Super-Simple Simultaneous Single-Ballot Risk-Limiting Audits

EVT/WOTE 2010
Washington, DC 10 August 2010

Philip B. Stark
Department of Statistics
University of California, Berkeley

Everything should be made as simple as possible, but no simpler—A.A. Einstein.

Many thanks to Eric Rescorla, Joe Hall, and Mark Lindeman for comments on earlier drafts!

Abstract: Simultaneous risk-limiting audits of a collection of contests have a known minimum chance of leading to a full hand count if any outcomes are wrong. They are generally performed in stages. Each stage involves drawing a sample of ballots, comparing a hand count of the votes on those ballots with the original count, and assessing the evidence that the original outcomes agree with the outcomes that a full hand count would show. If the evidence is sufficiently strong, the audit can stop; if not, more ballots are counted by hand and the new evidence is assessed. This paper derives simple rules to determine how many ballots must be audited to allow a simultaneous risk-limiting audit to stop at the first stage if the error rate in the sample is sufficiently low. The rules are of the form "audit at least ρ/μ ballots selected at random." The value of ρ depends on the simultaneous risk limit and the amount of error to be tolerated in the first stage without expanding the audit. It can be calculated once and for all without knowing anything about the contests. The number μ is the "diluted margin": the smallest margin of victory in votes among the contests, divided by the total number of ballots cast across all the contests. The initial sample size does not depend on any details of the contests, just the diluted margin. This is far simpler than previous methods.

For instance, suppose we are auditing a collection of contests at simultaneous risk limit 10%. In all, N ballots were cast in those contests. The smallest margin is V votes: The diluted margin is $\mu = V/N$. We want the audit to stop at the first stage provided the fraction of ballots in the sample that overstated the margin of some winner over some loser by one vote is no more than $\mu/2$ and no ballot overstates any margin by two votes. Then an initial sample of $15.2/\mu$ ballots suffices. If the sample shows any two-vote overstatements or more than 7 ballots with one-vote overstatements, more sampling might be required, depending on which margins have errors. If so, simple rules that involving only addition, subtraction, multiplication, and division can be used to determine when to stop.

Goal: Truly simple audit rules that allow elections officials to confirm that the outcomes of most contests are right, with one (small) sample.

Risk-limiting: large chance of correcting any outcomes that are wrong—i.e., that disagree with the outcome full hand count of the audit trail would show. (Correct them by conducting a full hand count.)

Exploit statistical efficiency of *single-ballot auditing*, which compares CVR with human interpretation of individual ballots.

Spend some of that efficiency on logistic and computational simplicity. Method still "cheap."

Have to match CVRs to physical ballots. Requires new voting systems or *transitive auditing* using parallel systems (e.g., Clear Ballot Group, Humboldt ETP, TrueBallot) *a la* Calendrino et al. (2007)

Advantages

- Audit entire collection of contests with one sample.
- Super simple: initial sample size is a constant divided by the "diluted margin." Constant set once and for all: doesn't depend on any particulars of the contests, margins, etc.
- Audit expands if too many ballots with errors that overstate a margin by one vote, or any ballots that overstate a margin by two votes. Determining when to stop is simple.
- Chance of correcting all wrong outcomes is guaranteed to be at least as high as claimed.

Definitions

Outcome: set of winners, not exact vote totals.

Machine-count outcome, apparent outcome: outcome that will become officially final unless an audit or other action intervenes.

Apparent winner: won according to apparent outcome.

Hand-count outcome, true outcome: outcome a full manual tally of the audit trail would show.

True winner: won according to full hand tally.

. . . more definitions . . .

Correct outcome: hand-count outcome—even though hand counting and audit trail aren't perfect.

Risk-limiting audit: guaranteed minimum chance of correcting a wrong outcome (by counting the whole audit trail). Endorsed by ASA, CC, VV, LWV, CEIMN, . . .

Risk: maximum chance that the audit fails to correct an apparent outcome that is incorrect, no matter what caused the outcome to be incorrect.

Risk-measuring audit: reports the strength of the evidence that the outcome is correct, but does not necessarily continue to count votes until that evidence is strong or all votes have been counted by hand.

Measured risk: P-value of the hypothesis that the outcome is incorrect, given the data collected by the audit.

still more . . .

Simultaneous risk-limiting audit: guaranteed minimum chance of correcting *all* the contests that have incorrect apparent outcomes.

Simultaneous risk of a simultaneous risk-limiting audit: the maximum chance that the audit will fail to correct one or more of the apparent outcomes that are incorrect, no matter what caused them to be incorrect.

Six risk-limiting audits so far, in California: 2 in Marin, 1 in Santa Cruz, 3 in Yolo.

California AB 2023 passed last month—official pilot of risk-limiting audits in 2011.

Audits compare hand counts of randomly selected "batches" of ballots to apparent counts. Reducing batch size improves statistical efficiency.

Single-ballots audits: batch size one. Compare CVR with human interpretation for a sample of individual ballots.

Jelly bean analogy; soup analogy.

Need tech that makes that possible.

One single-ballot risk-limiting audit so far (Yolo County, CA).

Yolo County Measure P, November 2009 (not single-ballot)

Reg. voters	ballots	precincts	batches	yes	no
38,247	12,675	31	62	3,201	9,465

VBM and in-person ballots were tabulated separately (62 batches).

For risk-limit 10%, initial sample size 6 batches; gave 4 distinct batches, 1,437 ballots.

Single-ballot auditing saves *lots* of work

Can determine the initial sample size for a single-ballot audit even though the cast vote records (CVRs) were not available.

For risk limit 10%, would need to look at CVRs for 6 ballots. That's less than 0.05% of ballots cast–one twentieth of one percent.

For risk limit 1%, would need to look at CVRs for 12 ballots. That's less than 0.1% of ballots cast–one tenth of one percent.

Cf., 1,437 ballots (11.33% of ballots cast) for actual batch sizes.

	contest				
	1	2	3	4	5
CVR	undervote	winner	loser	not on ballot	not on ballot
Hand	loser	loser	winner	not on ballot	not on ballot
overstatement	1	2	-2	0	0

Hypothetical CVR and hand interpretation of a ballot that contains three of five contests under audit. "Winner" and "loser" denote an apparent winner and an apparent loser, respectively. The maximum overstatement is two votes.

	contest				
	1	2	3	4	5
CVR	undervote	winner	loser	winner	not on ballot
Hand	loser	overvote	loser	winner	not on ballot
overstatement	1	1	0	0	0

Hypothetical CVR and hand interpretation of a ballot that contains four of five contests under audit. "Winner" and "loser" denote an apparent winner and an apparent loser, respectively. In contest 3, the CVR and hand count found votes for one and the same apparent loser, and in contest 4, the CVR and hand count found votes for one and the same apparent winner. There are two overstatement errors, but the maximum overstatement is one vote.

Sufficient condition for all outcomes to be right:

> For every winner and loser, the overstatement errors minus the understatement errors amount to less than 100% of the margin between that pair of candidates.

MACRO (maximum across-race relative overstatement) summarizes overstatement errors within and across contests.

If the MACRO summed over all ballots is less than 100%, all outcomes of all contests are correct.

Method here uses conservative simplification of MACRO. Errs on the side of safety. True simultaneous risk is smaller than the nominal simultaneous risk limit.

New procedure requires setting 3 numbers:

- simultaneous risk limit α
- error inflation factor $\gamma \geq 100\%$
- error tolerance $\lambda < 100\%$.

α might be set in legislation.

γ and λ affect operating characteristics but not risk.

α, γ, λ determine sample size multiplier ρ. Find ρ once and for all, before audit.

Initial sample size is ρ/μ, where μ is the *diluted margin*: smallest margin in votes, divided by the total number of ballots cast across all the contests

Procedure

1. Pick α, e.g., 10%. This is the largest chance that an incorrect outcome will not be corrected by the audit.

2. Pick $\gamma \geq 100\%$, e.g., $\gamma = 110\%$. γ controls a tradeoff between initial sample size and additional counting required when the sample finds many overstatements, especially two-vote overstatements.

3. Pick $\lambda < 100\%$, e.g., $\lambda = 50\%$. λ is tolerable rate of one-vote maximum overstatements in the initial sample as a fraction of μ. If the percentage of ballots in the sample with of one-vote maximum overstatements is no more than $\lambda\mu$ and there are no two-vote overstatement, audit stops.

4. Calculate the sample-size multiplier ρ

$$\rho = \frac{-\log \alpha}{\frac{1}{2\gamma} + \lambda \log(1 - \frac{1}{2\gamma})}.$$

For $\alpha = 10\%$, $\gamma = 110\%$ and $\lambda = 50\%$, $\rho = 15.2$. ρ doesn't depend on the audit data or particulars of the contests.

5. Calculate the diluted margin μ.

6. Audit random sample of at least $n = \lceil \rho/\mu \rceil$ ballots. If fewer than $n\lambda\mu$ of those have one-vote maximum overstatements and none has a two-vote overstatement, stop. Otherwise, Kaplan-Markov P-value determines when to stop. If $\mu = 2\%$ and $\rho = 15.2$ we would audit $15.2/2\% = 760$ ballots. If fewer than 8 of those ($760\lambda\mu = 7.6$) have a maximum one-vote overstatement and none has a two-vote overstatement, stop.

Math

Based on Kaplan-Markov method (Markov's inequality applied to optionally stopped Martingale formed from iid observations) and MACRO, with conservative simplifications for single ballots.

Use universal upper bound on error in each ballot and base test on smallest margin. Draw n of the N ballots at random with replacement with equal probability.

Kaplan-Markov MACRO P-value

$$P_{KM} = \prod_{r=1}^{n} \frac{1 - 1/U}{1 - \frac{e_r}{2\gamma/V}}. \tag{1}$$

Audit with simultaneous risk limit α by hand counting until $P_{KM} \le \alpha$ (or until all ballots have been counted) by hand;

Simplification

$$\begin{aligned} P_{KM} &\leq P(n, n_1, n_2; U, \gamma) \\ &\equiv [1 - 1/U]^n \times [1 - 1/(2\gamma)]^{-n_1} \times [1 - 1/\gamma]^{-n_2}\,. \end{aligned}$$

Examples

diluted margin μ	$\lambda = 50\%$ risk limit α 10%	5%	1%	$\lambda = 20\%$ risk limit α 10%	5%	1%
5%	305	396	609	139	180	277
2%	761	989	1521	346	450	691
1%	1521	1978	3041	691	899	1382
0.5%	3041	3956	6081	1382	1798	2764
multiplier ρ	15.20	19.78	30.40	6.91	8.99	13.82

Initial sample sizes n and sample-size multipliers ρ for $\gamma = 110\%$. Column 1: diluted margin of victory μ. Columns 2–4: n for various simultaneous risk limits if the audit is to stop when the percentage of ballots in the sample that overstate a margin by one vote is not more than 50% of the diluted margin. Columns 5–7: n for various simultaneous risk limits if audit is to stop when the percentage of ballots in the sample that overstate a margin by one vote is not more than 20% of the diluted margin. Last row: In columns 2–7, the sample sizes n are equal to these "multipliers" divided by the diluted margins μ. n is computed using inequality **??**. The simultaneous risk bound $P(n, n_1, n_2; U, \gamma)$ is generally on the order of 2/3 of the nominal values in the column headings.

Conclusions Very simple formula for initial sample for risk-limiting audit. Allows audit to stop if, in the initial sample, rate of 1-vote maximum overstatements is at most a pre-specified fraction of the margin and there are no 2-vote overstatements.

Method requires choosing 3 numbers

Simple but somewhat inefficient (see Checkoway et al. 2010; Stark 2009): More ballots have to be counted by hand than if sharper bounds were used, but those methods require far more complex math.

Single-ballot audits are so efficient that total cost still low.

To implement single-ballot audits on a wide scale may require changes to vote tabulation systems: have to associate individual cast vote records (CVRs) with individual physical ballots.

Auditing using an unofficial vote tabulation system that does produce CVRs—such as those of Clear Ballot Group, the Humboldt Transparency Project, or TrueBallot—and confirming transitively that the apparent outcome is correct, might be the best interim option. (See Calendrino et al. 2007)