Risk-Limiting Audits for Party-List Elections

Philip B. Stark

Department of Statistics
University of California, Berkeley

IT University of Denmark
Copenhagen, Denmark
21 November 2013

Joint with:
Mark Lindeman, Carsten Schürmann, Vanessa Teague, Vince Yates

All vote counting methods can make mistakes

- Most concerns are with electronic vote tabulation, but hand counting errs, too.
- Denmark counts votes by hand, thrice (or more).
- Can we save effort by auditing?
- What roles could audits play in Danish elections?

What do we want an audit to do?

Quality control in general.

Ensure that the electoral outcome is correct;
If outcome is wrong, correct it before it's official.

Outcome means the set of winners, not exact counts.

How can an audit correct a wrong outcome?

- If there's an adequately accurate audit trail, the audit could count all the votes by hand.
- The goal is to correct the outcome if it is wrong, but to do as little counting as possible when the outcome is right.
- Use statistical techniques to decide whether you have checked enough.

Why not just count all votes by hand (repeatedly)?

- Unnecessarily expensive and slow; accuracy decreases with fatigue.
- Instead, make a first count, then check a random sample.
- Keep checking until there's convincing evidence that the outcome is right—or until all ballots have been hand counted.
- Fatigue, staff quality, etc., may make a full hand count less accurate than a focused audit of a small random sample.
- An audit of hundreds or thousands of ballots can be more transparent than a full count: Public could actually observe the whole process.

Controlling the chance of error

- Since the sample is random, there's a chance a wrong outcome will escape correction—but we can make that chance as small as we want. Statistics says how.
- *Risk* is the largest possible chance that the audit does not correct the outcome, if the outcome is wrong.
- *Risk-limiting audit* ensures that the largest possible chance is still a small chance, like 10%, 5%, 1%.
- Generally, have to check more to make chance smaller.

"Stirring" is key to reducing work

- Don't have to climb into the bathtub to tell if it's hot: can just stick your toe in—if the water is stirred well.
- Don't have to drink a whole pot of soup to tell if it's too salty: a teaspoon is enough—if the pot has been stirred. (Doesn't matter whether the pot holds 0.5 l or 100 l.)

How do you stir ballots?

Random sampling is stirring

- Imagine numbering the ballots.
- Write the numbers on ping-pong balls; put in a lotto machine.
- Lotto machine stirs the balls and spits some out.
- The ballots with the numbers on the selected balls are a random sample of ballots.
- Easier to stir balls than ballots. Even easier to generate random numbers.
- Still amounts to putting ballots into a huge mixer to stir them, then taking a "teaspoon" of ballots.

Risk is *not*

- The chance that the certified outcome is wrong.
- The fraction of certified outcomes that are wrong.

Risk limit

- *Assumes* outcome is wrong in the hardest-to-find way.
- Biggest chance a wrong outcome won't be corrected.

Paper rules—if it is right

- Correct wrong outcomes by counting the whole audit trail.
- Counting the whole audit trail won't give right answer unless it's adequately accurate and intact.
- Requires sound procedures for protecting, tracking, and accounting for ballots.
- Denmark is far better than the USA in ballot accounting.
- Does Denmark produce *ballot manifests*?

Ballot-polling Audits and Comparison Audits

- Ballot polling audit: sample ballots until there is strong evidence that looking at all of them would show the same election outcome.
  Like an exit poll—but of ballots, not voters.
- Comparison audit:
  1. Commit to vote subtotals, ideally, individual ballot interpretations
     (equivalent: commit to manifest of sorted, counted bundles)
  2. Check that the subtotals add up exactly to contest results
  3. Check subtotals by hand until there is strong evidence the outcome is right

Tradeoffs

- Ballot polling audit
  - Virtually no set-up costs
  - Requires nothing of voting system
  - Requires more counting than ballot-level comparison audit
  - Does not check tabulation: outcome could be right because errors cancel
- Comparison audit
  - Heavy demands for reporting and data export
  - Requires commitment to subtotals
  - Requires retrieving ballots that correspond to subtotals
  - Ballot-level not possible w/ current electronic systems (but might be for DK)
  - Checks tabulation
  - Ballot-level comparison audits require least hand counting

Both need *ballot manifest*.

Statistical formulation of RLAs

Hypothesis Test

Null: outcome is wrong (one or more apparent winners really lost)
Alternative: outcome is right

Reject null $\rightarrow$ conclude outcome is right.
Maximum significance level is the *risk*.
Maximum is over all ways the outcome could be wrong.

Sequential Testing

- Collect data until there's strong evidence that the outcome is right (or until there's a full hand count).
- Need to account for sequential data collection
- Strategy: express sufficient condition in terms of scalar properties of population of cast ballots

Parameters and Statistics

- Ballot polling: for each pair, difference in weighted tallies.
- Comparison: maximum relative overstatement of pairwise margins.
- Both reduce to nonparametric hypothesis that the mean of a finite, bounded, nonnegative population is ≥ 1.
- Surprisingly little work on "simple" problem.
- "Best" test so far is based on Wald's (1945) sequential probability ratio test

Divisors for common "highest averages" methods

Name	used in	d(1)	d(2)	d(3)	d(4)
D'Hondt	Belgium Denmark Luxembourg	1	2	3	4
Modified D'Hondt	Estonia	$1^{0.9}$ 1	$2^{0.9}$ 1.866	$3^{0.9}$ 2.688	$4^{0.9}$ 3.482
Sainte-Laguë	Germany	1	3	5	7
Modified Sainte-Laguë	Norway	1.4	3	5	7

party p	$t(p)/d(1)$	$t(p)/d(2)$	$t(p)/d(3)$	$t(p)/d(4)$
1	100,000	50,000	33,333	25,000
2	60,000	30,000	20,000	15,000
3	40,000	20,000	13,333	10,000
4	30,000	15,000	10,000	7,500
5	25,000	12,500	8,333	6,250

Hypothetical results for contest with $S = 4$ seats, $P = 5$ parties.

$t(p)$ is reported count for party p.
$d(s)$ is the divisor for column s; here $d(s) = s$ (D'Hondt).
$a(p)$ is actual (i.e., perfect) count for party p.

party p	$t(p)/1$	$t(p)/2$	$t(p)/3$	$t(p)/4$
1	100,000	50,000	33,333	25,000
2	60,000	30,000	20,000	15,000
3	40,000	20,000	13,333	10,000
4	30,000	15,000	10,000	7,500
5	25,000	12,500	8,333	6,250

Apparent winning "pseudo candidates," $S = 4$ seats, $P = 5$ parties

party p	$t(p)/1$	$t(p)/2$	$t(p)/3$	$t(p)/4$
1	100,000	50,000	33,333	25,000
2	60,000	30,000	20,000	15,000
3	40,000	20,000	13,333	10,000
4	30,000	15,000	10,000	7,500
5	25,000	12,500	8,333	6,250

Seat allocation is correct if, for the true tallies $a(p)$ (not just reported tallies $t(p)$) every blue cell is greater than every red cell

party p	$a(p)/1$	$a(p)/2$	$a(p)/3$	$a(p)/4$
1		$a(1)/2$	$a(1)/3$	
2	$a(2)$	$a(2)/2$		
3	$a(3)$	$a(3)/2$		
4	$a(4)$			
5	$a(5)$			

Inequalities to be checked by audit:
each blue cell $>$ all red cells in other rows.

$a(1)/2 > a(2)/2$; $a(1)/2 > a(3)/2$; $a(1)/2 > a(4)$; $a(1)/2 > a(5)$.

$a(2) > a(1)/3$; $a(2) > a(3)/2$; $a(2) > a(4)$; $a(2) > a(5)$;

$a(3) > a(1)/3$; $a(3) > a(2)/2$; $a(3) > a(4)$; $a(3) > a(5)$.

Remaining inequalities guaranteed arithmetically.

$$
\begin{aligned}
B &: \#\text{ ballots cast in the contest} \\
V &: \#\text{ votes per ballot each voter is allowed to cast} \\
P &: \#\text{ parties} \\
S &: \#\text{ seats to be assigned} \\
C_p &: \#\text{ candidates in party } p \\
t(p) &: \text{reported total votes for party } p \\
a(p) &: \text{actual total votes for party } p \\
e(p) &\equiv t(p) - a(p), \text{ error reported vote for party } p \\
t(p,c) &: \text{reported total votes for candidate } c \text{ in party } p \\
a(p,c) &: \text{actual total votes for candidate } c \text{ in party } p \\
e(p,c) &\equiv t(p,c) - a(p,c), \text{ error in reported vote for candidate } c \text{ in party } p \\
d(s) &: \text{divisor for column } s \\
p_{ps} &\equiv t(p)/d(s) \\
\pi_{ps} &\equiv a(p)/d(s) \\
\mathcal{W} &: \text{pairs } (p,s) \text{ with the } S \text{ largest values of } p_{ps} \\
\mathcal{L} &: \text{pairs } (p,s),\ p = 1, \ldots, P,\ s = 1, \ldots, S \text{ not in } \mathcal{W} \\
\mathcal{W}^P &: \text{parties } p \text{ that (apparently) won at least one seat} \\
\mathcal{L}^P &: \text{parties } p \text{ that (apparently) lost at least one seat} \\
\mathcal{W}_p &: \text{candidates } c \text{ in party } p \text{ who were seated} \\
\mathcal{L}_p &: \text{candidates } c \text{ in party } p \text{ who were not seated}
\end{aligned}
$$

Pseudo-candidates

- $P \times S$ pairs (p, s) of *pseudo-candidates*.
- Candidate (p, s) reported to have received $p_{ps} = t(p)/d(s)$ votes.
- Candidate (p, s) actually received $\pi_{ps} = a(p)/d(s)$ votes.
- $\mathcal{W}$ are "apparent winners" according to reported tally.
- *apparent outcome*: $\#$ seats each party gets according to reported totals $t(p)$, $p = 1, \ldots, P$.
- *true outcome*: $\#$ seats each party would get according to true totals $a(p)$, $p = 1, \ldots, P$.
- apparent outcome is correct iff

$$\forall (p_w, s_w) \in \mathcal{W}, \ \forall (p_\ell, s_\ell) \in \mathcal{L}, \ \pi_{p_w s_w} > \pi_{p_\ell s_\ell}. \tag{1}$$

Auditing inequalities

- Auditing consists of checking those $S^2(P-1)$ inequalities.
- Some entailed by others: $\pi_{ps} > \pi_{pt}$ for $s < t$, for any method with $d(s) < d(t)$.
- E.g., if $\pi_{p_w s_w} > \pi_{p_\ell s_\ell}$, then $\pi_{p_w s_w} > \pi_{p_\ell s}$ for all $s \geq s_\ell$, and $\pi_{p_w s} > \pi_{p_\ell s_\ell}$ for all $s \leq s_w$.

Which need checking?

For party p, define

$$\begin{aligned} s_w(p) &\equiv \max\{s : (p, s) \in \mathcal{W}\} \\ s_\ell(p) &\equiv \min\{s : (p, s) \in \mathcal{L}\}. \end{aligned}$$

These are the column indices of the last seat party p wins and the first seat party p loses, respectively. One or the other might not exist for a particular party p, if it won no seats or all S seats; at most $\min(2P, S + P)$ exist. Define

$$\begin{aligned} \mathcal{W}^P &\equiv \{p : \exists s \text{ s.t. } (p, s) \in \mathcal{W}\} \\ \mathcal{L}^P &\equiv \{p : \exists s \text{ s.t. } (p, s) \in \mathcal{L}\}. \end{aligned}$$

Audit to check whether

$$\forall p \in \mathcal{W}^P,\ \forall q \in \mathcal{L}^P \text{ s.t. } p \neq q,\ \pi_{p, s_w(p)} > \pi_{q, s_\ell(q)}. \tag{2}$$

Wald's sequential probability ratio test

- Sequence of IID trials
- If null H_0 is true, chance of "success" is γ_0
- If alternative H_1 is true, chance of "success" is γ_1
- Set $T = 1$
- Repeat:
    - conduct trial
    - if "succeed," $T \to T \times \gamma_1/\gamma_0$
    - if "fail," $T \to T \times (1-\gamma_1)/(1-\gamma_0)$
    - if $T > 1/\alpha$, reject H_0 at significance level α; stop.

Ballot-polling audit: derivation

- pair of pseudo-candidates $(p_w, s_w) \in \mathcal{W}$, $(p_\ell, s_\ell) \in \mathcal{L}$
- want to determine whether $\pi_{p_w s_w} > \pi_{p_\ell s_\ell}$
- i.e., $a(p_w)/d(s_w) > a(p_\ell)/d(s_\ell)$
- i.e., $a(p_w) > a(p_\ell)\frac{d(s_w)}{d(s_\ell)}$

Ballot-polling audit: derivation

- A_p: event that a randomly selected ballot shows a vote for party p.
- $\Pr(A_p) = a(p)/B$
- If outcome is correct,

$$\Pr(A_{p_w}) \geq \frac{d(s_w)}{d(s_\ell)} \Pr(A_{p_\ell}),$$

so

$$\Pr(A_{p_w} | A_{p_w} \cup A_{p_\ell}) \geq \frac{d(s_w)}{d(s_\ell)} \Pr(A_{p_\ell} | A_{p_w} \cup A_{p_\ell}),$$

- For the outcome to be correct, need

$$\begin{aligned} \pi_{p_w|p_w p_\ell} &> (1 - \pi_{p_w|p_w p_\ell}) d(s_w)/d(s_\ell) \\ \textit{i.e.,} \quad \pi_{p_w|p_w p_\ell} &> \frac{d(s_w)}{d(s_\ell) + d(s_w)}. \end{aligned}$$

Derivation, contd.

$$\pi_{p_w|p_w p_\ell} \equiv \frac{a(p_w)}{a(p_w) + a(p_\ell)}$$

and

$$\frac{t(p_w)}{t(p_w) + t(p_\ell)} > \frac{d(s_w)}{d(s_\ell) + d(s_w)}.$$

Derivation, contd.

- Use Wald's sequential probability ratio test to test H_0:

$$\frac{a(p_w)}{a(p_w)+a(p_\ell)} \leq \frac{d(s_w)}{d(s_\ell)+d(s_w)}$$

against H_1:

$$\frac{a(p_w)}{a(p_w)+a(p_\ell)} \geq \frac{t(p_w)}{t(p_w)+t(p_\ell)}.$$

- Rejecting H_0 confirms $\pi_{p_w s_w} > \pi_{p_\ell s_\ell}$.

Derivation, contd.

- For single draw, conditional on $A_{p_w} \cup A_{p_\ell}$, if the ballot shows a vote for p_w,

$$\mathrm{LR} = \frac{\frac{t(p_w)}{t(p_w)+t(p_\ell)}}{\frac{d(s_w(p_w))}{d(s_w(p_w))+d(s_\ell(p_\ell))}}.$$

- If the ballot shows a vote for p_ℓ,

$$\mathrm{LR} = \frac{1 - \frac{t(p_w)}{t(p_w)+t(p_\ell)}}{1 - \frac{d(s_w(p_w))}{d(s_w(p_w))+d(s_\ell(p_\ell))}}$$

Ballot-polling audit: algorithm

1 Select the risk limit $\alpha \in (0,1)$, and M, the maximum number of ballots to audit before proceeding to a full hand count. Define

$$\gamma^{+}_{ps_w(p)qs_\ell(q)} \equiv \frac{t(p)}{t(p)+t(q)} \cdot \frac{d(s_w(p))+d(s_\ell(q))}{d(s_w(p))}$$

and

$$\gamma^{-}_{ps_w(p)qs_\ell(q)} \equiv \left(1-\frac{t(p)}{t(p)+t(q)}\right) \times \left(1-\frac{d(s_w(p))+d(s_\ell(q))}{d(s_w(p))}\right).$$

Set $T_{ps_w(p)qs_\ell(q)} = 1$ for all $p \in \mathcal{W}^P$ and $q \in \mathcal{L}^P$. Set $m = 0$.

2 Draw a ballot uniformly at random with replacement from those cast in the contest and increment m.

Ballot-polling audit: algorithm

3. If the ballot shows a valid vote for a reported winner $p \in \mathcal{W}^P$, then for each q in $\mathcal{L}^P$ that did not receive a valid vote on that ballot multiply $T_{ps_w(p)qs_\ell(q)}$ by $\gamma^+_{ps_w(p)qs_\ell(q)}$. Repeat for all such p.
4. If the ballot shows a valid vote for a reported loser $q \in \mathcal{L}^P$, then for each p in $\mathcal{W}^P$ that did not receive a valid vote on that ballot, multiply $T_{ps_w(p)qs_\ell(q)}$ by $\gamma^-_{ps_w(p)qs_\ell(q)}$. Repeat for all such q.
5. If any $T_{ps_w(p)qs_\ell(q)} \geq 1/\alpha$, reject the corresponding null hypothesis for each such $T_{ps_w(p)qs_\ell(q)}$. Once a null hypothesis is rejected, do not update its $T_{ps_w(p)qs_\ell(q)}$ after subsequent draws.
6. If all null hypotheses have been rejected, stop the audit: The reported results stand. Otherwise, if $m < M$, return to step 2.
7. Perform a full hand count; the results of the hand count replace the reported results.

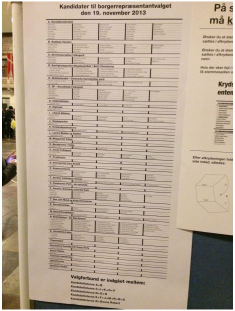
Auditing which candidates in a party are seated

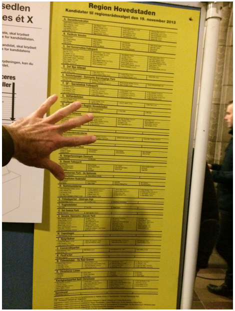
- Possible to audit this simultaneously, using the same sample.
- If a small number of votes separates two candidates in a party, required sample size may be very large.
- If ballots are sorted by party and candidate and there's a manifest, can reduce sample sizes substantially.
- Ballot-level comparison audits have much smaller sample sizes than ballot-polling audits when margins are small.
- $\exists$ sequential statistical methods for comparison audits as well.

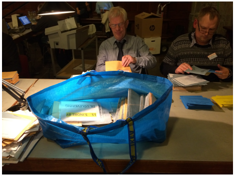
Denmark's elections are special

- Features that make auditing easier:
  - Paper ballots with excellent ballot accounting
  - Ballots have ≤ 1 [valid] vote for at most 1 party or candidate
  - Ballots are routinely sorted by party (and candidate?)
  - Bundles of ballots are small (≤ 100 ballots)
- OTOH, rules for "compensatory round" quite complicated.
  The "2%" rule is straightforward.
  "2 of 3" regional threshold requires more data.
  Collaborating with Carsten Schürmann on this.

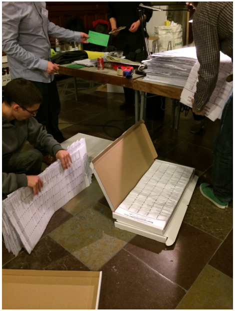
Danish Election, 19 November 2013





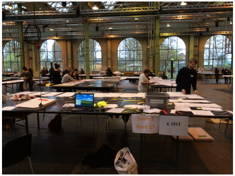
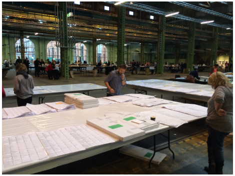


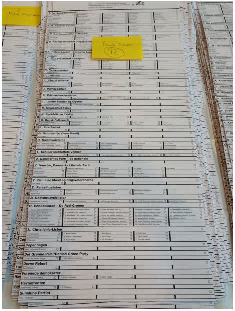


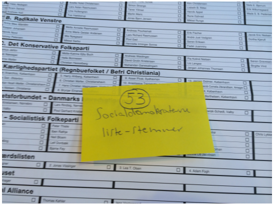







Team 1
Bord 3

Team 2
Bord 17
4. ØST






Bundles

- Do the bundles have identifiers?
- Is there a list of all sorted bundles with label info?
- Perfect ballot manifest for auditing!

Roles for Random Auditing in Denmark

- Quick rough results: snapshot of top ballot in each box of 300.
- Full ballot-polling audits.
  Requires ballot manifest but not sorting.
  Theory complete for D'Hondt rounds, Not complete for compensatory rounds.
- Ballot-level comparison audits.
  Requires ballot manifest.
  Relies on (and checks) manual sorting of ballots.
- Prepare for transition to electronic tallying?