Risk-limiting post-election audits

Department of Statistics
Kansas State University
Manhattan, KS
2 October 2008

Philip B. Stark
Department of Statistics
University of California, Berkeley
statistics.berkeley.edu/∼stark

[Election Leak]

Abstract: The apparent margin in an election can be inated by machine error, programming error, processing error, voter error or even deliberate fraud. Is the outcome still right?

Post-election audits hand-count ballots in a random sample of batches. Eighteen states require or allow post election audits; New Jersey is the latest. Oregon and other states are poised to require them. Generally, the state-mandated audits do not control the risk of certifying an incorrect election outcome. They do not guarantee any minimum chance that there will be a full manual count when the apparent outcome is wrong.

I will present a method that does. The method has been tested on data from a 2006 U.S. Senate race in MN and "live" on a 2008 ballot measure in Marin County, CA. Between 3 and 5 California counties will be trying the method this November.

The method couches the problem of conrming an election outcome as a statistical hypothesis test, and tests the hypothesis sequentially. Data are collected. If the data do not allow one to reject the null hypothesis is that the apparent outcome is not the outcome a full recount would show, the sample is enlarged. Eventually, either the hypothesis has been rejected or there has been a full manual count. Multiplicity is taken into account by adjusting the conditional levels of the sequential test.

Outline

- Voting systems: punchcard, optically scanned, DRE (VVPAT)
- Sample of sorrows: DC 2008, NJ 2008, OH 2004, FL 2004, CA 2004
- Laws: California, New Jersey
- Mechanical random selection
- Hypothesis testing framework: the math
- The realities
- Examples: 2006 MN Senate race; 2008 Marin Measure A.
- Complications & potential improvements
- References

Voting Systems

Punchcard & lever systems. Discouraged by *Help America Vote Act* of 2002. NY still uses—but not for much longer.

Optically scanned ballots: "bubble in" like a Scantron form. Produces auditable paper trail. Voter intent vs. machine scan.

Direct-recording Electronic (DRE): touchscreens, etc. VVPATs. Felten group, TTBR. De-certified in CA, CO, OH.

Washington, DC, 2008

Report Blames Speed In Primary Vote Error; Exact Cause of Defect Not Pinpointed by Nikita Stewart

Speed might have contributed to the Sept. 9 primary debacle involving thousands of phantom votes, according to a D.C. Board of Elections and Ethics report issued yesterday. . . . [T]he report does not offer a definitive explanation. . .

The infamous Precinct 141 cartridge "had inexplicably added randomly generated numbers to the totals that had been reported," according to the report written by the elections board's internal investigative team.

. . . 4,759 votes were reflected instead of the actual 326 cast there.

Washington Post, 2 October 2008; Page B02

New Jersey 2008

County finds vote errors: Discrepancies discovered in 5% of machines by *Robert Stern*

Five percent of the 600 electronic voting machines used in Mercer County during the Feb. 5 presidential primary recorded inaccurate voter turnout totals, county officials said yesterday . . .

23 February 2008, New Jersey Times

New Jersey 2008–contd.

Judge Suppresses Report on Voting Machine Security
by Andrew Appel

A judge of the New Jersey Superior Court has prohibited the scheduled release of a report on the security and accuracy of the Sequoia AVC Advantage voting machine. . . . [NJ] mostly uses Sequoia AVC [DRE] models. None of those DREs can be audited: they do not produce a voter verified paper ballot.

2 October 2008, Freedom to Tinker

Ohio 2004

Machine Error Gives Bush Thousands of Extra Ohio Votes by *John McCarthy*

COLUMBUS, Ohio – An error with an electronic voting system gave President Bush 3,893 extra votes in suburban Columbus, elections officials said. Franklin County's unofficial results had Bush receiving 4,258 votes to Democrat John Kerry's 260 votes in a precinct in Gahanna. Records show only 638 voters cast ballots in that precinct. Bush's total should have been recorded as 365.

5 November 2004, Associated Press

Florida 2004

Broward Machines Count Backward by *Eliot Kleinberg*

. . . Early Thursday, as Broward County elections officials wrapped up after a long day of canvassing votes, something unusual caught their eye. Tallies should go up as more votes are counted. Thats simple math. But in some races, the numbers had gone . . . down.

Officials found the software used in Broward can handle only 32,000 votes per precinct. After that, the system starts counting backward.

. . . The problem cropped up in the 2002 election. . . . Broward elections officials said they had thought the problem was fixed.

5 November 2004, The Palm Beach Post

California 2004

Lost E-Votes Could Flip Napa Race by *Kim Zetter*

Napa County in Northern California said on Friday that electronic voting machines used in the March presidential primary failed to record votes on some of its paper ballots, which will force the county to re-scan over 11,000 ballots and possibly change the outcome of some close local races.

. . . Napa Registrar of Voters John Tuteur said they discovered the problem on Thursday while conducting a manual recount of 1 percent of precincts, . . . they discovered that the machine wasn't recording certain votes.

. . . the machine was calibrated to detect carbon-based ink, but not dye-based ink commonly used in gel pens, . . . a Sequoia technician ran test ballots through the machine to calibrate its reading sensitivity, but failed to test for gel ink.

12 March 2004, Wired News

Machine (Voting System) Counting

- Want to count votes by machine: saves time and money (or so we are told).

- Machine counts are subject to various kinds of error. (So are hand counts, but they're the gold standard. Progress on accuracy, too.)

- Counting errors $\Rightarrow$ risk that machines name the wrong winner.

Statistical Audits

Can *limit* and *quantify* that risk.

Could guarantee that,

> If the outcome is wrong, there's a 99% chance of a full manual count

even if an evil adversary built the hardware and wrote the software.

(Of course, could just manually count 99% of *all* contests at random, but that's a lot of counting: avoidable by statistics.)

Essential that voters create an audit trail.

Essential to select batches at random.

California Elections Code §15360

. . . the official conducting the election shall conduct a public manual tally of the ballots tabulated by those devices, including absent voters' ballots, cast in 1 percent of the precincts chosen at random by the elections official . . .

The elections official shall use either a random number generator or other method specified in regulations . . .

The official conducting the election shall include a report on the results of the 1 percent manual tally in the certification of the official canvass of the vote. This report shall identify any discrepancies between the machine count and the manual tally and a description of how each of these discrepancies was resolved . . .

NJ S507 [1R] (Gill)

. . . shall conduct random hand counts of the voter-verified paper records in at least two percent of the election districts where elections are held for federal or State office . . .

Any procedure designed, adopted, and implemented by the audit team shall be implemented to ensure with at least 99% statistical power that for each federal, gubernatorial or other Statewide election held in the State, a 100% manual recount of the voter-verifiable paper records would not alter the electoral outcome reported by the audit . . .

[procedures] shall be based upon scientifically reasonable assumptions . . . including but not limited to: the possibility that within any election district up to 20% of the total votes cast may have been counted for a candidate or ballot position other than the one intended by the voters . . .

Say what?

Selecting batches at random

Software pseudo-random number generators:
not transparent, hackable.

One ticket per precinct:
hard to verify; hard to mix (Vietnam draft).

10-sided dice (Marin County) [Roll 1] [Roll 2]

Ping-pong balls (Alameda County) [Static] [Tumbling]

Alameda has 1204 precincts. Pick 1s digit, 10s, 100s.
If result is between 205 and 999, stop.
Else, remove 2–9 & pick 1000s digit.

Unintended consequences?

How to commit election fraud (if you must)

- make sure the election uses DREs w/o VVPATs; hack the software.

- if the jurisdiction uses DREs w/ VVPATS, hack the software and spoil the VVPATs with "household chemicals" (TTBR report)

- if you know that the audit will be based on whether any errors are found in a simple random sample, hide the fraud in as few precincts as possible. (But in Alameda County, CA, avoid precincts 205–1000.)

- target a jurisdiction where audits are illegal

General principles

Margin small $\Rightarrow$ less error required to produce it erroneously.

Sample small $\Rightarrow$ can be likely that sample will find few or no errors, even if machines named the wrong winner.

No look, no see: absence of evidence is not evidence of absence.

Smaller margins $\Rightarrow$ lower confidence.

Smaller samples $\Rightarrow$ lower confidence.

Larger discrepancies in sample $\Rightarrow$ lower confidence.

Sample big (compared with margin) $\Rightarrow$ likely to see big discrepancies in the sample if machines named wrong winner.

Rigorous statistical audit

If it's very likely that the audit would have found larger discrepancies than it did find, had the machines named the wrong winner, confirm the outcome.

Otherwise, keep counting.

If the outcome is confirmed, either the correct winner was named, or something very unlikely happened.

Complete procedure says:

- how many batches to audit initially
- given the discrepancies in the audit sample, whether to confirm the outcome or expand the audit
- eventually declares "outcome confirmed" or "full recount."
- change of full recount if outcome is wrong is at least 99%, e.g.

Only one basic approach so far does that.

Notation

f	# winners (vote for f)
P	# audit batches in the contest
K	# candidates in contest, after "pooling"
$\mathcal{K}_w$	indices of the f apparent winners
$\mathcal{K}_\ell$	indices of the $K - f$ apparent losers
a_{kp}	actual vote for candidate k in batch p
$A_k \equiv \sum_p a_{kp}$	actual total vote for candidate k
$A_{w\ell} \equiv A_w - A_\ell$	actual margin of candidate w over candidate ℓ
b_p	upper bound on a_{kp}
v_{kp}	reported vote for candidate k in batch p
$V_k \equiv \sum_p v_{kp}$	total vote reported for candidate k
$V_{w\ell} \equiv V_w - V_\ell$	apparent margin of candidate w over candidate ℓ

Sufficient condition for correct outcome:

The apparent winners are the actual winners if

$$\min_{w \in \mathcal{K}_w, \ell \in \mathcal{K}_\ell} A_{w\ell} > 0. \tag{1}$$

Define

$$e_{w\ell p} \equiv \frac{(v_{wp} - v_{\ell p}) - (a_{wp} - a_{\ell p})}{V_{w\ell}}. \tag{2}$$

Outcome must be right unless

$$\sum_{p=1}^{P} e_{w\ell p} \geq 1 \text{ for some } w \in \mathcal{K}_w,\ \ell \in \mathcal{K}_\ell. \tag{3}$$

Maximum relative overstatement of pairwise margins (MRO) in batch p:

$$e_p \equiv \max_{w \in \mathcal{K}_w, \ell \in \mathcal{K}_\ell} e_{w\ell p}. \tag{4}$$

Bounding the error in each batch

$$\max_{w\in\mathcal{K}_w,\ell\in\mathcal{K}_\ell}\sum_{p=1}^{P} e_{w\ell p} \le \sum_{p=1}^{P}\max_{w\in\mathcal{K}_w,\ell\in\mathcal{K}_\ell} e_{w\ell p} = \sum_{p=1}^{P} e_p. \qquad (5)$$

b_p: bound on a_{kp} from pollbooks, # registered voters, ballot accounting, etc.

$$e_p \le \max_{w\in\mathcal{K}_w,\ell\in\mathcal{K}_\ell} \frac{v_{wp} - v_{\ell p} + b_p}{V_{w\ell}} \equiv u_p. \qquad (6)$$

The whole shebang

1. Pick the min chance β of full manual count when result is wrong

2. Pick the max # of stages S, "escalation probabilities" $\beta_1, \beta_2, \ldots, \beta_S$ s.t. $\pi_s \beta_s = \beta$.

3. Select subtotals that comprise batches, & strata.
   B_c is # batches in stratum c, $c = 1, \ldots, C$.

4. Fnd upper bounds b_p on the number of votes per candidate per batch from voter registrations, pollbooks, or an accounting of ballots.

5. Set $s = 1$ (stage).
   $P_s = P$ (un-audited batches at stage s).

6. Find pairwise margins:

$$V_{w\ell} = \text{ (votes for winner } w) \; - \; \text{(votes for loser } \ell). \quad (7)$$

   Use the semi-official results for the P_s batches that have not yet been audited, and the audit results for the $P - P_s$ audited batches.

   If $\min_{w \in \mathcal{K}_w, \ell \in \mathcal{K}_\ell} V_{w\ell} \leq 0$, the list of winners has changed. Abort the audit and count all the votes by hand.

7. For each batch p that has not been audited, compute

$$u_p \equiv \max_{w \in \mathcal{K}_w, \ell \in \mathcal{K}_\ell} \frac{V_{wp} - V_{\ell p} + b_p}{V_{w\ell}}. \tag{8}$$

8. If there are a few un-audited batches p with especially large u_p audit them and return to step 6.

9. Set the tolerable level of error, $t \in [0, 1)$. If any margin is overstated by t or more, the audit will progress to the next stage.

10. Find the incremental sample sizes.

    For the P_s batches p not yet audited, define:
    $t_p \equiv \min(t, u_p)$; $T \equiv \sum_p T_p$; $\tilde{u}_p = u_p - t_p$.

    (a) Starting with the largest value of $\tilde{u}_p$, add successively smaller values of $\tilde{u}_p$ just until the sum of those values is $\geq 1 - T$. q is # terms in the sum.

    (b) Find the smallest whole number n such that

$$\left(\frac{P_s - q}{P_s}\right)^n \leq 1 - \beta_s. \tag{9}$$

    (c) Sample size n_c for stratum c is

$$n_c \equiv \left\lceil n \times \frac{\#\text{unaudited batches in stratum } c}{P_s} \right\rceil. \tag{10}$$

$$n^* = n_1 + n_2 + \cdots + n_C \geq n. \tag{11}$$

11. Select batches using a transparent, mechanical, verifiable source of randomness, such as fair 10-sided dice. For each stratum $c = 1, \ldots, C$, draw n_c batches from the as-yet-unaudited batches in stratum c, count votes by hand.

12. For each of the n^* batches p audited in this stage, find

$$e_{w\ell p} = \frac{v_{wp} - v_{\ell p} - (a_{wp} - a_{\ell p})}{V_{w\ell}} \tag{12}$$

for all pairs (w, ℓ) of semi-official winners w and losers ℓ. There are $n^* \times w \times \ell$ of those values.

$$t_s \equiv \max_{p,w,\ell} e_{w\ell p}. \tag{13}$$

13. If $t_s \leq t$, certify the election and stop. If $t_s > t$ and $s = S$, count all the votes by hand. Otherwise, increment s; perform any desired targeted auditing; set P_s tto # batches not yet audited; return to step 6.

Hypothetical example: cartoon of U.S. House Race

2 stages.

400 precincts split across 2 counties: 300 and 100.

Stratify by mode of voting (in-precinct or by mail) and county: 4 strata, sizes 300, 300, 100 and 100 batches

3 candidates, overvotes, undervotes.

t corresponds to 3 vote overstatement of margin of victory.

Audit batches equal size (255 votes), equal numbers of reported votes including 13 votes reported for candidate 3, 2 overvotes and 3 undervotes.

Worst-case erroneous escalation: all "tainted" in one stratum. (Random taint much less likely to trigger escalation.)

(1) V_{12}	(2) β	(3) β_1	(4) β_2	(5) n	(6) n^*	(7) f	(8) $\gamma_{0.01}$	(9) $\omega_{0.01}$	(10) $\gamma_{0.005}$	(11) $\omega_{0.005}$	(12) n_2^*
5.2%	75%	76.0%	98.9%	37	38	4.75%	34.7%	23.8%	18.8%	7.2%	108
		86.6%	86.6%	51	54	6.75%	45.3%	17.7%	25.5%	4.9%	50
	90%	91.0%	98.9%	61	62	7.75%	50.0%	34.9%	28.7%	11.3%	108
		94.9%	94.9%	76	78	9.75%	58.3%	31.2%	34.8%	9.5%	68
10.0%	75%	76.0%	98.9%	18	20	2.50%	22.3%	9.3%	11.6%	2.4%	54
		86.6%	86.6%	25	28	3.50%	28.7%	6.3%	15.3%	1.5%	26
	90%	91.0%	98.9%	29	30	3.75%	28.7%	12.1%	15.3%	3.1%	54
		94.9%	94.9%	36	38	4.75%	34.7%	11.2%	18.8%	2.8%	36
19.6%	75%	76.0%	98.9%	9	12	1.50%	15.4%	4.0%	7.9%	1.0%	28
		86.6%	86.6%	13	14	1.75%	15.4%	2.1%	7.9%	0.5%	14
	90%	91.0%	98.9%	15	16	2.00%	15.4%	4.0%	7.9%	1.0%	30
		94.9%	94.9%	18	20	2.50%	22.3%	4.5%	11.6%	1.1%	20

(1) Margin between candidates 1 and 2. 5.2%: 125 votes for candidate 1 and 112 for candidate 2 in each batch; 10.0%: 131 versus 106; 19.6%: 143 versus 94. (2) Min chance of full count if the outcome is wrong. (3) Min chance audit goes from stage 1 to stage 2 if the outcome is wrong. (4) Min chance audit goes from stage 2 to a full count if the outcome is wrong, if it gets to stage 2. (5) Stage 1 sample size before adjusting for stratification. (6) Stage 1 sample size adjusted for stratification. (7) $n^*/800 \times 100\%$. (8) Max chance audit gets to stage 2 if 1% of audit batches overstate V_{12} by more than 3 votes. (9) Max chance of a full count if 1% of audit batches overstate V_{12} by more than 3 votes, and the stage 1 net error is zero. (10) Same as (8), but 0.5% of batches have large overstatements of V_{12}. (11) Same as (9), but 0.5% of batches have large overstatements of V_{12}. (12) Stage 2 sample size if the net error in stage 1 is zero.

Logistical issues: stratification, etc.

Samples for different counties drawn independently: stratified.

VBM and absentee ballots not counted right away.

Makes sense to start with a uniform sampling rate, then escalate as necessary.

Can test separately in each stratum for proportional share of M.

Reject overall hypothesis if all reject; conservative.

OR, P-value for proportional sample $\leq$ P-value for unstratified sample w/ replacement.

November 2006 Minnesota U.S. Senate Race

MN requires:
Counties with <50,000 registered voters audit ≥ 2 precincts;
counties with 50,000–100,000 registered voters audit ≥ 3;
counties with $\geq$100,000 registered voters audit ≥ 4.

≥ 1 precinct audited in each county must have $\geq$150 votes cast.

$C = 87$ counties, $P = 4,123$ precincts, $n = 202$ audited.
Audited precincts had from 2–2,393 ballots cast.

Voters	under& invalid	Fitzgerald Indep	Kennedy Repub	Klobuchar D/F/L	Cavlan Green	Powers Constit	Write-ins
2,217,818	15,099	71,194	835,653	1,278,849	10,714	5,408	901
$V_{w\ell}$	N/A	1,207,655	443,196	N/A	1,268,135	1,273,441	1,277,948

Pool Cavlan, Powers, write-ins: pseudo-candidate apparently lost to Klobuchar by 1,261,773 votes; $K = 4$.
$\max_p u_p = 0.0097$; $\max_p e_p = 4.5 \times 10^{-6}$; $q = 166$.

Conservative P-value

Pretend sample was drawn with replacement from all 4,123 precincts, but that only 78 precincts were drawn, as if the population sampled using the minimum sampling fraction among counties

1.9% sample ($n = 78$) w/ replacement	SRS ($n = 202$) w/o replacement
4.05%	0.02%

Sharper treatment of stratification (with Mike Higgins) decreases conservative P-value to 1.9%.

5 February 2008 Marin County Measure A

First election ever audited to attain target level of confidence in the result.

Audited to attain 75% confidence that a full manual recount would find the same outcome.

Required 2/3 majority to pass. Margin 298 votes.

Stratified random sample: 6 polling-place counts, 6 VBM counts.

Marin Measure A data

precinct	registered	type	ballots	yes	no	bound	audited
2001	1326	IP	391	278	101	286	yes
		VBM	657	438	193	456	no
2004	893	IP	284	204	66	214	yes
		VBM	389	257	116	268	yes
2010	6	VBM	6	4	2	4	no
2012	740	IP	218	167	43	173	yes
		VBM	342	242	89	250	no
2014	983	IP	299	214	75	221	no
		VBM	420	306	95	319	yes
2015	905	IP	217	167	44	171	yes
		VBM	483	332	131	346	yes
2019	1048	IP	295	215	70	222	yes
		VBM	567	395	160	403	yes
2101	923	IP	265	169	79	181	no
		VBM	439	275	133	296	yes
2102	900	IP	223	144	68	152	yes
		VBM	410	233	142	257	yes
All	7724	PRO	252	176	54	191	no

Marin Measure A audit timeline

Milestone	Date
Election day	5 February
Polling place results available	7 February
Random selection of polling place precincts	14 February
VBM results available	20 February
Random selection of VBM precincts	20 February
Hand tally complete	20 February
Provisional ballot results available	29 February
Computations complete	3 March

Costs:

$1,501, including salaries and benefits for 4 people tallying the count, a supervisor, support staff to print reports, resolve discrepancies, transport ballots and locate and retrieve VBM ballots from the batches in which they were counted.

$0.35 per ballot audited. $1\frac{3}{4}$ days.

Other stuff

Expanding test to 3–5 California counties in November.

PPEB and connection to financial auditing (with Luke Miratrix).

False Discovery Rate.

Small races? Lower confidence? Only audit random sample of races?

Sharper treatment of stratification (with Mike Higgins)

Auditing entire ballots, not contests.

Recap

- Vote counting is not perfect; errors can affect outcomes
- Auditing laws that address the problem fall short
- There's a way to fix them using Statistics
- It seems practical/workable in examples

References

California Secretary of State Debra Bowen Voting System Review page. http://www.sos.ca.gov/elections/elections_vsr.htm

Aslam, J.A., R.A. Popa and R.L. Rivest, 2007. *On Auditing Elections When Precincts Have Different Sizes*, Computer Science and Artificial Intelligence Laboratory, MIT. http://people.csail.mit.edu/rivest/AslamPopaRivest-OnAuditingElectionsWhenPrecinctsHaveDifferentSizes.pdf

California Voter Foundation http://www.calvoter.org

Felten group at Princeton University: http://itpolicy.princeton.edu/voting/

Ginnold, E., J.L. Hall and P.B. Stark, 2008. *A confidence-driven audit of Measure A in the February 2008 Marin County election*, in preparation.

Norden, L., A. Burstein, J.L. Hall and M. Chen, 2007. *Post-election audits: restoring trust in elections*, Brennan Center. http://www.brennancenter.org/dynamic/subpages/download_file_50089.pdf

Jefferson, D., K. Alexander, E. Ginnold, A. Lehmkuhl, K. Midstokke and P.B. Stark, 2007. *Post-Election Audit Standards Working Group: Report to California Secretary of State Debra Bowen*. http://www.sos.ca.gov/elections/peas/final_peaswg_report.pdf

Stark, P.B., 2008. CAST: Canvass audits by sampling and testing. http://statistics.berkeley.edu/~stark/Preprints/cast08.pdf

Stark, P.B., 2008. A sharper discrepancy measure for post-election audits, *Annals of Applied Statistics*, in press. http://statistics.berkeley.edu/~stark/Preprints/pairwise08.pdf

Stark, P.B., 2008. Conservative Statistical Post-Election Audits, *Annals of Applied Statistics, 2*, 550–581. http://arxiv.org/abs/0807.4005

Stark, P.B., 2008. *Election audits by sampling with probability proportional to an error bound: dealing with discrepancies*, working paper. http://statistics.berkeley.edu/~stark/Preprints/ppebwrwd08.pdf

More voting-related links: http://statistics.berkeley.edu/~stark/Vote/index.htm