf}(\alpha) + 1.000 \; ; \tag{7}$$

The CLIPAUDIT user may find it convenient to derive $\beta(n, \alpha)$ using formula (6) or formula (7) rather than using the table.

5   Expected Final Sample Size

Let m denote the actual fractional margin in favor of A—the fraction of votes for A minus the fraction of votes for B. If the actual fractional margin in favor of A is m, then the expected value of S_t is mt.

Thus, $S_t \geq \beta\sqrt{t}$ is expected to occur when

$$t \geq \frac{\beta^2}{m^2} \; . \tag{8}$$

Compare with Bravo result: equation (7) in [8] says that Bravo should terminate once

$$t \geq \frac{2\ln(1/\alpha)}{m^2} \; . \tag{9}$$

Comparing CLIPAUDIT's β^2 with Bravo's $2\ln(1/\alpha) \approx 6$, (for $\alpha = 0.05$), we see that Bravo has an advantange for $\beta > \sqrt{6} = 2.449$, *assuming that the reported margin is correct.*

Example   Consider auditing an election with $n = 50000$ cast ballots, of which 60% are reported for candidate A and 40% are reported for candidate B, with a risk limit of $\alpha = 0.10$.

For Bravo, formula (9) gives an estimated sample size of 115; experimental results (shown in Table 1 of Lindeman et al. [8]) show an average sample size (ASN) of 119, assuming that the reported fraction of votes for each candidate is correct.

For ClipAudit, we use the value $\beta = 2.568$ from formula (6); then formula (8) gives an estimated sample size of 165; experimental results show an average sample size of 143.

We see that (at least for this example), using ClipAudit rather than Bravo incurs a performance penalty (in terms of number of ballots examined) on the order of 20%.

However, one advantage of ClipAudit is that its performance is *insensitive* to the reported vote fractions. If the reported vote fractions had been 70% for candidate A and 30% for candidate B (while the true fractions remained at 60% and 40%), Bravo would have performed a *full recount* with high probability, examining all 50000 ballots, whereas the expected workload by ClipAudit would be unchanged. Intuitively, Bravo is trying to see if the true vote fractions are "closer" to the reported vote fractions or to vote fractions representing a tied outcome; if the true vote fractions are "closer" to a tie (even if they support the reported outcome), then the Bravo audit will typically be forced into a doing a full recount.

Thus, one might adopt ClipAudit in situations where the reported vote fractions are a bit suspect, and/or where there is a desire to avoid the possibility of a full recount being caused by incorrect reported vote fractions.

6 Variants

Multiple candidates. The ClipAudit method can be easily extended to handle more than two candidates, possibly with multiple winners. Suppose the election has C candidates, and that the top W vote-getters will be declared the *winners*, and that the remaining $L = C - W$ will be declared *losers*. Here either W or L or both may be larger than 1.

Let $A_1, \ldots, A_W$ denote the W reported winners, and let $B_1, \ldots, B_L$ denote the L reported losers.

We simply modify ClipAudit to concurrently audit that each A_i defeats B_j, for $1 \leq i \leq W$ and $1 \leq j \leq L$. Each subaudit uses the same sequence of audited ballots, paying attention only to those ballots relevant to its task. Each audit uses the same overall risk limit α. The overall audit stops and accepts the reported outcomes if and only if all of the subaudits have stopped

and accepted that their A_i defeats their B_j.

(The overall audit is an RLA with risk limit α because the chance that the overall audit accepts an incorrect outcome is at most the chance that a particular one of the subaudits with an incorrect hypothesis to check accepts it as correct.[1])

Multiple contests   Multiple contests can be audited concurrently in a similar way. Each contest is audited using the overall risk limit α as its individual risk limit. Ballots are sampled uniformly from the set of all ballots having relevant (still-being-audited) contests on them. When a ballot is examined, choices for all relevant (still-being-audited) contests are determined by hand, and the audits for the relevant contests make progress (and possibly terminated). When all audits have terminated, or all ballots examined for all audits still being audit, the audit stops.

Each individual contest is audited in a risk-limiting manner. The fact that the different audits are using the same ballot samples doesn't affect this argument (although it may mean that the different audit results are correlated, which shouldn't be a problem).

7   Discussion

CLIPAUDIT may be easier to explain or understand than Bravo, and doesn't have any dependence on the unofficial margin.

It would be nice to prove a simple formula for $\beta(n, \alpha)$. Nonetheless, $\beta(n, \alpha)$ can be adequately estimated for practical use of CLIPAUDIT.

8   Related Work

DiffSum[11] has an identical structure, but the constant corresponding β there has no formal definition, and the method is not provably risk-limiting.

We note for the record that there are statistical post-election audits that don't seem to quite fit the "risk-limiting audit" definition, but which nonethe-

---

[1]Thanks for Philip B. Stark for providing this argument.

less have sound probabilistic foundations. In particular, the "Bayesian audit" of Rivest and Shen [12] is of this character. We refer the reader to that paper for details and discussion.

Acknowledgments

References

[1] J. Benaloh, D. Jones, E. Lazarus, M. Lindeman, and P.B. Stark. SOBA: Secrecy-preserving observable ballot-level audit. In *Proceedings 2011 Electronic Voting Technology Workshop/Workshop on Trustworthy Elections (EVT/WOTE '11)*, 2011. `http://static.usenix.org/events/evtwote11/tech/final_files/Benaloh.pdf`.

[2] J. Bretschneider, S. Flaherty, S. Goodman, M. Halvorson, R. Johnston, M. Lindeman, R.L. Rivest, P. Smith, and P.B. Stark. Risk-limiting post-election audits: Why and how?, Oct. 2012. (ver. 1.1) `http://people.csail.mit.edu/rivest/pubs.html#RLAWG12`.

[3] S. Checkoway, A. Sarwate, and H. Shacham. Single-ballot risk-limiting audits using convex optimization. In D. Jones, J.-J. Quisquater, and E. Rescorla, editors, *Proceedings 2010 EVT/WOTE Conference*. USENIX/ACCURATE/IAVoSS, August 2010.

[4] J. L. Hall, L. W. Miratrix, P. B. Stark, M. Briones, E. Ginnold, F. Oakley, M. Peaden, G. Pellerin, T. Stanionis, and T. Webber. Implementing risk-limiting post-election audits in California. In *Proc. 2009 Electronic Voting Technology Workshop/Workshop on Trustworthy Elections (EVT/WOTE*

'09, Montreal, Canada). USENIX, Aug 2009. `http://www.usenix.org/event/evtwote09/tech/full_papers/hall.pdf`.

[5] K. Johnson. Election verification by statistical audit of voter-verified paper ballots. `http://ssrn.com/abstract=640943`, Oct. 31 2004.

[6] M. Lindeman, M. Halvorseon, P. Smith, L. Garland, V. Addona, and D. McCrea. Principle and best practices for post-election audits. `www.electionaudits.org/files/best%20practices%20final_0.pdf`, 2008.

[7] Mark Lindeman and Philip B. Stark. A gentle introduction to risk-limiting audits. *IEEE Security and Privacy*, 10:42–49, 2012.

[8] Mark Lindeman, Philip B. Stark, and Vincent S. Yates. BRAVO: Ballot-polling risk-limiting audits to verify outcomes. In Alex Halderman and Olivier Pereira, editors, *Proceedings 2012 EVT/WOTE Conference*, 2012.

[9] Lawrence Norden, Aaron Burstein, Joseph Lorenzo Hall, and Margaret Chen. Post-election audits: Restoring trust in elections. Technical report, Brennan Center for Justice and Samuelson Law, Technology & Public Policy Clinic, 2007.

[10] California Secretary of State. Post-election risk-limiting audit pilot program, 2011-2013. `http://www.sos.ca.gov/elections/voting-systems/oversight/post-election-auditing-regulations-and-reports/post-election-risk-limiting-audit-pilot-program/`.

[11] Ronald L. Rivest. DiffSum – a simple post-election risk-limiting audit. arXiv abs/1509.00127, 2015.

[12] Ronald L. Rivest and Emily Shen. A Bayesian method for auditing elections. In J. Alex Halderman and Olivier Pereira, editors, *Proceedings 2012 EVT/WOTE Conference*, 2012.

[13] A. D. Sarwate, S. Checkoway, and H. Shacham. Risk-limiting audits and the margin of victory in nonplurality elections. *Politics and Policy*, 3(3):29–64, 2013.

[14] P. B. Stark. Risk-limiting vote-tabulation audits: The importance of cluster size. *Chance*, 23(3):9–12, 2010.

[15] P. B. Stark and D. A. Wagner. Evidence-based elections. *IEEE Security and Privacy*, 10(05):33–41, Sep-Oct 2012.

[16] Philip B. Stark. Conservative statistical post-election audits. *Ann. Appl. Stat.*, 2:550–581, 2008.

[17] Philip B. Stark. A sharper discrepancy measure for post-election audits. *Ann. Appl. Stat.*, 2:982–985, 2008.

[18] Philip B. Stark. CAST: Canvass audits by sampling and testing. *IEEE Trans. Inform. Forensics and Security*, 4(4):708–717, Dec. 2009.

[19] Philip B. Stark. Efficient post-election audits of multiple contests: 2009 California tests. `ssrn.com/abstract=1443314`, 2009. 2009 Conference on Empirical Legal Studies.

[20] Philip B. Stark. Risk-limiting post-election audits: P-values from common probability inequalities. *IEEE Trans. on Information Forensics and Security*, 4:1005–1014, 2009.

[21] Philip B. Stark. Super-simple simultaneous single-ballot risk-limiting audits. In *Proc. 2010 EVT/WOTE Workshop*, 2010. `http://www.usenix.org/events/evtwote10/tech/full_papers/Stark.pdf`.

[22] Philip B. Stark. Tools for comparison risk-limiting election audits. `http://www.stat.berkeley.edu/~stark/Vote/auditTools.htm`, 2015.

[23] Philip B. Stark and Vanessa Teague. Veriable european elections: Risk-limiting audits for d'hondt and its relatives. *USENIX Journal of Election Technology and Systems (JETS)*, 1(3):18–39, 2014.