Sharper p-Values for Stratified Election Audits

Michael J. Higgins[1] mikehiggins83@gmail.com

Based on work with Ronald L. Rivest[2] and Philip B. Stark[1]

1) Department of Statistics, University of California at Berkeley
2) Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology

Abstract

Votes in an election are often counted by machines. Winners are legally determined by what a full hand count would show. *Risk-limiting election audits* can help validate correct election outcomes. Moreover, when hand-count winners differ from machine-count winners, regardless of the source of the differences, these audits have a guaranteed pre-specified chance of escalating to a full hand-count thereby correcting the election outcome. Most states that have election audits require each county to independently select a random set of precincts for audit. For contests that span multiple counties, this amounts to a stratified random sample of precincts. Ballots in each audited precinct are counted by hand and differences between the hand count and the machine count are recorded. Using these differences, we test the null hypothesis that the machine-count winners differ from the hand-count winners. We compute sharp p-values by solving a 0-1 knapsack problem, a well-studied NP-complete integer programming problem. For most elections, a p-value can be computed in fractions of a second.

Auditing

- Machine counts and hand counts within a precinct may differ for a number of reasons: human error, software or hardware bugs, deliberate fraud, etc.
- When machines work as intended, machine counts and hand counts will be very close. Otherwise there may be large differences.
- If an audit does not uncover any precincts with large differences, and if "enough" precincts have been audited, we'd like to attest that machine-count winners and hand-count winners are the same.
- PROBLEM: If machine-count outcome deliberately different from hand-count outcome, differences between counts may be allocated to make detection by audit difficult.
- How can we quantify our confidence that the winners are the same?

Risk-Limiting Audits

- SOLUTION: Risk-limiting election audits [2].
- Risk-limiting audits have a guaranteed pre-specified chance to catch and correct differences between machine-count and hand-count winners, regardless of the source of the differences, while minimizing the audit workload.
- A risk-limiting election audit is a hypothesis test:
  H_0: Machine-count winners differ from hand-count winners.
  H_A: Winners are the same.
- Stop audit and conclude that winners are the same only if p-value is less than pre-specified α. Otherwise, keep auditing.
- Eventually, either audit stops or a full hand count occurs.

Nuts and Bolts

- Consider a contest with a single winning and losing candidate.
- Let d_j denote the *difference* in precinct j—the number of votes that the losing candidate would gain if precinct j was counted by hand.
- The machine-count winner is different than the hand-count winner if and only if

$$\sum_{j=1}^{n} d_j \geqslant \text{Margin of Victory} \qquad (1)$$

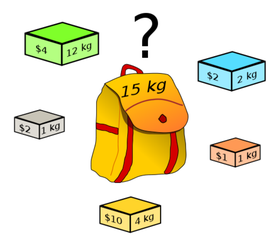
- Use $t = \max_{j \text{ audited}} d_j$ as the test statistic—good when sampling from heavily skewed distributions comprised almost entirely of near-zero values, and few, if any large values. Small $t \rightarrow$ did not observe any large differences between hand count and machine count.
- When differences satisfy (1), a p-value is calculated as follows:

$$P(T \leqslant t) = \prod_{c=1}^{C} \frac{\binom{N_c - \#_c(d>t)}{n_c}}{\binom{N_c}{n_c}} = \prod_{c=1}^{C} \prod_{k=1}^{\#_c(d>t)} \frac{\binom{N_c-k}{n_c}}{\binom{N_c-k+1}{n_c}}$$
$$= \prod_{c=1}^{C} \prod_{k=1}^{\#_c(d>t)} \left(\frac{N_c - n_c - k + 1}{N_c - k + 1} \right)$$

- C: Number of counties.
- N_c: Number of precincts in county c.
- n_c: Number of precincts sampled in county c.
- $\#_c(d > t)$: Number of precincts in county c with difference greater than t.
- PROBLEM: Need to know all values of differences to compute p-value.
- SOLUTION: Maximize p-value across all differences satisfying (1).

0-1 Knapsack Problem

- Maximizing this p-value is equivalent to solving a 0-1 knapsack problem [3].
- 0-1 Knapsack: You have a backpack that can hold up to M pounds of items. There are n items. Item j has weight w_j and price c_j.
  Choose a subset of items to put into the backpack so that the sum of weights is at most M and the sum of prices is maximized.



- For election auditing: a weight is the maximum difference that a precinct can hold and a price is an increased probability of detection by audit if that precinct has a large difference.

  Sum of weights no more than M $\iff$ Sum of differences greater than margin of victory
  Sum of prices is maximized $\iff$ Chance of detection by audit is minimized

- To compute p-value, solve the following 0-1 knapsack problem:

$$\lambda = \min_{x_j \in [0,1]} \left\{ \sum_{j=1}^{n} c_j x_j : \sum_{j=1}^{n} (w_j - t) x_j \geqslant M \right\}$$

where

$$c_j = -\log\left(\frac{(N_c - n_c - \ell + 1) \vee 0}{N_c - \ell + 1} \right).$$
$$w_j = \text{Largest difference possible between hand count and machine count.}$$
$$M = \text{Margin of Victory} - nt$$

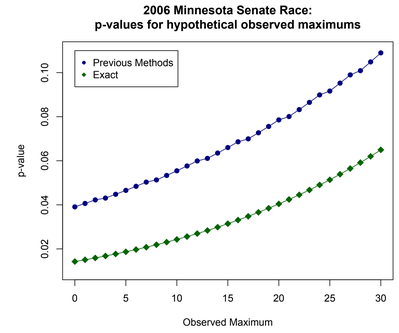
and where j is the precinct in county c with the ℓth largest value of w. Our p-value is $e^{-\lambda}$.

Data: 2006 Minnesota U.S. Senate Race

- Amy Klobuchar was the winner and Mark Kennedy was the runner-up.
- 2,217,818 ballots cast in 4,123 precincts spanning 87 counties.
- Klobuchar's margin of victory was 443,196 votes ($\approx 20\%$).
- 202 precincts audited in total. Counties randomly selected 2 to 8 precincts for audit, depending on the size of the county.
- Hennepin County, which has the most precincts (426), audited 8 precincts.
- Largest difference found: 2 votes.

Results

Exact p-values for hypothetical maximum observed differences for the 2006 Minnesota U.S. Senate race. Our p-values are substantially smaller than those found using previous methods [2].



Choosing Sample Sizes for Audits

- Can exploit 0-1 knapsack structure to reduce workload required by an audit.
- For example, by choosing sample sizes proportional to the number of precincts within a county, the audit workload can be reduced by 40% (202 $\rightarrow$ 122) while still achieving the power of the original audit.
- More complicated methods can reduce the audit workload even further.

Future Work

- R package `elec.strat` now available on CRAN.
- Apply these methods to other types of audits (e.g. financial audits).
- Single-ballot auditing may be able to reduce workload even further.

References

1 M.J. Higgins, R.L. Rivest, and P.B. Stark. "Sharper p-Values For Stratified Election Audits." *Statistics, Politics, and Policy*, 2(1), 2011. Article 7.

2 P.B. Stark. "Conservative statistical post-election audits." *Ann. Appl. Stat*, 2(2):550581, 2008.

3 Pisinger, D. and P. Toth (1998): Knapsack problems, in D.-Z. Du and P. Pardalos, eds., *Handbook of Combinatorial Optimization*, volume 1, Kluwer, Dordrecht, The Netherlands, 299428.