Efficient Post-Election Audits of Multiple Contests: 2009 California Tests

Conference on Empirical Legal Studies
University of Southern California
Gould School of Law
Los Angeles, CA
20–21 November 2009

Philip B. Stark

Problem: Any way of counting votes makes mistakes.

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

If there's a complete, accurate audit trail, 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.

Current audit laws have the wrong focus: Virtually useless for fixing wrong outcomes.

Efficiency is primarily about batch sizes: Need data plumbing.

Wrong Focus

Current and proposed laws focus on how big an initial sample to draw.

Heated debates over fixed percentages, tiered percentages depending on the margin, or sample sizes that vary continuously with the margin and depend on batch sizes.

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

Can't fix wrong outcomes without counting the whole audit trail.

Risk-Limiting Audits

> If the outcome is wrong, there's a at least a [pre-specified] minimum chance of a full manual count,
> no matter what caused the outcome to be wrong.

The *risk* is the maximum chance that there won't be a full hand count when a full hand count would show that the apparent outcome is wrong.

"Wrong" means disagrees with what a full hand count would show: presupposes accurate & complete audit trail, secure chain of custody, etc. Nontrivial.

Null hypothesis: outcome is wrong.
Control Type I error rate.

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

Persistent idea that only the initial sample matters, not the errors the sample finds.
E.g., Holt bill.

Essential that voters create complete, durable, accurate audit trail.

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

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

Need a plan for dealing with discrepancies, possibly leading to full count. "Explaining" or "resolving" isn't enough.

Current audit laws do not limit risk.

Compliance audits vs. materiality audits.

Assessing Evidence

How strong is the evidence that the outcome is correct, given how the sample was drawn, the margin, the errors found, etc.?

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

(The definition of "little" differs across sampling methods, etc.)

P-value of the hypothesis that the apparent outcome of one or more contests is wrong.

Sampling Designs

Simple

Stratified (by county, voting method, other)

PPEB

NEGEXP

Stratified PPEB?

Sampling scheme affects choice of test statistic—analytic tractability

Weighted max, binning for simple & stratified sampling, NEG-EXP, PPEB.

More efficient choices possible for PPEB: Kaplan-Markov

Taint

e_p: error in batch p (max % overstatement of any margin)
u_p: upper bound on e_p; $U = \sum_p u_p$.

The *taint* of batch p is

$$\tau_p = \frac{e_p}{u_p} \leq 1. \tag{1}$$

Suppose draw batches with replacement s.t., in each draw,

$$\mathbb{P}\{\text{draw batch } p\} = u_p/U. \tag{2}$$

Taint of jth draw is T_j.

$\{T_j\}$ are iid, $\mathbb{E}T_j = E/U$.

Can stop the audit if can reject the hypothesis $\mathbb{E}T_j \geq 1/U$.

Hypothesis about the mean of a bounded random variable.

Sequential risk-limiting audit using Kaplan-Markov bound

0. Calculate error bounds $\{u_p\}$, U. Set $n = 1$. Pick $\alpha \in (0, 1)$ and $m > 0$.

1. Draw a batch using PPEB. Audit it if it has not already bee audited.

2. Find $T_n \equiv t_p \equiv e_p/u_p$, taint of the batch p drawn at stage n.

3. Compute

$$P_n \equiv \prod_{j=1}^{n} \frac{1 - 1/U}{1 - T_j}. \tag{3}$$

4. If $P_n < \alpha$, stop; report apparent outcome. If $n = m$, audit remaining batches. If all batches have been audited, stop; report known outcome. Else, $n \leftarrow n + 1$ and go to 1.

This sequential procedure is risk-limiting:

If outcome is wrong,

$$\mathbb{P}\{\text{stop without auditing every batch}\} < \alpha.$$

Chance $\geq 1 - \alpha$ of fixing wrong outcome by full hand count.

Remarkably efficient if batches are not too big.

Pilot Audits in California

Marin County (February 2008; November 2008, 2009)

Yolo County (November 2008, 2009)

Santa Cruz County (November 2008)

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

Contest sizes 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 (IP) ballots were tabulated separately (62 batches).

$U = 3.0235$.

For $\alpha = 10\%$, 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 $\alpha = 10\%$ would need to look at CVRs for $n = 6$ ballots.

For $\alpha = 1\%$, $n = 12$ ballots.

C.f., 1,437 ballots 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 u_ps.
Pessimistic assumption $u_p = 0.125$ for every ballot.
$U = 187 \times 0.125 = 23.375$.

Initial sample $n = 32$ ballots, for $\alpha = 25\%$.

If mean u_p for sample were true for all 187, $U = 16.874$. Then:

$n = 23$ would have sufficed to limit the risk to $\alpha = 25\%$.

$n = 32$ would give $\alpha = 14.2\%$.

What do we need for efficient audits?

Laws that allow/require risk-limiting audits, 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.

Extra slides (time is unlikely to permit)

Notation

N batches (possibly single ballots), C contests.

Contest c has K_c "candidates," votes for up to f_c candidates.

Reported vote for candidate k in batch p is v_{kp}

$V_k \equiv \sum_{p=1}^{N} v_{kp}$.

$\mathcal{W}_c$: indices of apparent winners of contest c.

$\mathcal{L}_c$: indices of apparent losers of contest c.

Reported margin of $w \in \mathcal{W}_c$ over $\ell \in \mathcal{L}_c$:

$$V_{w\ell} \equiv V_w - V_\ell > 0. \tag{4}$$

More notation

Actual vote for candidate k in batch p is a_{kp}.

$A_k \equiv \sum_{p=1}^{N} a_{kp}$.

Actual margin of $w \in \mathcal{W}_c$ over $\ell \in \mathcal{L}_c$:

$$A_{w\ell} \equiv A_w - A_\ell. \tag{5}$$

Apparent winners of all C contests are the true winners iff

$$\min_{c \in \{1,\ldots,C\}} \min_{w \in \mathcal{W}_c, \ell \in \mathcal{L}_c} A_{w\ell} > 0. \tag{6}$$

Still more notation . . .

For $w \in \mathcal{W}_c$, $\ell \in \mathcal{L}_c$, define

$$e_{pw\ell} \equiv \begin{cases} \frac{(v_{wp}-v_{\ell p})-(a_{wp}-a_{\ell p})}{V_{w\ell}}, & \text{if ballots in batch } p \text{ contain contest } c \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

If any apparent outcome is wrong,

$$\exists\ (c \in \{1, \ldots, C\},\ w \in \mathcal{W}_c,\ \ell \in \mathcal{L}_c) \text{ with}$$

$$\sum_{p=1}^{N} e_{pw\ell} \geq 1. \tag{8}$$

MACRO

Maximum across-contest relative overstatement in batch p:

$$e_p \equiv \max_{c \in \{1,\ldots,C\}} \max_{w \in \mathcal{W}_c, \ell \in \mathcal{L}_c} e_{pw\ell}. \tag{9}$$

Now

$$\max_{c \in \{1,\ldots,C\}} \max_{w \in \mathcal{W}_c, \ell \in \mathcal{L}_c} \sum_{p=1}^{N} e_{pw\ell} \le \sum_{p=1}^{N} \max_{c \in \{1,\ldots,C\}} \max_{w \in \mathcal{W}_c, \ell \in \mathcal{L}_c} e_{pw\ell} \tag{10}$$

$$= \sum_{p=1}^{N} e_p \equiv E. \tag{11}$$

E is *maximum across-contest relative overstatement* (MACRO).

If $E < 1$, all C apparent outcomes are right.

Controlling the familywise error rate

C null hypotheses,

the outcome of contest c *is incorrect,* $c = 1, \ldots, C$.

If $E < 1$, the entire family of C null hypotheses is false: all apparent outcomes are right.

Test of hypothesis $E \geq 1$ at significance level α is a test of the C hypotheses with familywise error rate no larger than α.

Bounding e_p

If number of valid ballots cast in batch p for contest c is at most b_{cp} then

$$e_{pw\ell} \leq (v_{wp} - v_{\ell p} + b_{cp})/V_{w\ell}, \tag{12}$$

and so

$$e_p \leq \max_{c \in \{1,\ldots,C\}} \max_{w \in \mathcal{W}_c, \ell \in \mathcal{L}_c} \frac{v_{wp} - v_{\ell p} + b_{cp}}{V_{w\ell}} \equiv u_p. \tag{13}$$

u_p is a limit on the relative overstatement of *any* margin that can be concealed in batch p, the MACRO in batch p.

$U \equiv \sum_p u_p$, bound on total error.

Cartoon

	precincts	batches	ballots	winner	loser	margin	IP batches winner	IP batches loser	VBM batches winner	VBM batches loser
A	200	400	120,000	60,000	54,000	6,000	200	180	100	90
B	100	200	60,000	30,000	24,000	6,000	200	160	100	80
C	60	120	36,000	18,000	12,600	5,400	200	140	100	70

Contest A: entire jurisdiction, 200 precincts.

Contest B: 100 precincts.

Contest C: 60 precincts; 30 of those are also in contest B.

Each precinct is divided into two batches, 400 ballots cast in-precinct (IP) and 200 ballots cast by mail (VBM).

Valid votes, undervotes, and invalid ballots.

Cartoon, contd.

		FWER				PCER			
	U	n	expected batches	expected ballots	expected votes	n	expected batches	expected ballots	expected votes
A	21.00	52	48.49	16,074.23	16,074.23	33	31.58	10,488.77	10,488.77
B	11.00	28	26.01	8,615.69	8,615.69	17	16.27	5,402.16	5,402.16
C	7.67	19	17.50	5,795.81	5,795.81	12	11.41	3,787.51	3,787.51
all			85.13	28,038.26	30,485.73		56.38	18,649.98	19,678.44
M	22.72	36	34.30	11,387.29	20,617.68				

Independent and simultaneous audits controlling FWER and PCER to risk $\alpha = 25\%$.
The bottom row is MACRO

Cartoon, contd.

α	Single-ballot audit sharp	Single-ballot audit conservative	Batch audit
25%	39.99	60.98	9,878.64
10%	66.97	101.96	16,065.45
1%	132.90	202.83	29,566.79

Expected initial sample sizes, in ballots.