Close enough for government [to] work:
Risk-Limiting Post-Election Audits

Philip B. Stark

Department of Statistics, UC Berkeley

12 April 2011
Joint Berkeley-Stanford Statistics Colloquium
Stanford University

Outline

News reports

Legislation

The problem

Risk-limiting audits

Ballot-level audits

Pilot audits

Conclusions

NY SD 7

Abstract

How can we tell whether votes were tallied accurately enough to determine the correct winners of an election? Risk-limiting post election audits, endorsed by the American Statistical Association, Common Cause, The League of Women Voters, Verified Voting, and other groups concerned with election integrity, can help. A *risk-limiting audit* has a guaranteed minimum chance of correcting electoral errors. There have been eight risk-limiting audits so far, seven in California and one in Colorado. California AB 2023, which requires an official pilot of risk-limiting audits, became law in July, 2010 after unanimous, bipartisan votes in both houses. I will discuss the theory behind risk-limiting audits (couching auditing as a sequential nonparametric hypothesis test about the mean of a bounded population); factors that affect efficiency and cost; "transitive auditing," which uses a shadow system rather than the system of record; and lessons learned conducting audits and working with elections officials and election integrity activists. If time permits, I'll ramble about getting AB 2023 endorsed and passed and about the recent tangle in New York Senate District 7, which shifted the balance of power in the NY senate: They are interesting case studies at the intersection of statistics, policy, legislation, public administration, jurisprudence, and politics.

[Election Leak] [CNN: DC hack] [Voting Machine Wins] [Homer Votes—sort of]

Things to keep in mind

Sufficiently advanced cluelessness is indistinguishable from malice.
(Clark's Law)

The difference between theory and practice is smaller in theory than
it is in practice. (Various)

The purpose of elections is to convince the losers that they lost.
(D. Wallach)

The purpose of election audits is to convince me that the losers lost.
(PBS)

Saguache County, Colorado crimes

Grand jury has its hands full with Saguache election case, by Troy Hooper

A disputed election in south-central Colorado is now in the hands of a grand jury that is reviewing allegations that the clerk and other officials committed crimes when they tallied the votes.

The officials under investigation stood to benefit from the election's outcome — most notably Saguache County Clerk Melinda Myers — who, along with County Commissioner Linda Joseph, at first lost but then won their races after Myers declared the races had to be retabulated due to a technical glitch.

[Myers won't let the Colorado Secretary of State inspect the ballots.] "There are processes that we are avowed to protect," [Colorado County Clerks] association president Scott Doyle said. "One of them is preserving the sanctity of ballots. The cornerstone of our democracy is based on those ballots. It's what we stand for as clerks."

"The clerks are using the false argument about 'secrecy of ballots' as a scare tactic or sympathy evoking tool to try to get a trusting public to side with them in their effort to block public verification of elections," Al Kolwicz of the Colorado Voter Group said in an email. "Why exactly clerks oppose public verification is unknown."

Officials in Saguache County stand accused of more than 30 misdemeanors.

THE COLORADO INDEPENDENT, 25 MARCH 2011,

http://coloradoindependent.com/80819/grand-jury-has-its-hands-full-with-saguache-election-case

Waukesha County, WI: Oops!

Wisconsin Election Surprise: David Prosser Gains 7,500 Votes After 'Human Error' In Waukesha County, by Amanda Terkel

In a dramatic turn of events on Thursday, the Waukesha County clerk announced that the vote total announced for Tuesday's Wisconsin Supreme Court race had been mistaken – and that the corrected numbers changed the outcome of the entire election.

There were 3,456 missing votes for Democratic-backed challenger JoAnne Kloppenburg and 11,059 for incumbent GOP-backed Justice David Prosser. Kloppenburg has previously been beating Prosser by just 200 votes of the roughly 1.5 million cast statewide.

In the city of New Berlin, the total for one ward was recorded as 37 votes for Prosser, but it was actually 237, she said. In the town of Lisbon, a "typing error" resulted in both candidates losing votes. The most significant error, however, occurred in the city of Brookfield.

"The spreadsheet from Brookfield was imported into a database that was provided by the Government Accountability Board, but it inadvertently was not saved," Nickolaus said. "As a result, when I ran the report to show the aggregate numbers that were collected from all the municipalities, I assumed that the city of Brookfield was included. It was not. The city of Brookfield cast 14,315 votes on April 5 – 10,859 votes went for Justice David Prosser, 3,456 went for JoAnne Kloppenburg."

. . . prior to the election, Nickolaus "was heavily criticized for her decision to keep the county results on an antiquated personal computer, rather than upgrade to a new data system being utilized statewide."

"Nickolaus cited security concerns for keeping the data herself . . . "

HUFFINGTON POST, 7 APRIL 2011,

http://www.huffingtonpost.com/2011/04/07/david-prosser-wisconsin-supreme-court_n_846431.html

Vote-flipping in North Carolina

NC GOP leader: Touchscreen voting machines have programming flaw, by Michael Biesecker

The chairman of the N.C. Republican Party alleged Thursday that a programming flaw with touchscreen voting machines used for early voting in 36 counties is causing votes intended for GOP candidates to be counted for Democrats.

Tom Fetzer, the Republican chairman, said that if the State Board of Elections does not enact a list of demands intended to remedy the problem by the end of today, the party's lawyers will be in federal court Friday morning seeking a statewide injunction. . . .

Johnnie McLean, deputy director of the state elections board, said Thursday that her office has received no widespread reports of problems.

"In every election we will have scattered reports of machines where the screens need to be recalibrated," McLean said. "That sort of comes with the territory with touch-screen technology."

NEWS OBSERVER, 28 OCTOBER 2010, http://www.newsobserver.com/2010/10/28/766257/
nc-republican-party-chair-touchscreen.html#ixzz13gTJCCvp

Humboldt County CA, 2008

Serious Error in Diebold Voting Software Caused Lost Ballots in California County, by Kim Zetter

Election officials in a small county in California discovered by chance last week that the tabulation software they used to tally votes in this year's general election dropped 197 paper ballots from the totals at one precinct. The system's audit log also appears to have deleted any sign that the ballots had ever been recorded.

Premier has acknowledged ...its software caused the system to delete votes. The company has apparently known about the problem since 2004 ...

[RoV] Crnich would never have discovered the problem through her standard canvassing procedures ...nor would she have discovered it while conducting a mandatory manual audit that California counties are required to do.

Crnich discovered the missing ballots only because she happened to implement a new and innovative auditing system this year that was spearheaded by members of the public who helped her develop it.

WIRED NEWS, 8 DECEMBER 2008, http://blog.wired.com/27bstroke6/2008/12/unique-election.html

Polk County NC, 2008

Owens victory in Polk is in doubt, by Times-News staff

Ted Owens went to sleep Tuesday night thinking he had earned another term . . . A recount Wednesday showed he may not have. . . .

Computer software initially displayed figures that were different than those shown by the voting machines . . .

The software installed in the stand-alone computer that ballot results are fed into was the problem . . . [Elections Director Dale Edwards] said there was no explanation as to why the computer counted the wrong numbers, and no one is at fault.

BLUERIDGENOW.COM TIMES-NEWS, 6 NOVEMBER 2008, http://www.blueridgenow.com/article/20081106/NEWS/811050255

Santa Clara County, CA, 2008

Few problems reported in area despite record turnout, by Karen de Sá and Lisa Fernandez

Record-high voting in the Bay Area on Tuesday mostly defied predictions of unwieldy waits and overwhelmed polls. But in Santa Clara County, concerns about touch-screen voting machines will likely increase following significant malfunctions.

Fifty-seven of the county's Sequoia Voting Systems machines failed on Election Day, resulting in hourslong delays before replacements arrived.

MERCURY NEWS, 4 NOVEMBER 2008, http:
//www.mercurynews.com/elections/ci_10901166?nclick_check=1

Leon County, FL, 2008

Ballots not being recorded at two Leon County polling places,
by Angeline J. Taylor

Leon County Supervisor of Elections Ion Sancho has reported that
ballots . . . are not being read properly. The problem, he said, rests
with a new machine that has been purchased for polling sites
throughout the state. . . .

"Certain ballots are being rejected across the state," he said. . . . If the
machine reads the ballot card as too long, the . . . machine will simply
not read the card.

TALLAHASSEE DEMOCRAT, 20 OCTOBER 2008, http://www.
tallahassee.com/article/20081020/BREAKINGNEWS/81020024

Palm Beach County, FL, 2008

Florida Primary Recount Surfaces Grave Voting Problems One Month Before Presidential Election, by Kim Zetter

At issue is an August 26 primary election in which officials discovered, during a recount of a close judicial race, that more than 3,400 ballots had mysteriously disappeared after they were initially counted on election day. The recount a week later, minus the missing ballots, flipped the results of the race to a different winner.

. . . officials found an additional 227 ballots that were never counted on election day . . . in boxes in the county's tabulation center.

Palm Beach County was using new optical-scan machines that it recently purchased from Sequoia Voting Systems for $5.5 million.

Palm Beach County, FL, 2008, cont'd

[In a re-scan of ballots the machines had rejected] [o]fficials expected the machines would reject the same ballots again. But that didn't happen. During a first test of 160 ballots, the machines accepted three of them. In a second test of 102 ballots, the machines accepted 13 of them ... When the same ballots were run through the machines again, 90 of the ballots were accepted.

[T]he county then re-scanned two batches of 51 ballots each that had initially been rejected for having no vote cast in the judicial race, but that were found in a manual examination to contain legitimate votes for one candidate or the other. The first batch of 51 ballots were found to have legitimate votes for Abramson. The second batch of 51 ballots were found to have legitimate votes for Wennet.

In the first batch of 51 ballots ... 11 of the ballots that had previously been rejected as undervotes were now accepted ... the remaining 40 ballots were rejected as having no votes. In the second batch of 51 ballots ... the same machine accepted 2 ballots and rejected 49.

Palm Beach County, FL, 2008, cont'd

The same two batches of ballots were then run through the second
...machine. [I]n the first batch ...the machine accepted 41 ...and rejected
10 others. In the second batch ...the machine accepted 49 of the ballots
and rejected 2—the exact opposite of the results from the first machine.

WIRED NEWS, 7 OCTOBER 2008, `http://blog.wired.com/27bstroke6/2008/10/florida-countys.html`

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

see also hearings at
`http://www.octt.dc.gov/services/on_demand_video/`
`channel13/October2008/10_03_08_PUBSVRC_2.asx`

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

Ohio 2004

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

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

[E]arly 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 Elections Code §15360

[T]he 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)

[Officials] 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?

Others

Oregon and New Mexico have audit laws that allow the sample (of races and/or ballots) to be selected before the election.

Rep. Rush Holt has proposed federal legislation that has tiered sampling fractions, depending on the margin—but no requirement for followup if errors are found.

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

What should an election audit law do?

Legislation should enunciate *principles*, not *methods*.

Methods are best left to regulation: Easier to improve, fix, etc.

Mutual distrust among election integrity advocates, elections officials, and legislators is an unfortunate but important consideration.

What's the issue?

- Any way of counting votes makes mistakes.

- If there are enough mistakes that overstated the margin, apparent outcome is wrong.

- If there's an audit trail that shows the right outcome, can correct wrong outcomes—by counting all the paper by hand.

- Check the addition (more generally, the algorithm); check what was added (more generally, the vote data).

- Sum should be perfect (or call the feds!).

- Summands need to be accurate enough to determine correct winner. (If DRE results aren't perfect, call the feds!)

What's the issue?

- Any way of counting votes makes mistakes.

- If there are enough mistakes that overstated the margin, apparent outcome is wrong.

- If there's an audit trail that shows the right outcome, can correct wrong outcomes—by counting all the paper by hand.

- Check the addition (more generally, the algorithm); check what was added (more generally, the vote data).

- Sum should be perfect (or call the feds!).

- Summands need to be accurate enough to determine correct winner. (If DRE results aren't perfect, call the feds!)

What's the issue?

- Any way of counting votes makes mistakes.

- If there are enough mistakes that overstated the margin, apparent outcome is wrong.

- If there's an audit trail that shows the right outcome, can correct wrong outcomes—by counting all the paper by hand.

- Check the addition (more generally, the algorithm); check what was added (more generally, the vote data).

- Sum should be perfect (or call the feds!).

- Summands need to be accurate enough to determine correct winner. (If DRE results aren't perfect, call the feds!)

What's the issue?

- Any way of counting votes makes mistakes.

- If there are enough mistakes that overstated the margin, apparent outcome is wrong.

- If there's an audit trail that shows the right outcome, can correct wrong outcomes—by counting all the paper by hand.

- Check the addition (more generally, the algorithm); check what was added (more generally, the vote data).

- Sum should be perfect (or call the feds!).

- Summands need to be accurate enough to determine correct winner. (If DRE results aren't perfect, call the feds!)

What's the issue?

- Any way of counting votes makes mistakes.

- If there are enough mistakes that overstated the margin, apparent outcome is wrong.

- If there's an audit trail that shows the right outcome, can correct wrong outcomes—by counting all the paper by hand.

- Check the addition (more generally, the algorithm); check what was added (more generally, the vote data).

- Sum should be perfect (or call the feds!).

- Summands need to be accurate enough to determine correct winner. (If DRE results aren't perfect, call the feds!)

What's the issue?

- Any way of counting votes makes mistakes.
- If there are enough mistakes that overstated the margin, apparent outcome is wrong.
- If there's an audit trail that shows the right outcome, can correct wrong outcomes—by counting all the paper by hand.
- Check the addition (more generally, the algorithm); check what was added (more generally, the vote data).
- Sum should be perfect (or call the feds!).
- Summands need to be accurate enough to determine correct winner. (If DRE results aren't perfect, call the feds!)

What's the question?

- Compliance audits vs. materiality audits.

- Detection paradigm: If the outcome is wrong, ensure a big chance of finding at least one error.

- But audits almost invariably find at least one error. What then?

- What do we want audits to accomplish?

- One possibility: correct wrong electoral outcomes.

- Risk-limiting paradigm: If the outcome is wrong, ensure a big chance of correcting it.

What's the question?

- Compliance audits vs. materiality audits.

- Detection paradigm: If the outcome is wrong, ensure a big chance of finding at least one error.

- But audits almost invariably find at least one error. What then?

- What do we want audits to accomplish?

- One possibility: correct wrong electoral outcomes.

- Risk-limiting paradigm: If the outcome is wrong, ensure a big chance of correcting it.

What's the question?

- Compliance audits vs. materiality audits.

- Detection paradigm: If the outcome is wrong, ensure a big chance of finding at least one error.

- But audits almost invariably find at least one error. What then?

- What do we want audits to accomplish?

- One possibility: correct wrong electoral outcomes.

- Risk-limiting paradigm: If the outcome is wrong, ensure a big chance of correcting it.

What's the question?

- Compliance audits vs. materiality audits.
- Detection paradigm: If the outcome is wrong, ensure a big chance of finding at least one error.
- But audits almost invariably find at least one error. What then?
- What do we want audits to accomplish?
- One possibility: correct wrong electoral outcomes.
- Risk-limiting paradigm: If the outcome is wrong, ensure a big chance of correcting it.

What's the question?

- Compliance audits vs. materiality audits.
- Detection paradigm: If the outcome is wrong, ensure a big chance of finding at least one error.
- But audits almost invariably find at least one error. What then?
- What do we want audits to accomplish?
- One possibility: correct wrong electoral outcomes.
- Risk-limiting paradigm: If the outcome is wrong, ensure a big chance of correcting it.

What's the question?

- Compliance audits vs. materiality audits.
- Detection paradigm: If the outcome is wrong, ensure a big chance of finding at least one error.
- But audits almost invariably find at least one error. What then?
- What do we want audits to accomplish?
- One possibility: correct wrong electoral outcomes.
- Risk-limiting paradigm: If the outcome is wrong, ensure a big chance of correcting it.

Risk-limiting audits

- Crucial question: When to *stop* auditing [not how big a sample to start with].

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

- Eventually, either have strong evidence that the outcome is right, or the whole contest has been counted by hand and correct outcome is known.

- Sequential test of the null hypothesis that the outcome is wrong. "Risk" is chance of type I error: concluding a wrong outcome is right. Can control rigorously. No possibility of a type II error.

Risk-limiting audits

- Crucial question: When to *stop* auditing [not how big a sample to start with].

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

- Eventually, either have strong evidence that the outcome is right, or the whole contest has been counted by hand and correct outcome is known.

- Sequential test of the null hypothesis that the outcome is wrong. "Risk" is chance of type I error: concluding a wrong outcome is right. Can control rigorously. No possibility of a type II error.

Risk-limiting audits

- Crucial question: When to *stop* auditing [not how big a sample to start with].

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

- Eventually, either have strong evidence that the outcome is right, or the whole contest has been counted by hand and correct outcome is known.

- Sequential test of the null hypothesis that the outcome is wrong. "Risk" is chance of type I error: concluding a wrong outcome is right. Can control rigorously. No possibility of a type II error.

Risk-limiting audits

- Crucial question: When to *stop* auditing [not how big a sample to start with].

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

- Eventually, either have strong evidence that the outcome is right, or the whole contest has been counted by hand and correct outcome is known.

- Sequential test of the null hypothesis that the outcome is wrong. "Risk" is chance of type I error: concluding a wrong outcome is right. Can control rigorously. No possibility of a type II error.

Role of statistics

Limiting the risk is easy

No statistics needed: just count all the ballots by hand.

Statistics lets you do less counting when the outcome is right, but still ensure a big chance of a full hand count when outcome is wrong.

California AB 2023 (Saldaña, sponsored by SoS Bowen)

Unanimous bipartisan support in both houses.
11 counties committed to pilots; 20 interested.

(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.

http://www.leginfo.ca.gov/pub/09-10/bill/asm/ab_
2001-2050/ab_2023_bill_20100325_amended_asm_v98.html

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, correct outcome: outcome a full manual tally of the audit trail would show (by definition or by law)

True winner: would win according to full hand tally, if there were a full hand tally

more definitions . . .

Risk-limiting audit: audit with guaranteed minimum chance of correcting wrong outcomes (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 wrong, no matter what caused the outcome to be wrong.

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

Simultaneous risk: the maximum chance that the audit won't correct one or more of the apparent outcomes that are incorrect

Crucial ingredients for risk-limiting audits

- 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. (Smaller batches are better.)

- 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.

- Only one current audit law limits risk: AB 2023. [CO HB 09-1335 requires "risk-limiting audits" but doesn't define "risk."]

Crucial ingredients for risk-limiting audits

- 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. (Smaller batches are better.)

- 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.

- Only one current audit law limits risk: AB 2023. [CO HB 09-1335 requires "risk-limiting audits" but doesn't define "risk."]

Crucial ingredients for risk-limiting audits

- 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. (Smaller batches are better.)

- 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.

- Only one current audit law limits risk: AB 2023. [CO HB 09-1335 requires "risk-limiting audits" but doesn't define "risk."]

Crucial ingredients for risk-limiting audits

- 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. (Smaller batches are better.)

- 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.

- Only one current audit law limits risk: AB 2023. [CO HB 09-1335 requires "risk-limiting audits" but doesn't define "risk."]

Crucial ingredients for risk-limiting audits

- 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. (Smaller batches are better.)

- 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.

- Only one current audit law limits risk: AB 2023. [CO HB 09-1335 requires "risk-limiting audits" but doesn't define "risk."]

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? (Define "little" for tractability and power.)

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

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? (Define "little" for tractability and power.)

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

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? (Define "little" for tractability and power.)

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

MACRO

Sufficient condition for all outcomes to be right:

For every (winner, loser) pair, net overstatement of the margin between them is less than 100% of the reported margin between them.

Bound: (sum of max) $\geq$ (max of sum).

Simple sufficient condition.

MACRO

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

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

If any apparent outcome is wrong,

$$
\exists c \in \{1, \ldots, C\} \text{ s.t. } \exists (w \in \mathcal{W}_c, \ \ell \in \mathcal{L}_c) \text{ with } \sum_{p=1}^{N} e_{pw\ell} \geq 1. \tag{1}
$$

Test based on sufficient condition

$$e_p \equiv \max_c \max_{w \in \mathcal{W}_c,\, \ell \in \mathcal{L}_c} e_{pw\ell}.$$

All outcomes must be correct if

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

Maximum across-contest relative overstatement of pairwise margins (MACRO)

Controlling the familywise error rate

C null hypotheses,

the outcome of contest c is incorrect, c = 1, . . . , 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 the overstatement error in each batch

A priori bounds are crucial.

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

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

Hence,

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

$U \equiv \sum_p u_p$, upper bound on total MACRO.

Sampling Designs

- Most jurisdictions that have audits use stratified cluster sampling.

- For most voting technology, limited to some kind of cluster sample (c.f., Yolo, Orange audits).

- Simple, Stratified (by county, voting method, other), PPEB/PPS, NEGEXP, Stratified PPEB?

- Sampling scheme affects choice of test statistic—analytic tractability

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

- More efficient choices possible for PPEB: Kaplan-Markov, Feige?

Sampling Designs

- Most jurisdictions that have audits use stratified cluster sampling.
- For most voting technology, limited to some kind of cluster sample (c.f., Yolo, Orange audits).
- Simple, Stratified (by county, voting method, other), PPEB/PPS, NEGEXP, Stratified PPEB?
- Sampling scheme affects choice of test statistic—analytic tractability
- Weighted max, binning for simple & stratified sampling, NEGEXP, PPEB.
- More efficient choices possible for PPEB: Kaplan-Markov, Feige?

Sampling Designs

- Most jurisdictions that have audits use stratified cluster sampling.

- For most voting technology, limited to some kind of cluster sample (c.f., Yolo, Orange audits).

- Simple, Stratified (by county, voting method, other), PPEB/PPS, NEGEXP, Stratified PPEB?

- Sampling scheme affects choice of test statistic—analytic tractability

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

- More efficient choices possible for PPEB: Kaplan-Markov, Feige?

Sampling Designs

- Most jurisdictions that have audits use stratified cluster sampling.
- For most voting technology, limited to some kind of cluster sample (c.f., Yolo, Orange audits).
- Simple, Stratified (by county, voting method, other), PPEB/PPS, NEGEXP, Stratified PPEB?
- Sampling scheme affects choice of test statistic—analytic tractability
- Weighted max, binning for simple & stratified sampling, NEGEXP, PPEB.
- More efficient choices possible for PPEB: Kaplan-Markov, Feige?

Sampling Designs

- Most jurisdictions that have audits use stratified cluster sampling.
- For most voting technology, limited to some kind of cluster sample (c.f., Yolo, Orange audits).
- Simple, Stratified (by county, voting method, other), PPEB/PPS, NEGEXP, Stratified PPEB?
- Sampling scheme affects choice of test statistic—analytic tractability
- Weighted max, binning for simple & stratified sampling, NEGEXP, PPEB.
- More efficient choices possible for PPEB: Kaplan-Markov, Feige?

Sampling Designs

- Most jurisdictions that have audits use stratified cluster sampling.
- For most voting technology, limited to some kind of cluster sample (c.f., Yolo, Orange audits).
- Simple, Stratified (by county, voting method, other), PPEB/PPS, NEGEXP, Stratified PPEB?
- Sampling scheme affects choice of test statistic—analytic tractability
- Weighted max, binning for simple & stratified sampling, NEGEXP, PPEB.
- More efficient choices possible for PPEB: Kaplan-Markov, Feige?

Taint & PPEB Sampling

taint of batch *p*

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

Independent draws. In each draw,

$$\mathbf{P}\{\text{draw batch } p\} = u_p/U.$$

PPS, used in financial auditing.

Taint of jth draw is T_j. $\{T_j\}$ are iid. $\mathbf{E} T_j = E/U$.

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

Reduces auditing to testing 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 batch if it has not already been audited.

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

3. Compute

$$P_n \equiv \prod_{j=1}^{n} \frac{1 - 1/U}{1 - T_j}. \quad \text{See November 2010 WIRED, p.56} \quad (2)$$

4. If $P_n < \alpha$, report apparent outcomes and stop. If $n = m$, audit remaining batches, report then-known outcomes and stop.. Else, $n \leftarrow n + 1$ and go to 1.

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 batch if it has not already been audited.

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

3. Compute

$$P_n \equiv \prod_{j=1}^{n} \frac{1 - 1/U}{1 - T_j}. \quad \text{See November 2010 WIRED, p.56} \quad (2)$$

4. If $P_n < \alpha$, report apparent outcomes and stop. If $n = m$, audit remaining batches, report then-known outcomes and stop.. Else, $n \leftarrow n + 1$ and go to 1.

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 batch if it has not already been audited.

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

3. Compute

$$P_n \equiv \prod_{j=1}^{n} \frac{1 - 1/U}{1 - T_j}. \quad \text{See November 2010 WIRED, p.56} \quad (2)$$

4. If $P_n < \alpha$, report apparent outcomes and stop. If $n = m$, audit remaining batches, report then-known outcomes and stop.. Else, $n \leftarrow n + 1$ and go to 1.

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 batch if it has not already been audited.

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

3. Compute

$$P_n \equiv \prod_{j=1}^{n} \frac{1 - 1/U}{1 - T_j}. \quad \text{See November 2010 WIRED, p.56} \quad (2)$$

4. If $P_n < \alpha$, report apparent outcomes and stop. If $n = m$, audit remaining batches, report then-known outcomes and stop.. Else, $n \leftarrow n + 1$ and go to 1.

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 batch if it has not already been audited.

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

3. Compute

$$P_n \equiv \prod_{j=1}^{n} \frac{1 - 1/U}{1 - T_j}. \text{ See November 2010 WIRED, p.56} \quad (2)$$

4. If $P_n < \alpha$, report apparent outcomes and stop. If $n = m$, audit remaining batches, report then-known outcomes and stop.. Else, $n \leftarrow n + 1$ and go to 1.

This sequential procedure is risk-limiting

Chance $\geq 1 - \alpha$ of correcting wrong outcomes by full hand count

If any outcome is wrong,

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

Remarkably efficient if batches are not too big.

Super-simple simultaneous audits

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 *ballot-level auditing*, which compares CVR with human interpretation of individual ballots.

Spend some efficiency to buy logistic and computational simplicity.

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 of super-simple method

- Audit entire collection of contests with one simple random sample of ballots.

- Super simple: initial sample size is a constant—the sample size multiplier ρ—divided by the "diluted margin." ρ 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.

Advantages of super-simple method

- Audit entire collection of contests with one simple random sample of ballots.

- Super simple: initial sample size is a constant—the sample size multiplier ρ—divided by the "diluted margin." ρ 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.

Advantages of super-simple method

- Audit entire collection of contests with one simple random sample of ballots.

- Super simple: initial sample size is a constant—the sample size multiplier ρ—divided by the "diluted margin." ρ 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.

Advantages of super-simple method

- Audit entire collection of contests with one simple random sample of ballots.

- Super simple: initial sample size is a constant—the sample size multiplier ρ—divided by the "diluted margin." ρ 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.

Requires picking risk and 2 parameters:

- simultaneous risk limit α. Might be set by legislation.

- error inflation factor $\gamma \geq 100\%$. Controls tradeoff between initial sample size and additional counting when the sample finds many overstatements. γ affects operating characteristics but not risk.

- error tolerance $\lambda < 100\%$. If rate of ballots in the sample with 1-vote maximum overstatements is no more than $\lambda\mu$ and there are no 2-vote overstatement, audit stops. λ affects operating characteristics but not risk.

Requires picking risk and 2 parameters:

- simultaneous risk limit α. Might be set by legislation.

- error inflation factor $\gamma \geq 100\%$. Controls tradeoff between initial sample size and additional counting when the sample finds many overstatements. γ affects operating characteristics but not risk.

- error tolerance $\lambda < 100\%$. If rate of ballots in the sample with 1-vote maximum overstatements is no more than $\lambda\mu$ and there are no 2-vote overstatement, audit stops. λ affects operating characteristics but not risk.

Requires picking risk and 2 parameters:

- simultaneous risk limit α. Might be set by legislation.

- error inflation factor $\gamma \geq 100\%$. Controls tradeoff between initial sample size and additional counting when the sample finds many overstatements. γ affects operating characteristics but not risk.

- error tolerance $\lambda < 100\%$. If rate of ballots in the sample with 1-vote maximum overstatements is no more than $\lambda\mu$ and there are no 2-vote overstatement, audit stops. λ affects operating characteristics but not risk.

Super-simple simultaneous single-ballot procedure

1. Pick risk limit $\alpha \in (0, 1)$, $\gamma \geq 100\%$, $\lambda < 100\%$

2. Calculate the sample-size multiplier ρ

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

   ρ doesn't depend on the audit data or particulars of the contests.
   $\rho = 7$ gives 10% risk limit with $\gamma = 110\%$, $\lambda = 20\%$

3. Calculate the diluted margin μ.

4. Audit simple 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.

Super-simple simultaneous single-ballot procedure

1. Pick risk limit $\alpha \in (0, 1)$, $\gamma \geq 100\%$, $\lambda < 100\%$

2. Calculate the sample-size multiplier ρ

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

   ρ doesn't depend on the audit data or particulars of the contests.
   $\rho = 7$ gives 10% risk limit with $\gamma = 110\%$, $\lambda = 20\%$

3. Calculate the diluted margin μ.

4. Audit simple 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.

Super-simple simultaneous single-ballot procedure

1. Pick risk limit $\alpha \in (0, 1)$, $\gamma \geq 100\%$, $\lambda < 100\%$

2. Calculate the sample-size multiplier ρ

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

   ρ doesn't depend on the audit data or particulars of the contests.
   $\rho = 7$ gives 10% risk limit with $\gamma = 110\%$, $\lambda = 20\%$

3. Calculate the diluted margin μ.

4. Audit simple 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.

Super-simple simultaneous single-ballot procedure

1. Pick risk limit $\alpha \in (0, 1)$, $\gamma \geq 100\%$, $\lambda < 100\%$

2. Calculate the sample-size multiplier ρ

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

   ρ doesn't depend on the audit data or particulars of the contests.
   $\rho = 7$ gives 10% risk limit with $\gamma = 110\%$, $\lambda = 20\%$

3. Calculate the diluted margin μ.

4. Audit simple 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.

Secret sauce

To implement ballot-level 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)

Will try "transitive audit" in Monterey County, 9 May 2011, using Humboldt Transparency Project software on Sequoia ballots, office scanner.

Super-simple for arbitrary batches instead of individual ballots

No escalation if maximum observed taint is less than λ/U, $\lambda \in (0, 1)$:

$$n \geq \frac{-2\ln(\alpha)/(1 - \lambda)}{m}.$$

For $\lambda = 0.2$ and $\alpha = 0.091$, numerator $= 6$.

Simple rule:

1. Take the initial sample size $\geq 6/m$, sample with PPS
2. If no taint in sample is larger than $0.2/U = 0.1m$, can stop w/ risk $\leq 9.1\%$
3. If some observed taint is larger, use KM formula for P-value

Pilot Audits in California

Marin County (February 2008; November 2008, 2009)

Orange County (March 2011)

Yolo County (November 2008, 2009)

Santa Cruz County (November 2008)

Monterey County (to take place May 2011)

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.

2008 Yolo County, CA Measure W Audit

Precinct *100063*

Da   Joint Unified School District
M sure W
S   he Davis Joint Unified School District
p ve existing classroom programs
in   ng math and science, English, music,
pr   al education, librarians, secondary class
size duction, athletics and co-curricular
p grams including drama, debate, and
jou lism by being authorized to levy a
s   tax for a period of three years not to
e   the annual rate of $50.00 per dwelling
u   multi-dwelling parcels and $120.00 per
pa   'or all other parcels?

☐ Yes

☐ No

technologies; nursing and
programs; architecture, engineering and
construction management; computer sci
early childhood development; and fire ar
police public safety programs at the Ame
River, Cosumnes River, El Dorado, Fols
and Sacramento City College campuses

☐ Bonds Yes

☒ Bonds No

City of Davis
Measure N
Shall the Proposed Charter of the City of
Davis Be Adopted?

☐ Yes

☒ No

2009 Yolo County, CA Measure P Audit

Special Election November 2009
City of Davis
November 03, 2009   Precinct

Instruction Text:
Please use a black or blue ink pen to mark your choices on the ballot.
To vote for your choice in each contest, completely fill in the box
provided to the left of your choice.

MEASURE P
Shall Resolution No. 09-132, amending the Davis General Plan to
change the land use designations for the Wildhorse Ranch property from
agriculture to residential uses, as set forth in the Resolution and
establishing the Base Line Project Features for development of the
Wildhorse Ranch Project be approved?

■ Yes
☐ No

Neatness counts

10000170100059

Special Election November 2009
City of Davis
November 03, 2009

Instruction Text:
Please use a black or blue ink pen to mark your c
To vote for your choice in each contest, complete
provided to the left of your choice.

MEASURE P
Shall Resolution No. 09-132, amending the Davis
change the land use designations for the Wildhors
agriculture to residential uses, as set forth in the F
establishing the Base Line Project Features for de
Wildhorse Ranch Project be approved?

☐ Yes
■ No

10000170100059

2011 Orange County, first audit under AB 2023

Project

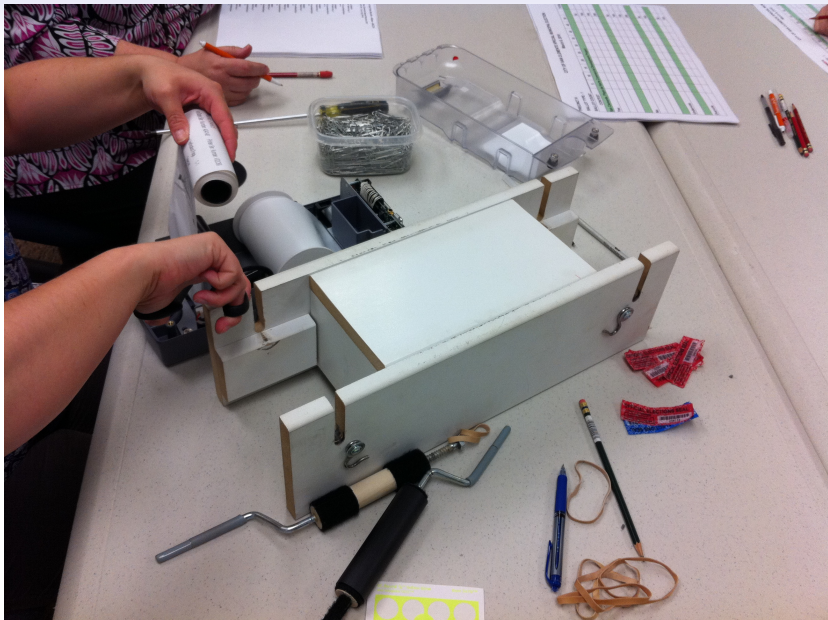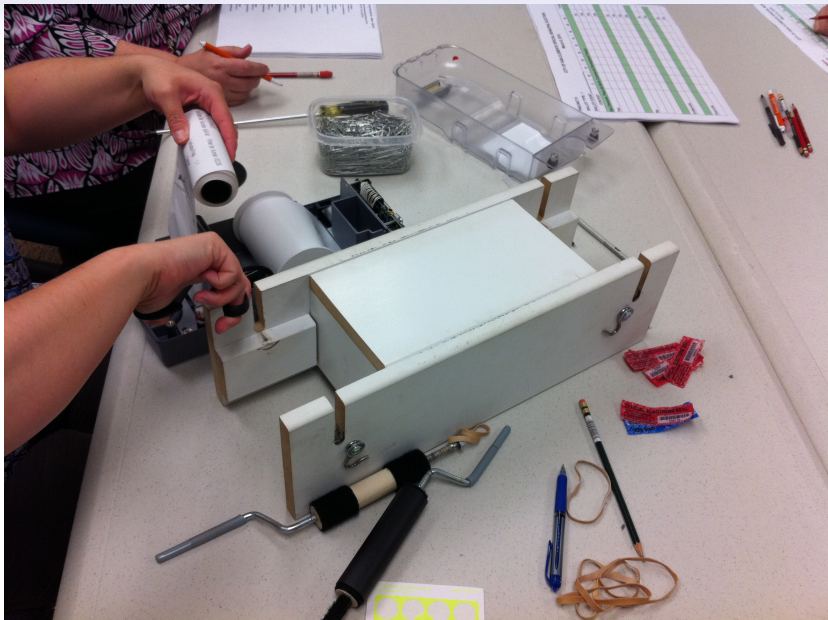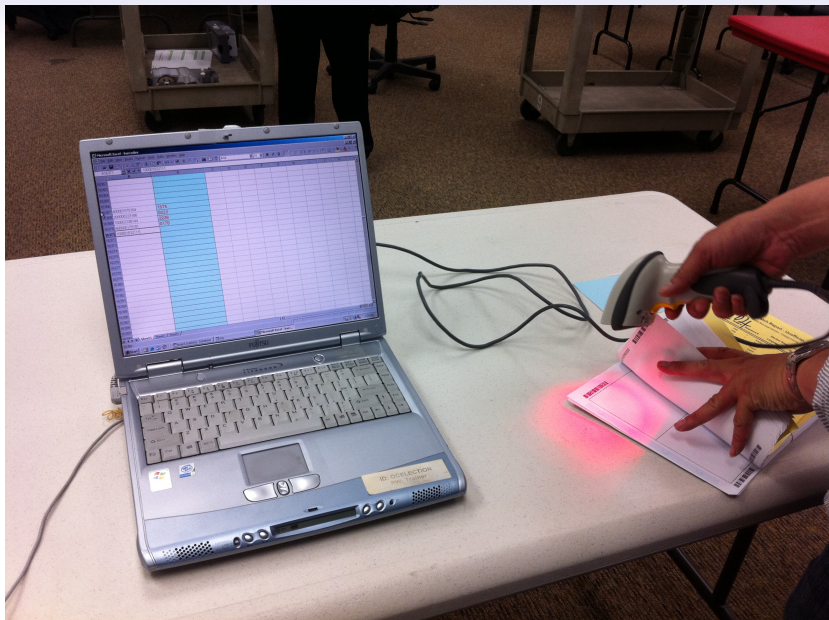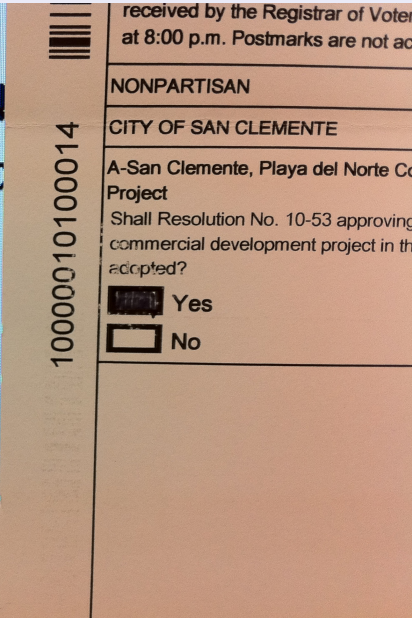Shall Resolution N

commercial develo

adopted?

◼ Yes

☐ No
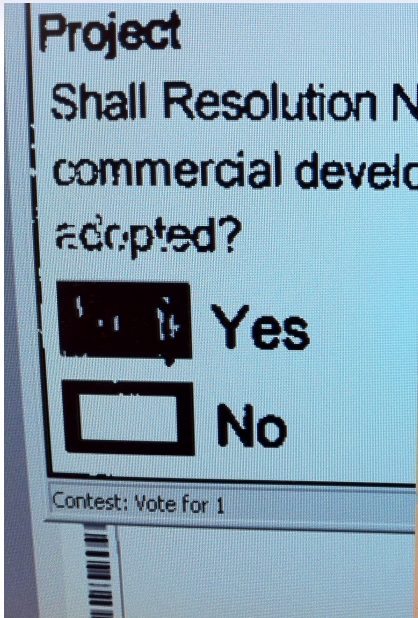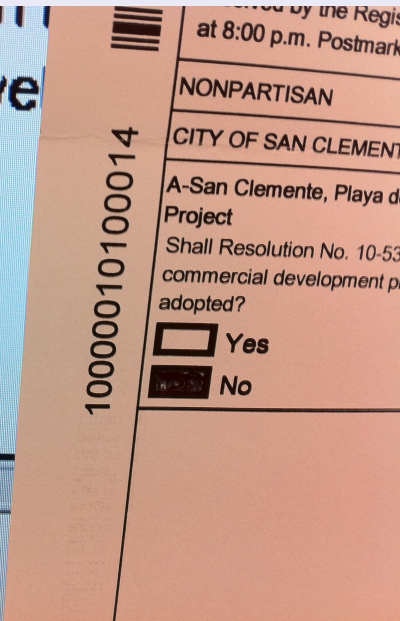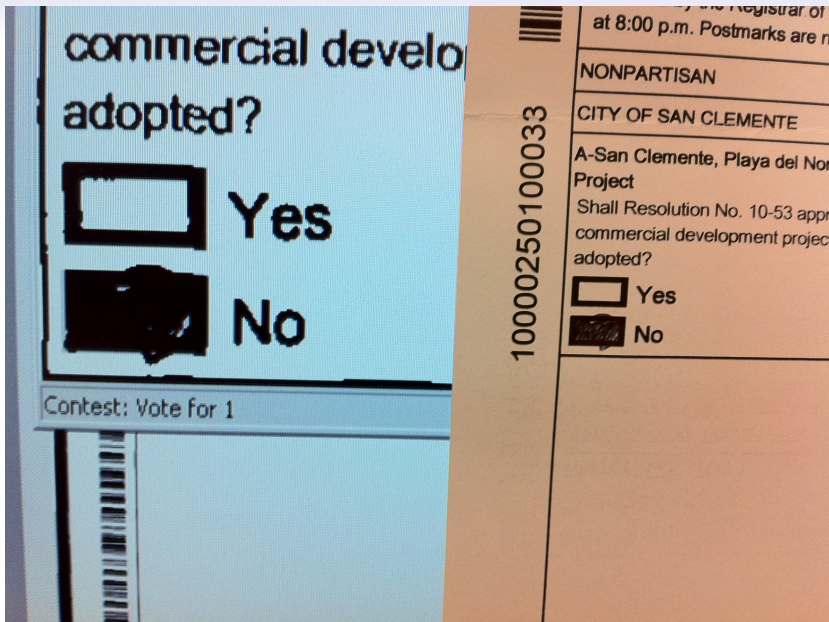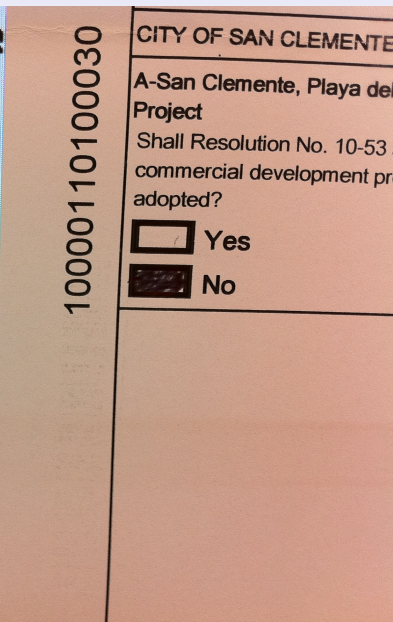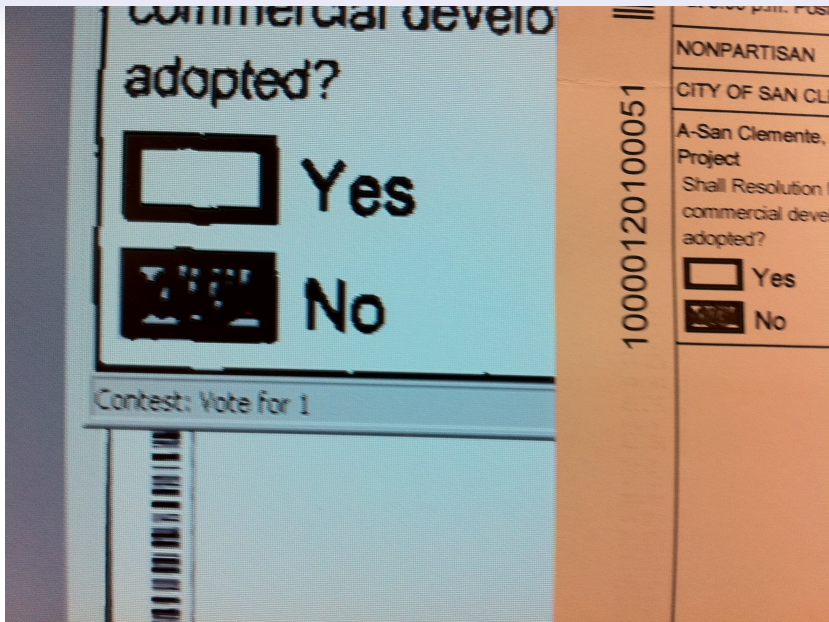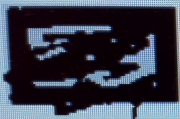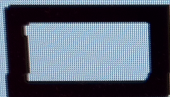
Contest: Vote for 1

1000001010100014

received by the Registrar of Voter
at 8:00 p.m. Postmarks are not ac

NONPARTISAN

CITY OF SAN CLEMENTE

A-San Clemente, Playa del Norte Co
Project
Shall Resolution No. 10-53 approving
commercial development project in th
adopted?

☐ Yes

☐ No

commercial develo

adopted?

☐ Yes

■ No

Contest: Vote for 1

at 8:00 p.m. Postmarks are n

10000250100033

NONPARTISAN

CITY OF SAN CLEMENTE

A-San Clemente, Playa del Nor
Project
Shall Resolution No. 10-53 appr
commercial development projec
adopted?

☐ Yes
☒ No
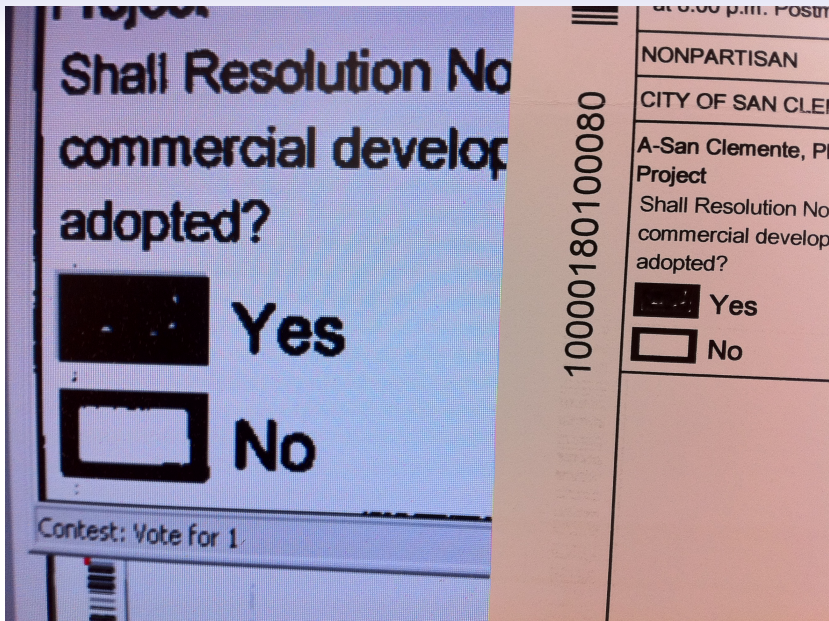
Yolo County Measure P, November 2009
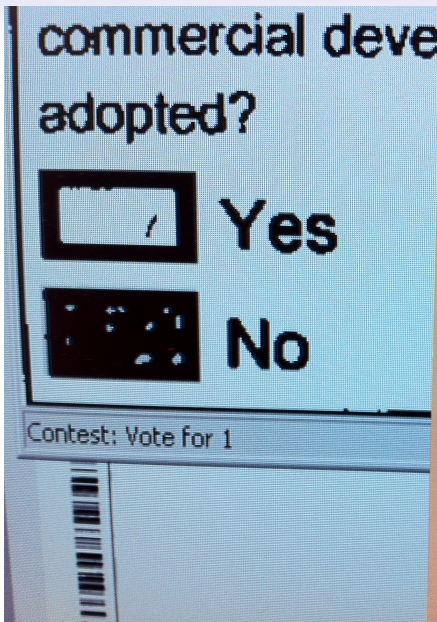
Reg. voters	ballots	precincts	batches	yes	no
38,247	12,675	31	62	3,201	9,465

(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.

Ballot-level auditing would save *lots* of work

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

Ballot-level auditing would save *lots* of work

Can determine the initial sample size for a Kaplan-Markov ballot-level 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, 2009

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%.

Orange County 2011 Audit design and sample

Left provisionals in machine ballot counts for error bounds. 5523 total.

One VBM-only precinct with 119 ballots. 158 election-day paper ballots. 38 rejected provisional ballots

Used a deck of cards to pick a 9-digit seed: shuffled cards well, counted Ace as 1, etc., 10 as 0, and ignored face cards, dealt until we had 9 digits. Used R implementation of Mersenne Twister.

Sample gave 12 eSlate machines with a total of 446 ballots, and 21 individual ballots. Total sample size 467 ballots (expected size was 384.8 ballots). One of the eSlates had already been audited as part of the statutory 1% audit.

Ordered ballots canonically: scanner A, B, C. which scanner, which batch, which ballot in the batch. from that, could look up a serial number for the ballot image use barcode scanner to verify that we had the right ballot then compare the ballot image (with that serial number) with the physical ballot to verify identity of ballot then confirm that the CVR matched our interpretation

1% Statutory Audit

Votes in one precinct counted by hand. No errors found.
Chance the 1% audit would find no errors even if the outcome is
wrong could be over 88%.

Statutory audit does little to limit risk, even if it required a full hand
count if errors were found.

Special steps

Pollworkers instructed to spread voters across machines (roughly 10 per precinct) so that machine batch sizes would be comparable and small.

Unable to export of subtotals by machine from the vote tabulation system. Downloaded counts of voters from each of 200 eSlates to determine sampling weights; about 2 hours work.

Initial sample sizes for various batching rules

San Clemente Measure A, 3/8/2011

batching rule	draws	expected batches	expected ballots
VBM by precinct IP by precinct	18	14.7	6370.2
VBM by ballot IP by precinct	28	27.4	1192.9
VBM by ballot IP by machine	32	31.7	376.6
SS: VBM by ballot IP by ballot	47	46.9	46.9
KM: VBM by ballot IP by ballot	33	33.0	33.0

Expected counting burden, 10% risk limit, no overstatement errors. All based on PPEB sampling using KM inequality. "By ballot" includes error bound "headroom" of 5% (2.1 vote maximum error per ballot). "By machine" error bound is twice the number of ballots. SS: "super-simple" method. Sample size 6.638/margin. KM: Kaplan-Markov using error bound of 2.1 votes per ballot.

Risk-limiting Audits: Costs

San Clemente Measure A, 3/8/2011

1% Statutory Audit: $257.68
Scales as the size of the contest: a contest twice as large would cost
about twice as much to audit.

Risk-limiting: $483.79 (does not include my time or airfare)
Would have cost essentially the same for any contest with the same
percentage margin, no matter how large the contest.

Research directions

- IRV/RCV (Shen, Rivest, Lindeman, Lundell), NPV

- "False winner rate" (Benjamini)

- Optimal stratified sampling rates given the test statistic and strata (Rivest, Higgins)

- Extending KM to stratified samples (Higgins)

- Sharper test given sampling design (Shacham et al. use KL distance for ballot-level)

- Optimal tests if sampling design is up for grabs. Concentration inequalities? Feige?

- Transparent ballot-level audits that maintain voter privacy (Lazarus, Lindeman, Benaloh)

- Auditing E2E encrypted systems (Wallach)

- Simpler, simpler, simpler

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.

New York's Recent Reforms

Moved to precinct-count optically scanned paper ballots.

Introduced audit laws, starting with 3% of machines (scanners).

Irreconcilable differences between hand count and machine count can lead to counting more ballots by hand: 5%, 12%, or all.

NY SD 7

- Balance of power in NY Senate: Either 31 seats for each party, or 32 for Republicans.

- Reported margin of 451 votes (0.5%) for Republican candidate Martins.

- Disagreement about purpose and requirements of auditing.

NY SD 7

- Balance of power in NY Senate: Either 31 seats for each party, or 32 for Republicans.

- Reported margin of 451 votes (0.5%) for Republican candidate Martins.

- Disagreement about purpose and requirements of auditing.

NY SD 7

- Balance of power in NY Senate: Either 31 seats for each party, or 32 for Republicans.
- Reported margin of 451 votes (0.5%) for Republican candidate Martins.
- Disagreement about purpose and requirements of auditing.

NY SD 7: Audit

- 7 of 249 of machines audited at random (3%).

- 3 of 7 (i.e., 43%) showed errors. Net error favored the apparent winner.

- Republicans: the errors were "reconciled": Machines are fine. Democrats disagree.

- Judge Warshawsky: "In my opinion, reconcilable would be 'Is there a clear reason why the deficiency occurred?' "

- Is "the machine was mis-programmed" a clear reason?

- In my opinion, reconcilable would be "We counted again by hand and found that the error was in the hand count, not the machine count."

- "Clear reason" is irrelevant for whether the apparent outcome of the contest is correct. Size of the difference matters.

NY SD 7: Audit

- 7 of 249 of machines audited at random (3%).

- 3 of 7 (i.e., 43%) showed errors. Net error favored the apparent winner.

- Republicans: the errors were "reconciled": Machines are fine. Democrats disagree.

- Judge Warshawsky: "In my opinion, reconcilable would be 'Is there a clear reason why the deficiency occurred?' "

- Is "the machine was mis-programmed" a clear reason?

- In my opinion, reconcilable would be "We counted again by hand and found that the error was in the hand count, not the machine count."

- "Clear reason" is irrelevant for whether the apparent outcome of the contest is correct. Size of the difference matters.

NY SD 7: Audit

- 7 of 249 of machines audited at random (3%).
- 3 of 7 (i.e., 43%) showed errors. Net error favored the apparent winner.
- Republicans: the errors were "reconciled": Machines are fine. Democrats disagree.
- Judge Warshawsky: "In my opinion, reconcilable would be 'Is there a clear reason why the deficiency occurred?' "
- Is "the machine was mis-programmed" a clear reason?
- In my opinion, reconcilable would be "We counted again by hand and found that the error was in the hand count, not the machine count."
- "Clear reason" is irrelevant for whether the apparent outcome of the contest is correct. Size of the difference matters.

NY SD 7: Audit

- 7 of 249 of machines audited at random (3%).

- 3 of 7 (i.e., 43%) showed errors. Net error favored the apparent winner.

- Republicans: the errors were "reconciled": Machines are fine. Democrats disagree.

- Judge Warshawsky: "In my opinion, reconcilable would be 'Is there a clear reason why the deficiency occurred?' "

- Is "the machine was mis-programmed" a clear reason?

- In my opinion, reconcilable would be "We counted again by hand and found that the error was in the hand count, not the machine count."

- "Clear reason" is irrelevant for whether the apparent outcome of the contest is correct. Size of the difference matters.

NY SD 7: Audit

- 7 of 249 of machines audited at random (3%).

- 3 of 7 (i.e., 43%) showed errors. Net error favored the apparent winner.

- Republicans: the errors were "reconciled": Machines are fine. Democrats disagree.

- Judge Warshawsky: "In my opinion, reconcilable would be 'Is there a clear reason why the deficiency occurred?' "

- Is "the machine was mis-programmed" a clear reason?

- In my opinion, reconcilable would be "We counted again by hand and found that the error was in the hand count, not the machine count."

- "Clear reason" is irrelevant for whether the apparent outcome of the contest is correct. Size of the difference matters.

NY SD 7: Audit

- 7 of 249 of machines audited at random (3%).

- 3 of 7 (i.e., 43%) showed errors. Net error favored the apparent winner.

- Republicans: the errors were "reconciled": Machines are fine. Democrats disagree.

- Judge Warshawsky: "In my opinion, reconcilable would be 'Is there a clear reason why the deficiency occurred?' "

- Is "the machine was mis-programmed" a clear reason?

- In my opinion, reconcilable would be "We counted again by hand and found that the error was in the hand count, not the machine count."

- "Clear reason" is irrelevant for whether the apparent outcome of the contest is correct. Size of the difference matters.

NY SD 7: Audit

- 7 of 249 of machines audited at random (3%).
- 3 of 7 (i.e., 43%) showed errors. Net error favored the apparent winner.
- Republicans: the errors were "reconciled": Machines are fine. Democrats disagree.
- Judge Warshawsky: "In my opinion, reconcilable would be 'Is there a clear reason why the deficiency occurred?' "
- Is "the machine was mis-programmed" a clear reason?
- In my opinion, reconcilable would be "We counted again by hand and found that the error was in the hand count, not the machine count."
- "Clear reason" is irrelevant for whether the apparent outcome of the contest is correct. Size of the difference matters.

Example NY SD7 Audit report

HO 13-24 machine 952:

It was evident that the discrepancy between the audit (manual count) of the votes reflected on the ballots in the ballot box, and the votes reflected on the scanner result tape, are attributable to the two additional ballots found in the ballot box. The disparity between the manual count, and the scanner result tape, are precisely equal to the votes reflected when counting all off [sic] the ballots in the ballot box – including the two additional ballots. This is not "scanner error," but is instead attributable to any one of a number of alternative possibilities . . . The presence of more ballots in the ballot box does not demonstrate that the scanner has "failed," merely that the machine operated as it was designed to do – but with the result that some number (in this case, two) ballots were not scanned. . . . PASS

More

2-4059 Machine 104

1 additional ballot found and explains all discrepancies.
. . . PASS

H18-12 Machine 259

One additional ballot was counted by the scanner than was
found in the bin . . . PASS

Proffered testimony

The audit results not surprising even if a full hand count would show Mr. Johnson to be the winner.

Substantial possibility that the machine with the largest error was not one of the machines that was audited. 97% chance that auditing 7 of 249 machines won't check the machine with the largest error.

Average of less than two errors per machine could account for the apparent margin of about 450 votes.

Average of one error per 200 ballots could account for the apparent margin.

Proffer, contd.

Not a surprising level of error in precinct-count optically scanned ballots. Consistent with the errors the audit did find, within the statistical variability expected from "the luck of the draw."

Large potential for error: the 242 unaudited machines could hold enough error to account for the apparent margin 186 times over. Sixty-six of the 242 unaudited machines could *individually* hold enough error to account for the apparent margin.

Proffer, contd.

Substantial chance that a 3% or 8% audit would find little or no error even if Sen. Johnson is the true winner.

If 30 of the 249 machines have errors of 15 votes or more—enough to account for the apparent margin—chance the 3% audit would have found any of those machines is under 60%.

If 20 of the 249 machines have errors of 23 votes or more–enough to account for the apparent margin–chance the 3% audit would have found any of those machines is under 45%.

If 20 of the 242 unaudited machines have errors of 23 votes or more (enough to account for the apparent margin) and an additional 5% of the machines are audited, chance the additional audit would find any of those 20 is under 69%.

Proffered testimony, contd.

Margin is so small compared to the possible errors that very large percentage of machines must be audited to give strong evidence that Mr. Martins is indeed the winner.

3% is not sufficient.

8% is not sufficient.

To have 90% statistical confidence that Mr. Martins won requires auditing a minimum of 90% of the machines selected randomly: an additional 218 machines.

This is true if the audit finds that those 218 machines have counted perfectly. If the audit of those 218 machines found many errors, still more machines would have to be audited.

Feige's Inequality—sharper than Kaplan-Markov?

He, Zhang, and Zhang, 2010 (Theorem 3.2)

$\{X_j\}_{j=1}^n$ independent; $\mathbf{E}X_j = 0$, $\forall j$. Fix $\Delta > 0$. Suppose $\exists c > 0$ s.t. $X_j \geq -c\Delta$ $\forall j$. Let $S \equiv \sum_{j=1}^n X_j$. Then for any $\tau > 0$,

$$\mathbf{P}\{S < \Delta\} \geq e^{-1/\tau} F(c, \tau \max(1, c)), \tag{3}$$

where

$$F(c_1, c_2) \equiv (2\sqrt{3} - 3)\frac{4(s(c_1, c_2) + 2)}{s^2(c_1, c_2) + 12s(c_1, c_2) + 24} \tag{4}$$

and

$$s(c_1, c_2) \equiv \max\{c_1^2 + 4c_1, c_2^2 - 4c_2, c_1^2 + c_2^2 - 4c_1 c_2 - 4(c_2 - c_1)\} \tag{5}$$