Sexy Audits and the Single Ballot

Election Verification Network Annual Conference
Washington, DC
25–27 March 2010

Philip B. Stark
http://statistics.berkeley.edu/~stark

This document:
http://statistics.berkeley.edu/~stark/Seminars/evn10.pdf

Problem: Any way of counting votes makes mistakes.

If there are enough mistakes, apparent winner could be wrong.

If there's an audit trail that reflects the right outcome, can ensure big chance of fixing wrong outcomes.

Crucial question: when to *stop* counting, not where to start.

Solution: If there's compelling evidence that outcome is right, stop; else, audit more.

Efficiency is primarily about batch sizes: Need data plumbing and vote tabulation systems (VTSs) designed for auditing.

Current Laws

Misplaced focus on how big an initial sample to draw.

Debates over fixed percentages, tiered percentages depending on the margin, or sample sizes that vary with the margin and batch sizes.

The important issue isn't where to start. It's when to stop.

Can't fix a wrong outcome without hand-counting the whole audit trail.

Risk-Limiting Audits

> If the electoral outcome is wrong, there's a known minimum chance of a full hand count (which fixes it), *no matter what caused the outcome to be wrong*.

The *risk* is the largest chance that an outcome that is wrong won't be fixed.

"Wrong" means the outcome isn't what a full hand count would show.

Role of statistics: Less counting when the outcome is right, but still a big chance of a full hand count when outcome is wrong.

Essential that voters create a durable audit trail that reflects the true outcome.

Essential that voting systems enable auditors to access reported results (total ballots, counts for each candidate) in auditable batches.

Essential to perform "ballot accounting" to ensure that no ballots appeared or disappeared.

Essential to select batches at random, *after* the results are posted or "committed." (Can supplement with "targeted" samples.)

Need a plan for when to count more ballots, possibly leading to full hand count. "Explaining" or "resolving" isn't enough. Plan must ensure that the chance of a full hand count is high whenever the outcome is wrong.

Compliance audits vs. materiality audits.

California AB 2023 (Saldaña, sponsored by SoS Bowen): http://www.leginfo.ca.gov/pub/09-10/bill/asm/ab_2001-2050/ab_2023_bill_20100325_amended_asm_v98.html

First proposed audit bill that limits risk!

15560. (a) The Secretary of State is authorized to establish a postcanvass risk-limiting audit pilot program in five or more counties to improve the accuracy of, and public confidence in, election results. The Secretary of State is encouraged to include urban and rural counties; counties from northern, central, and southern California, and counties with various different voting systems.

The volunteer counties audit one or more contests after each election in 2011.

The Secretary of State reports to the Legislature by March 2012 on the effectiveness, efficiency, and costs of risk-limiting audits.

California AB 2023, contd.:

(b)(3) "Risk-limiting audit" means a manual tally employing a statistical method that ensures a large, predetermined minimum chance of requiring a full manual tally whenever a full manual tally would show an electoral outcome that differs from the outcome reported by the vote tabulating device for the audited contest. A risk-limiting audit shall begin with a hand tally of the votes in one or more audit units and shall continue to hand tally votes in additional audit units until there is strong statistical evidence that the electoral outcome is correct. In the event that counting additional audit units does not provide strong statistical evidence that the electoral outcome is correct, the audit shall continue until there has been a full manual tally to determine the correct electoral outcome of the audited contest.

Amen!

Quantifying the Evidence the Audit Sample Gives

What is the biggest chance that—if the outcome is wrong—the audit would have found "as little" error as it did?

That chance depends on

- how the sample is drawn and its size
- batch sizes and reported votes in each batch
- the errors that are found

Chance can be big *even if no errors are found*—if the sample is small or the margin is small.

Don't stop counting until that chance is small!

Calculations

The calculations are not hard.

For the most efficient method so far (the *Kaplan-Markov method*), only need ordinary arithmetic: addition, subtraction, multiplication, and division.

Nothing fancy.

Pilot Audits in California

Marin County 2/08 (first ever); 11/08

Santa Cruz County 11/08

Yolo County 11/08, 11/09 (2, incl. 1st single-ballot audit)

Measures requiring super-majority, simple measures, multi-candidate contests, vote-for-n contests.

Contests ranged from about 200 ballots to 121,000 ballots.

Counting burden ranged from 32 ballots to 7,000 ballots.

Cost per audited ballot ranged from nil to about \$0.55.

Yolo County Measure P, November 2009

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 would save *lots* of work

Can determine the initial sample size for a Kaplan-Markov 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.

Director, Esparto Community Service District, Yolo County

Voters could select up to $f = 2$ candidates.

1 precinct; 988 registered voters; 187 ballots cast.

Reg. voters	ballots	Jordan	Pomeroy	Fescenmeyer	Moreland	under votes	over votes
988	187	95	80	64	62	57	8

Esparto, contd.

The smallest margin $80 - 64 = 16$ votes.

Did not have CVRs so could not compute sharp error bounds.
Pessimistic assumption: error bound 0.125 for every ballot.

Initial sample 32 ballots, for risk-limit 25%.

If mean error bound for sample held for all 187, then:

23 ballots would have sufficed to limit the risk to 25%.

32 ballots would give risk-limit 14.2%.

What do we need for efficient audits?

Laws that allow/require risk-limiting audits (such as CA AB 2023), but mostly . . .

Data plumbing:

Structured, small batch data export from VTSs.

A way to associate individual CVRs with physical ballots.

Reducing counting effort is mostly about reducing batch sizes.