

A Simple Rule of Thumb for Election Audit Size Determination

Ronald L. Rivest

Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139
rivest@mit.edu

Draft of November 3, 2007

Abstract

This note proposes a very simple “Rule of Thumb” for determining how many precincts to audit in a post-election audit. This rule says to audit

$$1.0/m \tag{1}$$

precincts (rounded up to the next integer), where m is the “margin of victory” (the difference in percentage of votes for the apparent winner and for the apparent runner-up). For example, if the margin of victory is 4%, then this rule of thumb suggests auditing $1.0/0.04 = 25$ precincts. If the formula suggests auditing more precincts than are available, all precincts are audited.

This Rule of Thumb gives a guaranteed “confidence level” of at least 92% when certain assumptions hold (when all precincts have the same size and the miscount within each precinct is at most 20% of the votes in that precinct). Under these assumptions, the actual confidence level achieved ranges from 92% to 100%.

This Rule of Thumb is simple enough to be computed in one’s head; it provides reasonable guidance for “back-of-the-envelope” estimates, and is useful when the simplicity of the formula is especially important.

1 Introduction

The purpose of this note is to put into writing a suggestion I made at a panel at the October 26, 2007 “Post-Election Audit Summit” *.

At this meeting, quite a few participants expressed a desire to have a *very* simple rule for determining audit size. The rule proposed here is about as simple as you get while still retaining some guaranteed statistical quality (under some assumptions).

To keep this note short, I’ll presume the reader has some familiarity with election auditing. Good introductions can be found in Norden et al. [6] or the Election-Line report [3]. Dopp [2] and Hall [4] provide helpful guidance to previous literature.

The presentation here is (intentionally) overly simplistic in many regards; see Section 4 for some discussion.

I recommend that the reader also consult Aslam et al. [1] for motivation and notation; the present paper should be viewed as a minor extension of that paper.

Suppose we have an election with n precincts.

In each precinct we have both electronic records and paper records for each voter. The electronic records are easy to tally.

A precinct can be “audited” by counting (by hand) the paper records of that precinct, to confirm that they (very nearly) match the electronic totals for that precinct. Roughly, an election can be audited as follows:

1. Determine a number u of precincts to be audited.
2. Draw a set S of u precincts at random.
3. Audit all precincts in S . If no significant discrepancies are found between the electronic tallies and the hand-count of the paper ballots, stop and declare the apparent winner (from the electronic tallies) to be the election winner. Otherwise, investigate further, and perform audits on a larger sample of precincts (or on all precincts) as deemed necessary.

The “confidence level” c for such an audit is the probability that significant discrepancies will be discovered during the audit of u precincts if there are enough discrepancies so that the electronic and paper records give different election outcomes.

The “margin of victory” m is the difference between the fraction of votes for the winning candidate and the fraction of votes for the runner-up.

To achieve a given confidence level c , without having the audit all precincts, the size u of the sample should depend on the margin of victory. As the margin decreases, the size of the sample should increase. See Aslam et al. [1] for discussion and an essentially exact formula for u as a function of m and n . A 95% confidence level ($c = 0.95$) might be a “typical” target confidence level.

(Built in to the such formulae is typically an assumption that an adversary “wouldn’t dare” to change more than some fraction s of votes in a precinct. For example,

*<http://electionaudits.org/index.html>

$s = 0.20$ is a typical assumption [5]. The value s is also denoted WPM (“Within-Precinct Miscount”).)

2 Derivation

The audit size Rule of Thumb proposed here is essentially an application of the (generalized) “Rule of Three” as discussed in Aslam et al. [1, equation (15)]. The (generalized) Rule of Three says to audit

$$-n \ln(1 - c)/b \tag{2}$$

precincts, where b is number of “corrupted” (bad) precincts one wishes to detect with probability at least c (our “confidence level”). (Here \ln is the natural logarithm function; the fact that $\ln(1 - 0.95) \approx -3$ explains the name “Rule of Three”).

If the margin of victory is m , and the apparent (electronic) winner is not the real (paper) winner, then there must be a total miscount (discrepancy) of at least $mn/2$ votes between the electronic and paper counts. If we assume that all precincts have the same size, and that each precinct’s miscount is at most a fraction s of its total count, then the number of miscounted (bad) precincts is at least $b = mn/2s$. Plugging this into equation (2), the number of precincts to audit is:

$$-2s \ln(1 - c)/m \tag{3}$$

Thus, when the following equation holds:

$$-2s \ln(1 - c) = 1 \tag{4}$$

we have the Rule of Thumb (equation (1)). For example, consider $s = 0.20$ and $c = 0.92$ (confidence level 92%). The left hand side of equation (4) is just the numerator of Rule of Thumb fraction (equation (1)).

3 Performance

What confidence does the Rule of Thumb provide?

Because the Rule of Thumb is derived from the Rule of Three with $c = 0.92$, the Rule of Thumb *guarantees* a confidence level of 92% (assuming that $s = 0.20$ and that all precincts have the same size).

A small Python program was written to explore the samples sizes and confidence levels for the Rule of Thumb for a wide range of “reasonable” values of n and m . Here n ranged from 2 to 10000 and m ranged from 0.001 to 0.5. The confidence levels ranged from 0.92 to 1.00; the median confidence level obtained was 0.996. Thus, the Rule of Thumb could be viewed as one that *guarantees* a confidence level of 92% (subject to assumptions), or one that *approximates* a confidence level of 96% or so (plus or minus up to four percent). Also The Rule of Thumb sample size was never more than a factor

of 1.42 too large (i.e. compared to the optimal sample size for a 0.92 confidence level); the median ratio was 1.09. (I don’t consider the “median” values quoted here particularly meaningful...)

4 Discussion

Note that this Rule of Thumb doesn’t specify a *percentage* of precincts to audit, it specifies a specific *number* of precincts to be audited.

While it is common in the election literature to talk about *percentage-level* audits (e.g. a 1% audit or a 10% audit), this is very misleading from a statistical point of view, since it is really the *number* of precincts audited that matters most. Note that the Rule of Thumb gives the same answer no matter how many precincts are available. (With the understanding that if the Rule specifies more precincts than are available, you stop once all precincts have been audited.) If the Rule of Thumb says to audit 50 precincts (e.g. for a 2% margin of victory), then this recommendation holds whether you are auditing just a county or the whole United States. Obviously, the percentage of precincts audited varies greatly in these two examples, but the confidence obtained with such 50-precinct audit is nearly the same in these two examples.

I would thus encourage others to specify and emphasize the *number* of precincts audited first, and then perhaps give the percentage parenthetically. (For example, say: “With a 20-precinct audit (2% audit), we obtained the following results...”)

Of course, one might reasonably have some sort of “baseline” audit (e.g. 10 precincts, or one precinct from each county) that is always performed, even if the margin of victory is large.

This note doesn’t describe or recommend procedures to be followed in the case that a significant discrepancy *is* discovered in the initial audit of u precincts (nor does it specify in any detail what a “significant” discrepancy is); Stark [8] does.

The confidence level c guaranteed (under the same assumptions as before) can be varied by changing the numerator of the Rule of Thumb (equation (1)) to something other than 1 (to $-2s \ln(1 - c)$):

- A numerator of 1.85 ensures $c \geq 0.99$.
- A numerator of 1.20 ensures $c \geq 0.95$.
- A numerator of 1.00 ensures $c \geq 0.92$.
- A numerator of 0.93 ensures $c \geq 0.90$.
- A numerator of 0.56 ensures $c \geq 0.75$.
- A numerator of 0.28 ensures $c \geq 0.50$.

The ratios of these quantities give a very rough guide as to how much you might have to expand an audit for a given increase the confidence level; for example, to go from a 90% confidence level to a 99% confidence level, you might have to roughly double the size of the audit (since $1.85/0.93$ is close to 2).

The Rule of Thumb doesn't take precinct size variation into account; see Rivest [7], McCarthy et al. [5], or Aslam, Popa, Rivest[†] for discussion; when precinct sizes vary the sample size must be increased in a way that seems difficult to incorporate into a simple formula.

The Rule of Thumb tends to be an overestimate for the correct sample size when all precincts have the same size, because the Rule of Three (from which the Rule of Thumb is derived) is for sampling with replacement, rather than sampling without replacement. On the other hand, when precincts sizes vary considerably, the Rule of Thumb may nonetheless underestimate the correct sample size. With luck, these effects will balance out...

Because the audit size Rule of Thumb is so very simple, it might be useful at times (perhaps some legislative situations) when the simplicity of the formula is more important than being as efficient as possible (i.e. as accurate as possible for a desired level of confidence).

5 Example

Suppose we have $n = 640$ precincts with a total vote count of 315,540, and a margin of victory of $m = 0.01$ (1%), and a desired confidence level of 92%.

If all precincts have the same size, then:

- The Rule of Thumb suggests auditing 100 precincts chosen uniformly at random, with an expected workload of 49303 votes to count.
- The optimal sample size (computed using the formula of Aslam, Popa, and Rivest[1]) is 93 precincts, with an expected workload of 45852 votes to count.

The Rule of Thumb is within 8% of the optimal workload.

Now suppose that we still $m = 0.01$, but that precincts have varying sizes: suppose that they have the distribution of precinct sizes from Ohio 2004 Congressional District 5 (OH-5), with 315,540 votes distributed into 640 precincts, where the largest has 1637 votes and the smallest has 132[‡].

- If we blindly use the Rule of Thumb, we'll sample 100 precincts uniformly at random, with expected workload of 49303, as before. But now the confidence level drops from 92% to 67%.

[†]On Auditing Elections When Precincts Have Different Sizes. To appear.

[‡]Thanks to Mark Lindeman for supplying this data, which is available at: <http://people.csail.mit.edu/rivest/pps/oh5votesonly.txt>.

- The SAFE method [5] adjusts for sample size variation; it finds that the adversary must corrupt at least 7 precincts, and then samples 193 precincts chosen uniformly, with an estimated workload of 95,155 votes, and confidence 92%.
- The logistic approach (see Rivest [7]) samples precincts with varying probability (roughly proportional to their size); it samples 93 precincts on the average, with an expected workload of 50,937 votes, and confidence 92%.

Thus, auditing in a manner that is nicely sensitive to varying precinct sizes can be beneficial. (But that is the subject of another paper...) It is perhaps interesting that in the end, the Rule of Thumb estimated a counting workload (49,303 votes) that is very near the expected workload (50,937 votes) for the logistic method that takes precinct size into account. Thus, the Rule of Thumb may sometimes provide a good estimate for the amount of work one has to do, even though the actual procedure used, involving non-uniform sampling, may be different than the uniform sampling with replacement procedure that is behind the Rule of Thumb.

References

- [1] 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.
- [2] 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>.
- [3] ElectionLine.org. Case study: Auditing the vote, March 2007. <http://electionline.org/Portals/1/Publications/EB17.pdf>.
- [4] Joseph Lorenzo Hall. Post-election manual auditing of paper records: Bibliography, 2007. http://www.josephhall.org/papers/auditing_biblio.pdf.
- [5] 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>.

- [6] 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, http://www.brennancenter.org/dynamic/subpages/download_file_50089.pdf.
- [7] Ronald L. Rivest. On auditing elections when precincts have different sizes, 2007. <http://people.csail.mit.edu/rivest/OnAuditingElectionsWhenPrecinctsHaveDifferentSizes.pdf>.
- [8] Philip B. Stark. Conservative statistical post-election audits, 2007. <http://statistics.berkeley.edu/~stark/Preprints/conservativeElectionAudits07.pdf>.