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 23, 2022

Outline

- Risk-Limiting Audits (RLAs)
  - BRAVO and MINERVA

Outline

- Risk-Limiting Audits (RLAs)
  - Bravo and Minerva
  - Both are options in Arlo, statistical election audit software used by election officials across the US

Outline

- Risk-Limiting Audits (RLAs)
  - Bravo and Minerva
  - Both are options in Arlo, statistical election audit software used by election officials across the US
- Experiments: simulated $10000 = 10^4$ audits for various margins with both
  - a correctly announced outcome
  - an underlying tie

Outline

- Risk-Limiting Audits (RLAs)
  - Bravo and Minerva
  - Both are options in Arlo, statistical election audit software used by election officials across the US
- Experiments: simulated $10000 = 10^4$ audits for various margins with both
  - a correctly announced outcome
  - an underlying tie
- Observed: stopping probability, maximum risk, number of ballots

Outline

- Risk-Limiting Audits (RLAs)
  - BRAVO and MINERVA
  - Both are options in Arlo, statistical election audit software used by election officials across the US
- Experiments: simulated $10000 = 10^4$ audits for various margins with both
  - a correctly announced outcome
  - an underlying tie
- Observed: stopping probability, maximum risk, number of ballots
- Results:
  - MINERVA requires fewer ballots over multiple rounds for both
    - high stopping probability (0.90)
    - low stopping probability (0.25)
  - less advantage for the lower stopping probability

Outline

- Risk-Limiting Audits (RLAs)
  - Bravo and Minerva
  - Both are options in Arlo, statistical election audit software used by election officials across the US
- Experiments: simulated $10000 = 10^4$ audits for various margins with both
  - a correctly announced outcome
  - an underlying tie
- Observed: stopping probability, maximum risk, number of ballots
- Results:
  - Minerva requires fewer ballots over multiple rounds for both
    - high stopping probability (0.90)
    - low stopping probability (0.25)
  - less advantage for the lower stopping probability
- Discussion and Future Work

Risk-Limiting Audits

- Scanners are used to tabulate ballots

Risk-Limiting Audits

- Scanners are used to tabulate ballots
  - Cannot trust the machines: bugs, configuration errors, hacking

Risk-Limiting Audits

- Scanners are used to tabulate ballots
  - Cannot trust the machines: bugs, configuration errors, hacking
- Compliance and tabulation audits

Risk-Limiting Audits

- Scanners are used to tabulate ballots
  - Cannot trust the machines: bugs, configuration errors, hacking
- Compliance and tabulation audits
- Risk-Limiting Audits
  - Given that the election outcome is incorrect, the probability with which the audit stops, declaring the outcome correct, is at most the risk limit, α.

Ballot Polling Risk-Limiting Audits: A Procedure

- Is a manual audit, which relies on a voter-verified paper trail and successfully completed compliance audits

Ballot Polling Risk-Limiting Audits: A Procedure

- Is a manual audit, which relies on a voter-verified paper trail and successfully completed compliance audits
- Sketch:
    1. Election results announced

Ballot Polling Risk-Limiting Audits: A Procedure

- Is a manual audit, which relies on a voter-verified paper trail and successfully completed compliance audits
- Sketch:
  1. Election results announced
  2. In a public procedure, sample ballots at random and manually interpret them

Ballot Polling Risk-Limiting Audits: A Procedure

- Is a manual audit, which relies on a voter-verified paper trail and successfully completed compliance audits
- Sketch:
  1. Election results announced
  2. In a public procedure, sample ballots at random and manually interpret them
  3. Compute a pre-specified error measure, the maximum risk, and compare to the risk limit
     - If smaller, stop the audit
     - Else, sample more (goto 2)

Bravo

- Most commonly used ballot polling RLA

Bravo

- Most commonly used ballot polling RLA
- In the two candidate case is an instance of Wald's Sequential Probability Ratio Test (SPRT)

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 the decision of whether to stop the audit is made after each ballot is drawn (ballot-by-ballot)

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 the decision of whether to stop the audit is made after each ballot is drawn (ballot-by-ballot)
- In real audits, decisions are taken after many ballots are drawn (round-by-round)

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 the decision of whether to stop the audit is made after each ballot is drawn (ballot-by-ballot)
- In real audits, decisions are taken after many ballots are drawn (round-by-round)
- BRAVO can be implemented as:
  - Selection-Ordered (SO) BRAVO, where ballot selection order is retained, and the decisions are taken as though the audit were ballot-by-ballot
  - End-of-Round (EoR) BRAVO, where the decision using the BRAVO stopping rule is taken once, after the entire round of ballots is drawn

Minerva

- Recent RLA designed for round-by-round use

Minerva

- Recent RLA designed for round-by-round use
- Bravo uses the likelihood ratio; that is the ratio of points on two probability distribution functions

Minerva

- Recent RLA designed for round-by-round use
- Bravo uses the likelihood ratio; that is the ratio of points on two probability distribution functions
- Minerva uses a ratio of the *tails* of the pdfs used in Bravo

Minerva

- Recent RLA designed for round-by-round use
- Bravo uses the likelihood ratio; that is the ratio of points on two probability distribution functions
- Minerva uses a ratio of the *tails* of the pdfs used in Bravo
  - Shown to be risk-limiting if all round sizes are pre-committed, before the audit begins

Minerva

- Recent RLA designed for round-by-round use
- Bravo uses the likelihood ratio; that is the ratio of points on two probability distribution functions
- Minerva uses a ratio of the *tails* of the pdfs used in Bravo
  - Shown to be risk-limiting if all round sizes are pre-committed, before the audit begins
  - In a first round chosen to give a 0.90 probability of stopping, Minerva requires
    - 50% as many ballots as EoR Bravo
    - 70-80% as many ballots as SO Bravo

Minerva

- Recent RLA designed for round-by-round use
- Bravo uses the likelihood ratio; that is the ratio of points on two probability distribution functions
- Minerva uses a ratio of the *tails* of the pdfs used in Bravo
  - Shown to be risk-limiting if all round sizes are pre-committed, before the audit begins
  - In a first round chosen to give a 0.90 probability of stopping, Minerva requires
    - 50% as many ballots as EoR Bravo
    - 70-80% as many ballots as SO Bravo
  - Unknown how the audits compare for smaller stopping probability or for rounds after the first

Experiments

- Use simulations to provide evidence for theoretical claims

Experiments

- Use simulations to provide evidence for theoretical claims
- R2B2 software library for round-by-round and ballot-by-ballot RLAs

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 was as announced
  - $10000 = 10^4$ trials assuming the underlying election was a tie

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 was as announced
  - $10000 = 10^4$ trials assuming the underlying election was a tie
- Risk limit: 10%

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 was as announced
  - $10000 = 10^4$ trials assuming the underlying election was 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

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 was as announced
  - $10000 = 10^4$ trials assuming the underlying election was 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)

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 was as announced
  - $10000 = 10^4$ trials assuming the underlying election was 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.

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)

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)
- The tie is the hardest incorrect outcome to detect

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)
- The tie is the hardest incorrect outcome to detect
- Probability of stopping given a tie should be low

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)
- The tie is the hardest incorrect outcome to detect
- Probability of stopping given a tie should be low
- Probability of stopping given a correctly announced outcome should be high for as few ballots as possible

Experiments

Definition (Maximum Risk)

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

Experiments

Definition (Maximum Risk)

The maximum risk R of audit $\mathcal{A}$ with sample $X \in \{0,1\}^*$ drawn from the 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 $S_j(\mathcal{A}) =$

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

Experiments

Definition (Stopping Probability)

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

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

Definition (Cumulative Stopping Probability)

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

Experiments

Definition (Stopping Probability)

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

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

Definition (Cumulative Stopping Probability)

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

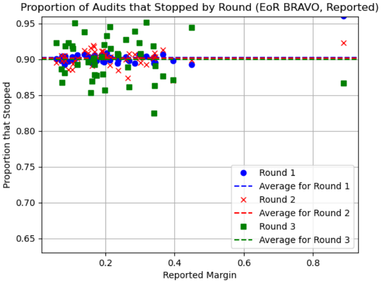
Definition (Conditional Stopping Probability)

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

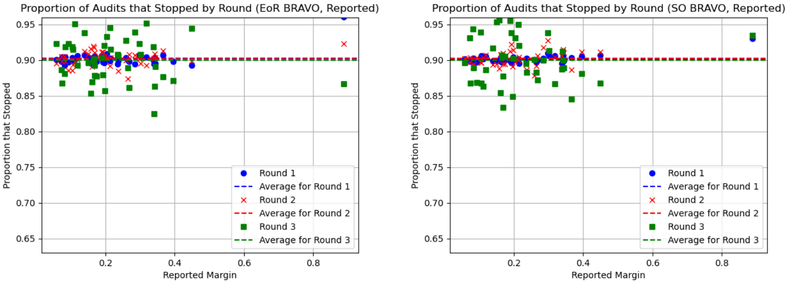
$$\Pr[\mathcal{A}(X) = \textit{Correct in round } j \ \mid H_a \wedge \mathcal{A}(X) \neq \textit{Correct previously}]$$

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

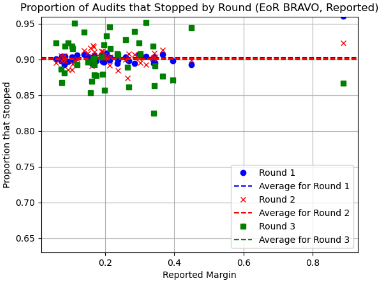
Results: Stopping Probability ($\chi_1 = 0.9$)

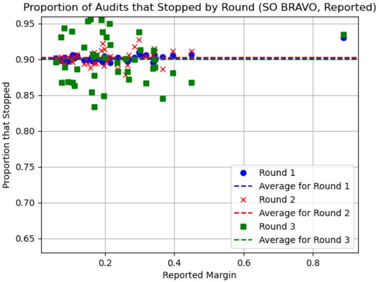


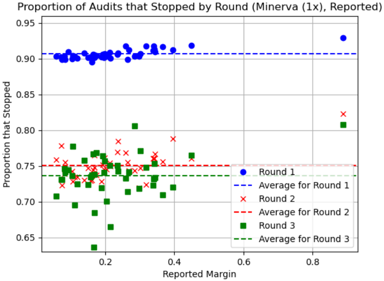
Results: Stopping Probability ($\chi_1 = 0.9$)



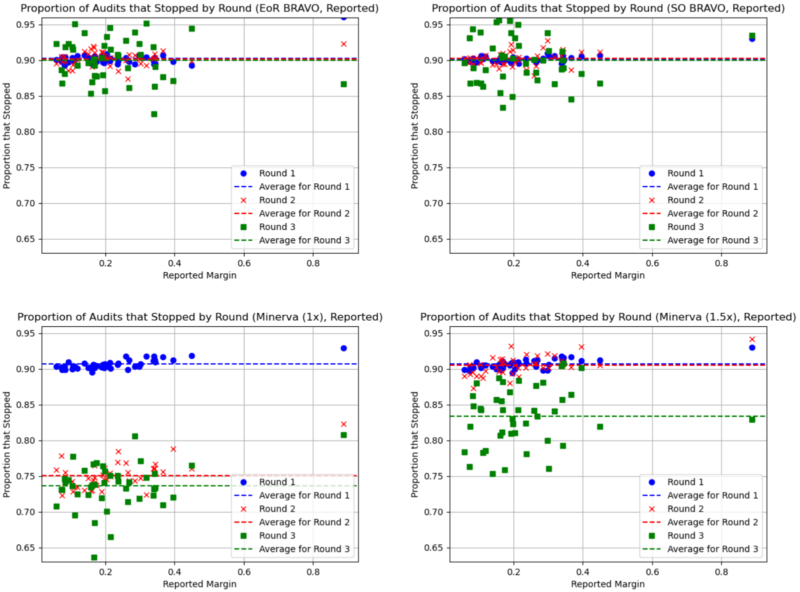
Results: Stopping Probability ($\chi_1 = 0.9$)



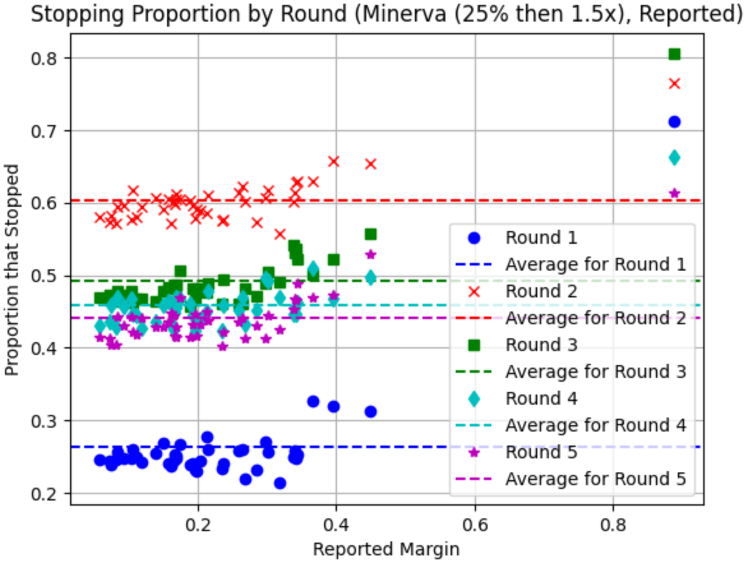




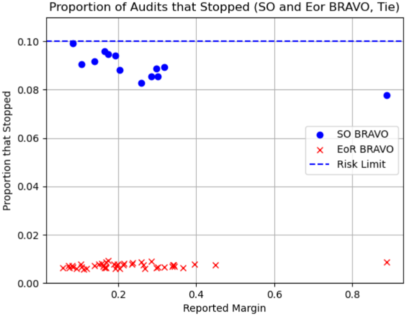
Results: Stopping Probability ($\chi_1 = 0.9$)



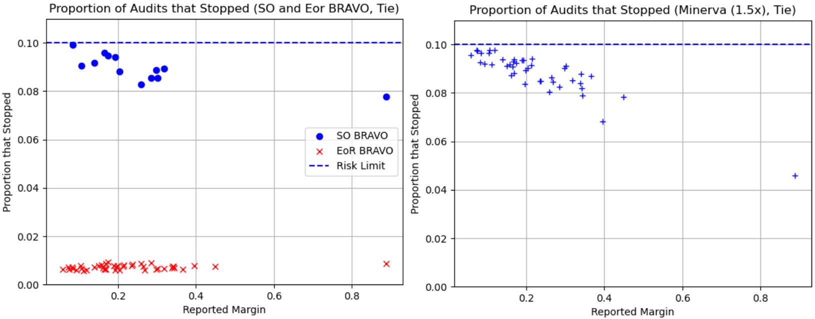
Results: Stopping Probability ($\chi_1 = 0.25$)



Results: Risk ($\chi_1 = 0.9$)

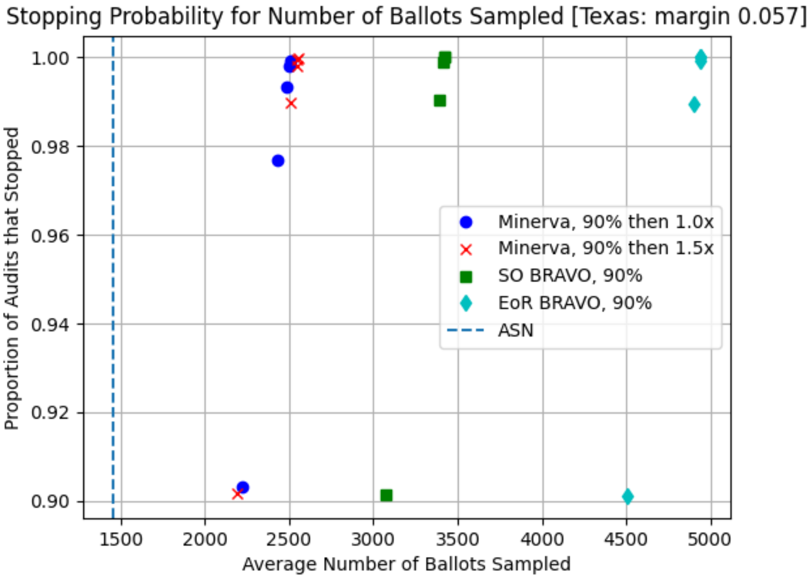


Results: Risk ($\chi_1 = 0.9$)

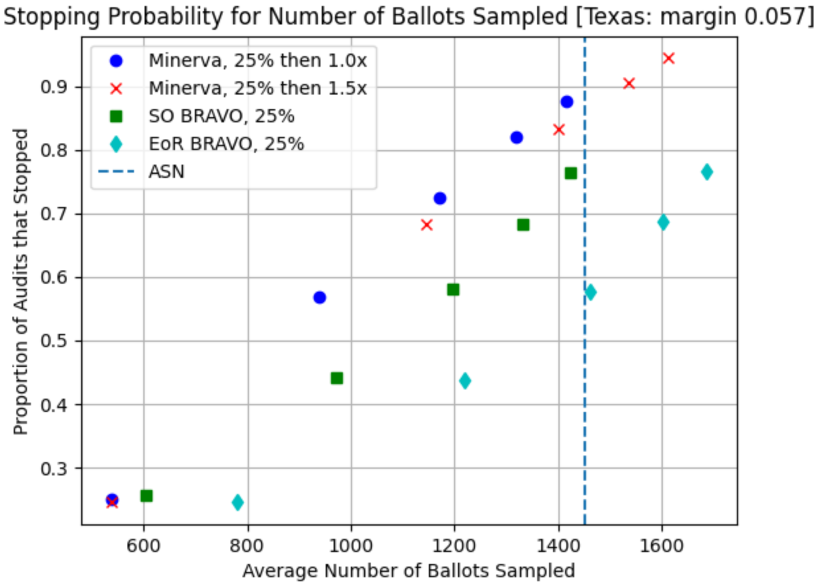


Results: Number of Ballots ($\chi_1 = 0.9$)

Results: Number of Ballots ($\chi_1 = 0.9$)



Results: Number of Ballots ($\chi_1 = 0.25$)



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

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 $\chi_1 = 0.9$

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 $\chi_1 = 0.9$
- SO BRAVO:
  - for $\chi_1 = 0.9$ requires a third more than does MINERVA
  - for $\chi_1 = 0.25$ requires a tenth more than does MINERVA

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 $\chi_1 = 0.9$
- SO BRAVO:
  - for $\chi_1 = 0.9$ requires a third more than does MINERVA
  - for $\chi_1 = 0.25$ requires a tenth more than does MINERVA
- EoR BRAVO:
  - for $\chi_1 = 0.9$ requires twice as many as MINERVA
  - for $\chi_1 = 0.25$ requires a fourth to a half more (depending on margin) than does MINERVA

Results: Minerva Stopping Probabilities

For $\chi_1 = 0.9$, Minerva consequent conditional stopping probabilities for rounds two and three are respectively:

Results: MINERVA Stopping Probabilities

For $\chi_1 = 0.9$, MINERVA consequent conditional stopping probabilities for rounds two and three are respectively:

- with multiplying factor 1, $\chi_2 \approx 0.75$ and $\chi_3 \approx 0.74$

Results: Minerva Stopping Probabilities

For $\chi_1 = 0.9$, Minerva consequent conditional stopping probabilities for rounds two and three are respectively:

- with multiplying factor 1, $\chi_2 \approx 0.75$ and $\chi_3 \approx 0.74$
- with multiplying factor 1.5, $\chi_2 \approx 0.91$ and $\chi_3 \approx 0.83$

Conclusion

- We describe use of the R2B2 library and simulator to characterize:

Conclusion

- We describe use of the R2B2 library and simulator to characterize:
  - maximum risk,

Conclusion

- We describe use of the R2B2 library and simulator to characterize:
  - maximum risk,
  - stopping probability, and

Conclusion

- We describe use of the R2B2 library and simulator to characterize:
  - maximum risk,
  - stopping probability, and
  - number of ballots

Conclusion

- We describe use of the R2B2 library and simulator to characterize:
  - maximum risk,
  - stopping probability, and
  - number of ballots
  - for various round schedules.

Conclusion

- We describe use of the R2B2 library and simulator to characterize:
  - maximum 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 for each round

Future Work

- More detailed study of the impact of different round schedules

Future Work

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

Thank you

odbroadrick@gmail.com