Simulations of Ballot Polling Risk-Limiting Audits

Oliver Broadrick[1] Sarah Morin[1] Grant McClearn[2]
Neal McBurnett Poorvi L. Vora[1] Filip Zagórski[3,4]

[1]Department of Computer Science, The George Washington University (odbroadrick@gmail.com)

[2]Department of Computer Science, Stanford University (grantmcc@stanford.edu)

[3]Wroclaw University of Science and Technology (filip.zagorski@gmail.com)

[4]Votifica

April 21, 2022

Outline

- Risk-Limiting Audits
  - Bravo and Minerva
- Experiments
- Results
  - Stopping Probability
  - Risk
  - Number of Ballots
- Discussion and Future Work

Post-tabulation audits

- Scanning machines are used to tabulate ballots
  - Cannot trust the machines: bugs, configuration errors, hacking
- Post-tabulation audits
- Risk-Limiting Audits

Risk-Limiting Audits

- Risk-Limiting Audit (RLA) is a post-tabulation audit that manualy checks a random sample of voters' ballots
- Relies on a voter-verified paper trail
- Sketch:
  1. election results announced
  2. sample ballots randomly
  3. check if the sample is 'statistically similar' to the announced tally
     - "Yes" - correct - stop the audit
     - "No" - incorrect - proceed to a full hand count
     - "Don't know yet" - undetermined - draw more samples (goto: 2)

Bravo

- Most commonly used ballot polling RLA
- In the two candidate case is an instance of Wald's Sequential Probability Ratio Test (SPRT)
- Is thus the most efficient RLA when ballots are drawn sequentially (i.e. ballot-by-ballot)
- Real audits are performed in rounds for which Bravo can be implemented as:
  - Selection-Ordered (SO) Bravo
  - End-of-Round (EoR) Bravo

MINERVA

- Recent RLA designed for round-by-round use
- Uses a ratio of the *tails* of the probability distributions used in BRAVO
- Known to require half the number of ballots as EoR BRAVO in a first round to achieve a large (0.90) probability of stopping
- Unknown how the audits compare for smaller stopping probability or for rounds after the first

Experiments

- Use simulations to provide evidence for theoretical claims
- R2B2 software library for round-by-round and ballot-by-ballot RLAs
- Simulate RLAs for election results from the 2020 Presidential election (all margins above 0.05)
  - $10000 = 10^4$ trials assuming the underlying election is as announced
  - $10000 = 10^4$ trials assuming the underlying election is a tie
- Risk limit: 10%
- Round schedules:
  - Bravo round sizes to achieve a chosen probability of stopping in each round given that the audit has already reached that round
  - Minerva first round sizes to achieve a chosen probability of stopping, and subsequent round sizes found by multiplying the previous round size by a constant (1.5 and 1)
- Stopping probabilities: 0.90 and 0.25

Experiments

Definition

An audit $\mathcal{A}$ takes a sample of ballots X as input and gives as output either (1) *Correct*: the audit is complete, or (2) *Uncertain*: continue the audit.

- Binary hypothesis test: H_0 (a tie) and H_a (announced results)

Definition (Risk)

The maximum risk R of audit $\mathcal{A}$ with sample $X \in \{0,1\}^*$ drawn from the true underlying distribution of ballots is $R(\mathcal{A}) = \Pr[\mathcal{A}(X) = \mathit{Correct} \mid H_0]$.

Definition (Risk Limiting Audit (α-RLA))

An audit $\mathcal{A}$ is a Risk Limiting Audit with risk limit α iff $R(\mathcal{A}) \leq \alpha$.

Experiments

Definition (Stopping Probability)

The stopping probability S_j of an audit $\mathcal{A}$ in round j is $S_j(\mathcal{A}) =$

$$\Pr[\mathcal{A}(X) = \mathit{Correct\ in\ round\ j}\ \wedge \mathcal{A}(X) \neq \mathit{Correct\ previously} \mid H_a]$$

Definition (Cumulative Stopping Probability)

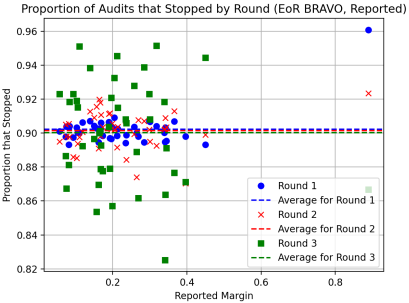
The cumulative stopping probability C_j of an audit $\mathcal{A}$ in round j is $C_j(\mathcal{A}) = \sum_{i=1}^{j} S_j$

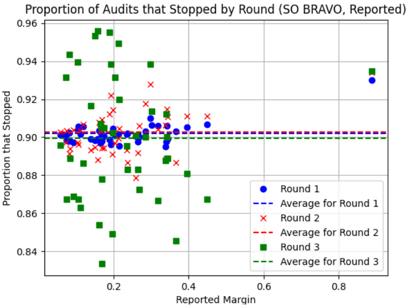
Definition (Conditional Stopping Probability)

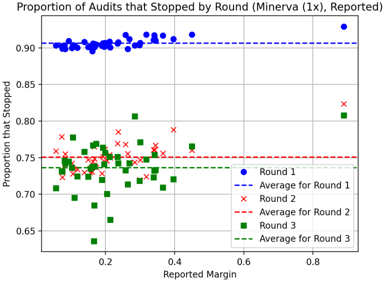
The conditional stopping probability of an audit $\mathcal{A}$ in round j is $\chi_j(\mathcal{A}) =$

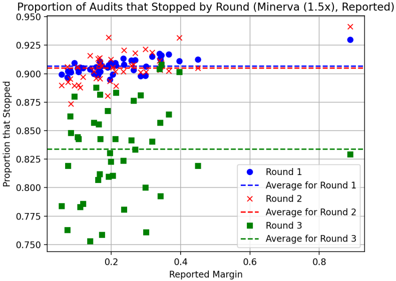
$$\Pr[\mathcal{A}(X) = \mathit{Correct\ in\ round\ j}\ \mid H_a \wedge \mathcal{A}(X) \neq \mathit{Correct\ previously}]$$

Results: Stopping Probability ($\chi_1 = 0.9$)

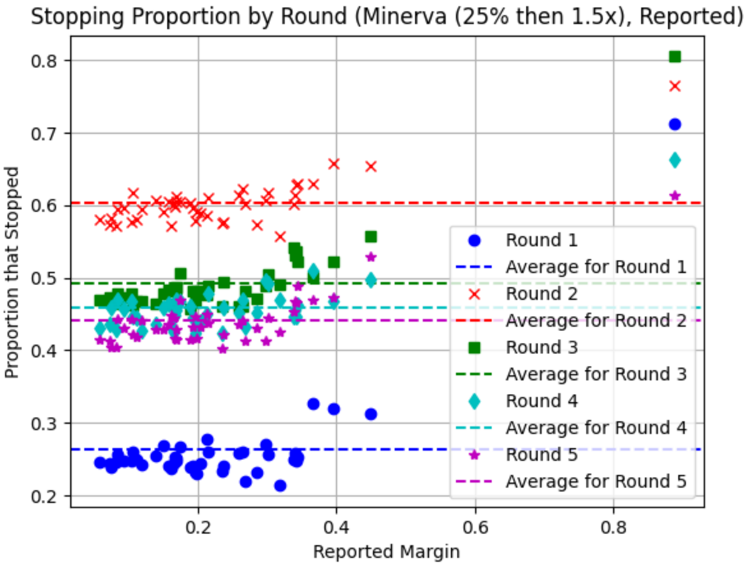




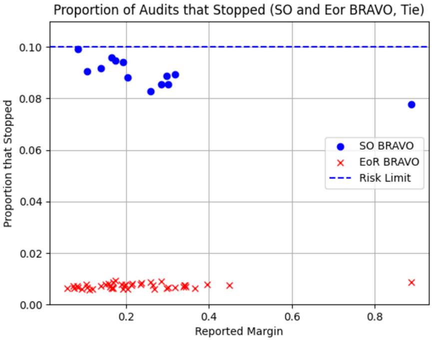


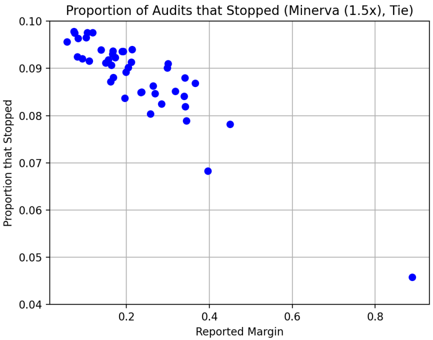


Results: Stopping Probability ($\chi_1 = 0.25$)

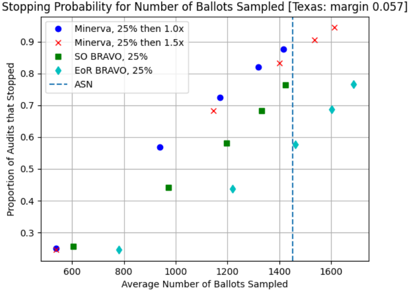


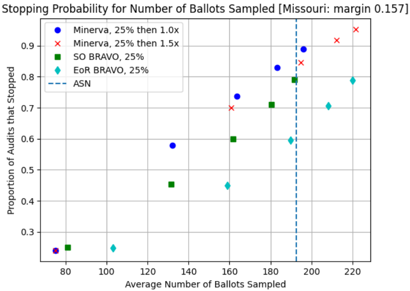
Results: Risk

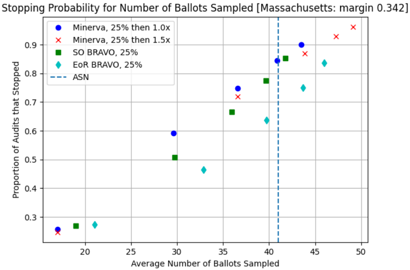


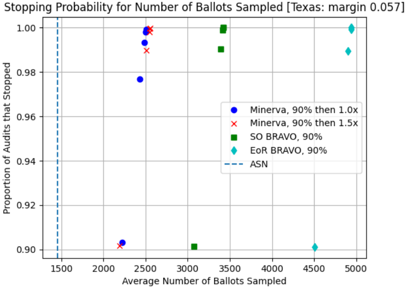


Results: Number of Ballots









Results: Round Size Proportions

For $\chi_1 = 0.25$, the number of ballots required for Minerva is smaller than that required by SO Bravo and EoR Bravo

- Improvement considerably smaller than that when the stopping probability is 0.9
- Number of ballots for SO Bravo for $\chi_1 = 0.9$ is about a third more than that required by Minerva, but for $\chi_1 = 0.25$, it requires only about a tenth more ballots than does Minerva
- Number of ballots for EoR Bravo for $\chi_1 = 0.9$ is about twice those required by Minerva, but for $\chi_1 = 0.25$, it requires only about a fourth to a half more ballots (depending on margin) than does Minerva

For $\chi_1 = 0.9$ with multiplying factor 1, Minerva consequent conditional stopping probabilities are about 0.75 and 0.74 respectively for rounds two and three.

- When the multiplying factor is 1.5, we see $\chi_2 \approx 0.91$ and $\chi_3 = 0.83$

Conclusion

- We describe use of the R2B2 library and simulator to characterize:
  - risk,
  - stopping probability, and
  - number of ballots
  - for various round schedules.
- MINERVA requires fewer ballots than either implementation of BRAVO in all cases we study, but the advantage decreases for a smaller stopping probability

Future Work

- More detailed study of the impact of different round schedules
- Simulations with other underlying distributions

Thank you