Size Matters: Smaller Batches Yield More Efficient Risk-Limiting Audits

Small-Batch Audit Meeting

Washington, DC

27–28 March 2010

Philip B. Stark

http://statistics.berkeley.edu/~stark

This document:

http:
//statistics.berkeley.edu/~stark/Seminars/smallBatch10.pdf

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.

Bigger reductions are possible when batches are smaller: Need data plumbing and vote tabulation systems (VTSs) that export machine-readable subtotals for small batches—ideally, individual ballots (CVRs).

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 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.

3

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—or the batches are big.

Don't stop counting until that chance is small!

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 audit can greatly reduce burden when the outcome is right.

For actual 62 batches (31 IP, 31 VBM), audit required counting 1,437 ballots (11.33% of ballots cast)

For risk-limit 10%, would need to look at CVRs for 6 ballots. 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. Less than 0.1% of ballots cast–one tenth of one percent.

(Assuming sufficiently few errors were found.)

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

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

Jelly Beans

100 4oz bags of various flavors of jelly beans—25lbs in all. Some assorted flavors, some a single flavor.

Want to estimate how many coconut beans there are in all.

- pour the 100 bags into a large pot and stir well. Draw 4oz without looking. Estimate the total number of coconut jelly beans to be the number in the sample, times 100.

- Select 1 of the 4oz bags at random. Estimate the total number of coconut jelly beans to be the number in that bag, times 100.

Jelly Beans, contd.

Both estimates are *unbiased* but the first has much lower *variability*.

Mixing disperses the coconut jelly beans pretty evenly throughout the pot. The sample is likely to contain coconut jelly beans in roughly the same proportion as the 100 bags do overall, so multiplying the number in the sample by 100 gives a reasonably reliable estimate of the total.

Single bag is quite likely to have coconut jelly beans in a proportion quite different from the overall proportion, so multiplying that number by 100 could easily be far from the total number of coconut jelly beans among the 100 bags.

More efficient to mix the beans before selecting 4oz. Then 4oz suffices to get a reasonably reliable estimate.

How salty is the stock?

100 12-ounce cans of stock, of a variety of brands, styles, and types: chicken, beef, vegetable, low salt, regular, etc.

We want to know how much salt there is in all 1,200 ounces of stock.

- Open all the cans, pour the contents into a cauldron, stir well, and remove a tablespoon of the mix. Assay that tablespoon, multiply by the total number of tablespoons in the 100 cans (1T = 0.5oz, so the total number of tablespoons in the 100 cans is $12 \times 100 \times 2 = 2,400T$).

- Select a can at random, determine the amount of salt in that can, and multiply by 100.

Salty stock, contd.

The single tablespoon is extremely likely to have some stock from all 100 cans. The salt is likely to be spread out quite evenly through all the stock in the cauldron; it is very unlikely that the tablespoon will consist almost entirely of salt or almost entirely of water. Rather, the tablespoon is likely to contain salt in roughly the same concentration as the 100 cans do on the whole.

Single can selected at random can be quite likely to contain salt in a concentration quite different from the 1,200 ounces of stock as a whole, unless all the cans have nearly identical concentrations of salt.

Connection to Election Auditing

A vote-tabulation error is like a coconut jelly bean or a fixed quantity of salt.

A precinct or other audit batch is like a bag of jelly beans or a can of stock.

Drawing the audit sample is like selecting a bag or a 4oz scoop of jelly beans or a tablespoon or can of stock.

1st approach is like single-ballot or small-batch auditing. All the ballots are mixed together well.

2nd approach is like auditing using precincts or other large batches of ballots.

Numerical Examples

50,000 ballots, 500 ballots cast in each of 100 precincts.

1,000 (i.e., 2%) tallied incorrectly; 49,000 tallied correctly.

- What is the chance the sample contains any misinterpreted ballots at all?

- What is the chance that the percentage of misinterpreted ballots in the sample is at least half the percentage of misinterpreted ballots in the contest as a whole?

Sampling schemes

- draw a single precinct (batch) of 500 ballots at random

- divide the precincts into 10 batches of 50 ballots each and draw 10 batches at random without replacement from the resulting 1,000 batches

- draw a simple random sample of 500 ballots at random.

Chance of finding at least one misinterpreted ballot

errors	1 pct	10 batches of 50	SRS of 500
10 in every pct	100%	100%	99.996%
10 in 98 pcts, 20 in 1 pct	99%	~100%	99.996%
20 in 50 pcts	50%	99.9%	99.996%
250 in 4 pcts	4%	33.6%	99.996%
500 in 2 pcts	2%	18.4%	99.996%

Phrased differently: if the sample has no misinterpreted ballots, the "confidence" that no more than 2% of the 50,000 ballots were misinterpreted is:

sampling method	1 pct	10 batches of 50	SRS of 500
"confidence"	2%	18.4%	99.996%

Chance percentage of misinterpreted ballots in the sample is at least 1%

errors by pct.	1 pct.	10 batches of 50	SRS of 500
10 in every pct	100%	100%	97.2%
10 in 98 pcts, 20 in 1 pct	99%	~100%	97.2%
20 in 50 pcts	50%	62.4%	97.2%
250 in 4 pcts	4%	5.7%	97.2%
500 in 2 pcts	2%	18.4%	97.2%

Phrased differently: if the sample has 1% misinterpreted ballots, the "confidence" that no more than 2% of the 50,000 ballots were misinterpreted is:

sampling method	1 pct	10 batches of 50	SRS of 500
"confidence"	2%	18.4%	97.2%

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 largely about reducing batch sizes.