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\(^3\)

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.

**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 \( \sum_{j=1}^{n} d_j \geq \text{Margin of Victory} \)
- Use \( t = \max_{j\in \text{audited}} d_j \) 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 \( t \) did not observe any large differences between hand count and machine count.
- When differences satisfy (1), a \( p \)-value is calculated as follows:

\[
P(T \leq t) = \sum_{c=1}^{C} \frac{C \cdot \binom{N_c - (N - k + 1)}{n}}{N_c - k + 1} 
\]

\[
P(T > t) = \sum_{c=1}^{C} \frac{C \cdot \binom{N_c - n_c - k + 1}{n_c}}{N_c - k + 1} 
\]

- **C:** Number of counties.
- **\( N_c \):** Number of precincts in county \( c \).
- **\( n_c \):** Number of precincts sampled in county \( c \).
- **\( N \):** Number of precincts in county \( c \) with differences greater than \( t \).

**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. \( H_0 \): Winners are the same.
- Stop audit and conclude that winners are the same only if \( p \)-value is less than pre-specified \( \alpha \). Otherwise, keep auditing.
- Eventually, either audit stops or a full hand count occurs.

**Results**

- **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.

**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.

**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.

**References**