PROVIDENCE: a Flexible Round-by-Round Risk-Limiting Audit

Your N. Here

Your Institution

Second Name

Second Institution

Abstract

A Risk-Limiting Audit (RLA) draws consecutive random samples (rounds) of paper election ballots. After each round, a rigorous error criterion is used to decide whether to stop (confirming the announced outcome) or continue (draw another round). The most commonly used ballot polling RLA, BRAVO [7], requires the fewest ballots when each round is just one ballot. Recent RLA MINERVA [17] draws significantly fewer ballots than BRAVO (half as many for larg round sizes) but requires the round schedule to be fixed before the audit begins. We present PROVIDENCE, an audit with the efficiency of MINERVA and flexibility of BRAVO. We prove PROVIDENCE is Risk-Limiting in the presence of an adversary who can choose subsequent round sizes given knowledge of previous samples, and we present simulation results demonstrating its efficiency. Finally, we describe the pilot use of PROVIDENCE by the Rhode Island Board of Elections. An implementation of PROVIDENCE is provided in the open source R2B2 library including novel functionality that allows election officials to select round sizes with information as to the probability of a round giving a misleading outcome, a desirable feature given political context[1].

1 Introduction

The literature contains numerous descriptions of vulnerabilities in deployed voting systems, and it is not possible to be certain that any system, however well-designed, will perform as expected in all instances. For this reason, *evidence-based elections* [13] aim to produce trustworthy and compelling evidence of the correctness of election outcomes, enabling the detection of problems with high probability. One way to implement an evidence-based election is to use a well-curated voter-verified paper trail, compliance audits, and a rigorous tabulation audit of the election outcome, known as a risk-limiting audit (RLA) [6]. An RLA is an audit which guarantees that the probability of concluding that an election outcome is correct, given that it is not, is below a pre-determined value known as the risk limit of the audit, independent of the true, unknown vote distribution of the underlying election. Over a dozen states in the US have seriously explored the use of RLAs—some have pilot programs, some allow RLAs to satisfy a general audit requirement and some have RLAs in statute.

This paper focuses on ballot-polling RLAs, which require a large number of ballots relative to comparison RLAs but do not rely on any special features of the election technology. Since comparison RLAs are not always feasible, ballot-polling audits remain an important resource and have been used in a number of US state pilots (California, Georgia, Indiana, Michigan, Ohio, Pennsylvania and elsewhere). In the general ballot-polling RLA, a number of ballots are drawn and tallied in what is termed a *round* of ballots [17]. A statistical measure is then computed to determine whether there is sufficient evidence to declare the election outcome correct within the pre-determined risk limit. Because the decision is made after drawing a round of ballots, the audit is termed a *round-by-round (R2) audit*. The special case when round size is one—that is, stopping decisions are made after each ballot draw—is a *ballot-by-ballot (B2) audit*.

The BRAVO audit is designed for use as a B2 audit: it requires the smallest expected number of ballots when the true tally of the underlying election is as announced and stopping decisions are made after each ballot draw. In practice, election officials draw many ballots at once, and the BRAVO stopping rule needs to be modified for use in an R2 audit that is not B2. There are two obvious approaches. The B2 stopping condition can be applied once at the end of each round: End-of Round (EoR) BRAVO. Alternatively, the order of ballots in the sample can be tracked by election officials and the B2 BRAVO stopping condition can be applied retroactively after each ballot drawn: Selection-Ordered (SO) BRAVO. SO BRAVO requires fewer ballots on average than EoR BRAVO but requires the work of tracking the order of ballots rather

---

than just their tally.

MINERVA was designed for R2 audits and applies its stopping rule once for each round. Thus it does not require the tracking of ballots that SO BRAVO does. Zagórski *et al.* [17] prove that MINERVA is a risk-limiting audit and requires fewer ballots to be sampled than EoR BRAVO when an audit is performed in rounds, the two audits have the same pre-determined (before any ballots are drawn) round schedule and the underlying election is as announced. They also present first-round simulations which show that MINERVA draws fewer ballots than SO BRAVO in the first round for first round sizes with a large probability of stopping when the (true) underlying election is as announced. Broadrick *et al.* provide further simulations that show MINERVA requires fewer ballots over multiple rounds and for lower stopping probability.

While more efficient than BRAVO, MINERVA requires that the round schedule is fixed in advance of the audit. This may lead to significant unnecessary work. For example, suppose the fixed round schedule for MINERVA is $10,000$ ballots per round (a reasonable number in a state-wide contest with a narrow margin), and the first round sample finds a number of ballots for the winner just short of that necessary for the audit to stop. It then may be sufficient to draw a very small second round and still stop with high probability in the second round, but the additional $10,000$ ballots must be drawn. In contrast, subsequent BRAVO round sizes can be chosen based on preceding samples.

An open question is whether a ballot polling RLA exists with the efficiency of MINERVA and this flexibility of BRAVO.

1.1 Our Contributions

We present PROVIDENCE, and provide the following:

1. Proof that PROVIDENCE is an RLA and resistant to an adversary who can choose subsequent round sizes with knowledge of previous samples

2. Simulations of PROVIDENCE, MINERVA, SO BRAVO, and EoR BRAVO which show that PROVIDENCE has similar efficiency to MINERVA, both greater than either implementation of BRAVO

3. Results and analysis from the use of PROVIDENCE in a pilot audit in Rhode Island

4. Open source implementation of PROVIDENCE

2 Related work

The BRAVO audit [7] is a well-known ballot polling audit which has been used in numerous pilot and real audits. When used to audit a two-candidate election, it is an instance of Wald's sequential probability ratio test (SPRT) [15], and inherits the SPRT property of being the most efficient test (requiring the smallest expected number of ballots) if the election is as announced. The model for BRAVO and the SPRT is, however, that of a sequential audit: a sample of size one is drawn, and a decision of whether to stop the audit or not is taken. Real election audits invest in drawing large numbers of ballot, called rounds, before making stopping decisions because sequentially sampling individual ballots has significant overhead (unsealing storage boxes and searching for individual ballots). It is possible to apply BRAVO to the sequence of ballots in a round if the sequential order is retained. This is not, however, the most efficient possible use of the drawn sample because information in consequent ballots is ignored when applying BRAVO to ballots that were drawn earlier in the sample.

We do know a great deal about the properties of BRAVO. The risk limiting property of BRAVO follows from the similar property of the SPRT. Stopping probabilities for BRAVO may be estimated as implemented in [14]; this method is due to Mark Lindeman and uses quadratic approximations. A later method for stopping probability estimates presented by Zagórski *et al.* [16, 17] uses a similar technique for narrow margins and a separate algorithm for wider margins, the results of which match simulation results reported by Lindeman *et al.* [7, Table 1].

The MINERVA audit [16, 17] was developed for large first round sizes which enable election officials to be done in one round with large probability. It uses information from the entire sample, and has been proven to be risk limiting when the round schedule for the audit is determined before the audit begins. That is, information about the actual ballots drawn in the first round cannot inform future round sizes. First-round sizes for a 0.9 stopping probability when the election is as announced have been computed for a wide range of margins and are smaller than those for EoR and SO BRAVO. First round simulations of MINERVA [16] demonstrate that its first-round properties—regarding the probabilities of stopping when the underlying election is tied and when it is as announced—are as predicted for first round sizes with stopping probability 0.9. Additional simulations [3] have shown that MINERVA requires fewer ballots than EoR and SO BRAVO over multiple rounds and for smaller stopping probability. As expected, the advantage of MINERVA decreases for smaller stopping probability (smaller round sizes) as such round schedules approach the B2 round schedule $(1,1,1,\dots)$ for which BRAVO is known to be most efficient.

Ballot polling audit simulations have been used to familiarize election officials and the public with the approach [12] and to investigate the efficiency of audits [2, 5, 8, 9]. Some workload measurements have been made [4]. While total ballots sampled can give naive workload estimates [11], Bernhard presents a more complex workload estimation model [1].

2.1 Model

An audit $\mathcal{A}$ is a function that takes as input the sample of ballots and outputs either (1) *Correct: stop the audit* or (2) *Undetermined: sample more ballots*. BRAVO and MINERVA are modeled as binary hypothesis tests where the null hypothesis H_0 corresponds to a true underlying tie and the alternative hypothesis H_a corresponds to the underyling election being as announced. (With an even total number of ballots, H_0 corresponds to the announced loser winning by one ballot.) Thus the null hypothesis is the incorrectly announced outcome which is most difficult to detect. A risk-limiting audit requires that in the event of an underlying tie, the probability of failing to detect the error is below the risk limit.

Definition 1 (Risk Limiting Audit (α-RLA)). *An audit $\mathcal{A}$ is a Risk Limiting Audit with risk limit α iff for sample X*

$$Pr[\mathcal{A}(X) = Correct \,|\, H_0] \leq \alpha$$

The stopping conditions of BRAVO and MINERVA rely on the following ratios.

Definition 2 (BRAVO Ratio). *The BRAVO audit uses the ratio σ. Consider a sample size of n ballots with k for the reported winner. The proportion of ballots for the reported winner under the alternative hypothesis and null hypothesis are p_a and p_0 respectively.*

$$\sigma(k, p_a, p_0, n) \triangleq \frac{p_a^k (1-p_a)^{n-k}}{p_0^k (1-p_0)^{n-k}} \tag{1}$$

In BRAVO, $p_0 = \frac{1}{2}$. A BRAVO audit outputs correct if and only if

$$\sigma(k, p_a, \frac{1}{2}, n) \geq \frac{1}{\alpha}.$$

If testing the BRAVO stopping condition after each individual ballot is drawn (a B2 BRAVO audit), σ is equivalent to the likelihood ratio:

$$\frac{Pr[K = k | H_a, n]}{Pr[K = k | H_0, n]} = \frac{\binom{n}{k} p_a^k (1-p_a)^{n-k}}{\binom{n}{k} (\frac{1}{2})^n} = \sigma(k, p_a, \frac{1}{2}, n)$$

It now becomes useful to have shorthand for a sequence of round sizes and a sequence of winner ballot tallies. We use:

$$\boldsymbol{k_j} \triangleq (k_1, k_2, \ldots, k_j)$$

$$\boldsymbol{n_j} \triangleq (n_1, n_2, \ldots, n_j)$$

Where BRAVO uses a ratio of points on conditional probability distributions, MINERVA uses a ratio of *tails* of conditional probability distributions.

Definition 3 (MINERVA Ratio). *The R2 MINERVA audit uses the ratio τ_j. We use cumulative round sizes $\boldsymbol{n_j}$, with corresponding $\boldsymbol{k_j}$ ballots for the reported winner in reach round.*

The proportion of ballots for the reported winner under the alternative hypothesis and null hypothesis are p_a and p_0 respectively.

$$\tau_j(k_j, p_a, p_0, \boldsymbol{n_j}, \alpha) \triangleq$$
$$\frac{Pr[K_j \geq k_j \wedge \forall_{i<j}(\mathcal{A}(X_i) \neq Correct) \,|\, H_a, \boldsymbol{n_j}]}{Pr[K_j \geq k_j \wedge \forall_{i<j}(\mathcal{A}(X_i) \neq Correct) \,|\, H_0, \boldsymbol{n_j}]} \tag{2}$$

3 PROVIDENCE

Definition 4 (($\alpha, p_a, p_0, k_{j-1}, n_{j-1}, n_j$)-PROVIDENCE). *For cumulative round size n_i for round i and a cumulative k_i ballots for the reported winner found in round i, the R2 PROVIDENCE stopping rule for the jth round is:*

$$\mathcal{A}(X_j) = \begin{cases} Correct & \omega_j(k_j, k_{j-1}, p_a, p_0, n_j, n_{j-1}) \geq \frac{1}{\alpha} \\ Undetermined & else \end{cases}$$

where $\omega_1 \triangleq \tau_1$ and for $j \geq 2$, we define ω_j as follows:

$$\omega_j(k_j, k_{j-1}, p_a, p_0, n_j, n_{j-1}) \triangleq$$
$$\sigma(k_{j-1}, p_a, p_0, n_{j-1}) \cdot \tau_1(k_j - k_{j-1}, p_a, p_0, n_j - n_{j-1}) \tag{3}$$

Notice that for $j \geq 2$, unlike τ_j, computing ω_j requires no convolution. We now we prove that PROVIDENCE is risk-limiting using lemmas from basic algebra in Appendix A.

Theorem 1. *An ($\alpha, p_a, p_0, k_{j-1}, n_{j-1}, n_j$)-PROVIDENCE audit is an α-RLA.*

Proof. Let $\mathcal{A} = (\alpha, p_a, p_0, k_{j-1}, n_{j-1}, n_j)$-PROVIDENCE. Let $\boldsymbol{n_j}$ be the cumulative roundsizes used in this audit, with corresponding cumulative tallies of ballots for the reported winner $\boldsymbol{k_j}$. For round $j = 1$, by Definitions 4 and 3, we see that the $\mathcal{A} = $ Correct (the audit stops) only when

$$\tau_1(k_1, p_a, p_0, n_1) = \frac{Pr[K_1 \geq k_1 \,|\, H_a, n_1]}{Pr[K_1 \geq k_1 \,|\, H_0, n_1]} \geq \frac{1}{\alpha}.$$

By Lemma 5, we see that this is equivalent to the following:

$$\frac{Pr[K_1 \geq k_{min,1} \,|\, H_a, n_1]}{Pr[K_1 \geq k_{min,1} \,|\, H_0, n_1]} \geq \frac{1}{\alpha}.$$

For any round $j \geq 2$, by Definition 4 and Lemma 5, $\mathcal{A} = $ Correct (the audit stops) only when

$$\omega_j(k_j, k_{j-1}, p_a, p_0, n_j, n_{j-1}, \alpha) \triangleq \sigma(k_{j-1}, p_a, p_0, n_{j-1}) \cdot \tau_1(k_j - k_{j-1}, p_a, p_0, \ldots$$

By Lemma 6 and Definition 3, this is equivalent to

$$\frac{Pr[\boldsymbol{k_{j-1}} \,|\, H_a] \cdot Pr[K_j \geq k_j \,|\, \boldsymbol{k_{j-1}}, H_a, nj]}{Pr[\boldsymbol{k_{j-1}} \,|\, H_0] \cdot Pr[K_j \geq k_j \,|\, \boldsymbol{k_{j-1}}, H_0, nj]} \geq \frac{1}{\alpha}.$$

By Lemma 5 and Definition 4, we see that there exists a $k_{min,j} \leq k_j$ for which

$$\frac{Pr[\boldsymbol{k_{j-1}} \,|\, H_a] \cdot Pr[K_j \geq k_j \,|\, \boldsymbol{k_{j-1}}, H_a, nj]}{Pr[\boldsymbol{k_{j-1}} \,|\, H_0] \cdot Pr[K_j \geq k_j \,|\, \boldsymbol{k_{j-1}}, H_0, nj]} \geq$$

$$\frac{\Pr[\boldsymbol{k}_{j-1} \mid H_a] \cdot Pr[K_j \geq k_{min,j} \mid \boldsymbol{k}_{j-1}, H_a, nj]}{\Pr[\boldsymbol{k}_{j-1} \mid H_0] \cdot Pr[K_j \geq k_{min,j} \mid \boldsymbol{k}_{j-1}, H_0, nj]} \geq$$

$$\frac{\Pr[\boldsymbol{k}_{j-1} \mid H_a] \cdot Pr[K_j \geq k_j \mid \boldsymbol{k}_{j-1}, H_a, nj]}{\Pr[\boldsymbol{k}_{j-1} \mid H_0] \cdot Pr[K_j \geq k_j \mid \boldsymbol{k}_{j-1}, H_0, nj]} \geq \frac{1}{\alpha}.$$

Taking the sum over all possible audit histories, we get

$$\frac{\sum_{\boldsymbol{k}_j} \Pr[\boldsymbol{k}_{j-1} \mid H_a] \cdot Pr[K_j \geq k_{min,j} \mid \boldsymbol{k}_{j-1}, H_a, nj]}{\sum_{\boldsymbol{k}_j} \Pr[\boldsymbol{k}_{j-1} \mid H_0] \cdot Pr[K_j \geq k_{min,j} \mid \boldsymbol{k}_{j-1}, H_0, nj]} \geq \frac{1}{\alpha}.$$

Finally, because the total probability of stopping the audit under the alternative hypothesis is less than 1, we get

$$\frac{Pr[\mathcal{A} = \text{Correct} \mid H_a]}{Pr[\mathcal{A} = \text{Correct} \mid H_0]} \geq \frac{1}{\alpha}$$

$$Pr[\mathcal{A} = \text{Correct} \mid H_0] \leq Pr[\mathcal{A} = \text{Correct} \mid H_a] \cdot \alpha \leq \alpha.$$

$\square$

3.1 Resistance against an adversary choosing round sizes

I don't feel very comfortable with how to word/present this argument... I have not yet made an attempt at doing so.

4 Simulations

We use simulations to provide additional evidence for theoretical claims and gain insight into audit behavior. As in [**?**], we use margins from the 2020 US Presidential election, state-wide pairwise margins between the leading two candidates of 5% or more. Narrower margins are computationally expensive, especially for the simulations with an underyling tie which quickly increase in sample size. We use the simulator in the R2B2 software library [10]. We perform $10000 = 10^4$ trials per margin for both an underyling announced outcomes and an underyling tie.

The simulations with an underlying tie give an estimate of maximum risk as shown in Figure 1. For all margins, the estimated risks were less than the risk limit.

Simulations with the underlying ballot distribution as announced provide insight into stopping probability and number of ballots drawn. Figure 2 shows that the stoping probabilities over the first rounds are near and slightly above 90% as expected, since our software chose round sizes to give at least a 90% conditional stopping probability. Figure 3 shows that the probability of stopping as a function of number of ballots sampled. Points above (higher probability of stopping) and to the left (fewer ballots) represent more efficient audits. As shown, PROVIDENCE has comparable efficiency to MINERVA, while both are significantly more efficient than either implementation of BRAVO. Note that a difference in twice as many ballots could be negligible in a contest with a wide margin. In a contest with a narrow margin (in the 2020 US
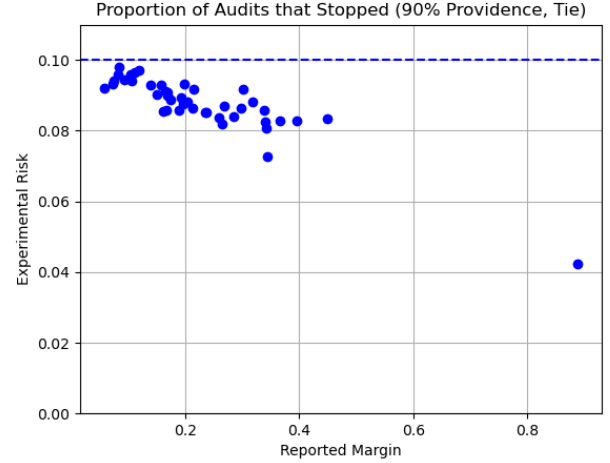

Figure 1: The proportion of simulated audits with an underlying tie that stopped in one of the five simulated rounds.

ballots	PROVIDENCE	MINERVA	EoR BRAVO	SO BRAVO
140	**4.18%**	**4.18%**	**5.41%**	36.6%

Table 1: Risk measures for the drawn first round of 140 ballots in the Providence, RI pilot audit. Risks in bold meet the risk-limit (10%) and thus correspond to audits that would stop.

Presidential election, several states had margins less than 3%) the difference in number of ballots sampled could be weeks of work. Section 6 discusses workload in more depth.

5 Pilot use

The Rhode Island Board of Elections performed a pilot audit in Providence in February 2022. The contest audited was a single yes-or-no question from the November 2021 special election. The question had announced margin 25.67%. The audit used risk-limit 10%. A first round size of 140 ballots with large probability of stopping (95%) was selected in order to give the potential for more interesting analysis afterwards; the large margin made a first round with large stopping probability practical. Selection order was tracked for the sake of analysis. As expected, the audit concluded in the first round with a PROVIDENCE risk of 4.18%. Table 1 shows risk measures for the drawn sample using MINERVA and BRAVO (both EoR and SO).

TODO: Add examples of how the audits perform for various hypothetical round schedules. I wait to do this until I'm done with the workload estimates since the examples here
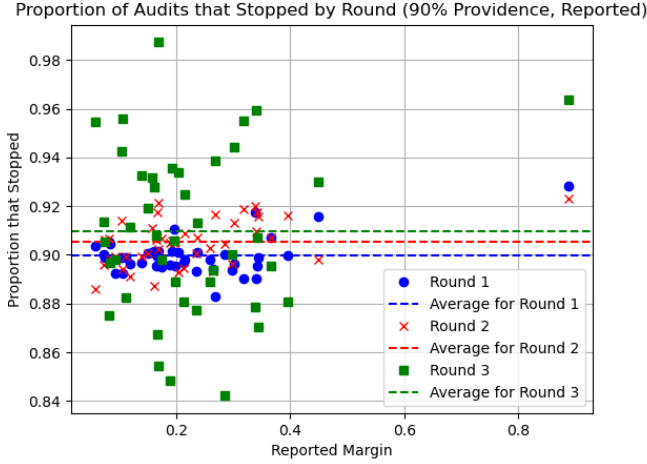
Figure 2: The proportion of simulated audits with an margin as announced that stopped in one of the five simulated rounds.
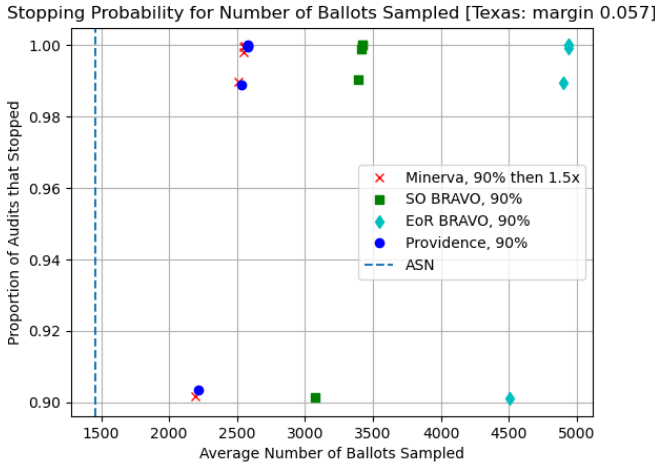

Figure 3: For all five rounds, the estimated stopping probability for average number of ballots drawn for PROVIDENCE, MINERVA, EoR BRAVO, and SO BRAVO.

should be chosen to motivate that section.

6 Multiple round audits

Some election audits have benefited from a one-and-done approach: draw a large sample with high probability of stopping in the first round and usually avoid a second round altogether. This is appealing for two reasons. Firstly, rounds have some overhead in time and effort. Thus the workload (cost and time) of an audit grows not just with the number of ballots sampled but also with the number of rounds. On the other hand, a one-and-done audit often draws more ballots than are necessary. We propose a simple workload model under which our simulation results suggest that a few smaller rounds gives lower average workload.

We begin with the simple assumption that part of the workload grows linearly with the number of ballots, giving us the per ballot cost parameter c_b. It is less clear how cost is related to the number of rounds, but we propose a simple model in which the round cost is also linear, with constant per round cost c_r. This is a reasonable assumption for contests with small margin in which all or nearly all ballot boxes need to be opened in each round. [@Poorvi, I need help justifying the workload model.]

So for an audit $\mathcal{A}$ with expected number of ballots E_b and expected number of rounds E_r, we get that the cost C of the audit is

$$C(\mathcal{A}) = E_b * c_b + E_r * c_r.$$

Analytical approximation of the expected audit behavior (E_b and E_r) is challenging since the number of possible sequences of samples grows exponentially with the number of rounds. Therefore we again use the standard approach of simulations.

Consider the US state of Pennsylvania. In the 2020 US Presidential contest, Pennsylvania had a margin of 99999%.

We let the per ballot cost be one, $c_b = 1$. Then the per round cost c_r tells us the cost of a round as a number of ballots. We begin with an example in which $c_r = 1000$. That is, we suppose that the overhead of a round is equal to the workload of sampling 1000 ballots.

We now consider a simple round schedule for our example. Each round is selected to give the same probability of stopping, p. If the audit does not stop in the first round, we find a round size which, given the sample drawn in the first round, will give a probability of stopping p in the second round. For this round schedule scheme, a one-and-done audit is achieved by choosing large p, say $p = .9$ or $p = .95$.

The simulation results for our example are given in Figure **??**. For various round schedules, given by the varying parameter p on the horizontal axis, we see that the average audit cost varies with a minimum at roughly *.xx*(TODO). In particular, the one-and-done $p = .9$ audits have an average of *zz%* higher cost than the $p = .xx$ audits.

5

Without extensive workload measurements, it is hard to know the most accurate value for c_r, and so we also evaluate the simulation results with varying c_r.

7 Conclusion

8 Availability

PROVIDENCE is implemented in the R2B2 software library for R2 and B2 audits.

References

[1] Matthew Bernhard. *Election Security Is Harder Than You Think*. PhD thesis, University of Michigan, 2020.

[2] Michelle L. Blom, Peter J. Stuckey, and Vanessa J. Teague. Ballot-polling risk limiting audits for IRV elections. In Robert Krimmer, Melanie Volkamer, Véronique Cortier, Rajeev Goré, Manik Hapsara, Uwe Serdült, and David Duenas-Cid, editors, *Electronic Voting - Third International Joint Conference, E-Vote-ID 2018, Bregenz, Austria, October 2-5, 2018, Proceedings*, volume 11143 of *Lecture Notes in Computer Science*, pages 17–34. Springer, 2018.

[3] Oliver Broadrick, Sarah Morin, Grant McClearn, Neal McBurnett, Poorvi L. Vora, and Filip Zagórski. Simulations of ballot polling risk-limiting audits. In *Seventh Workshop on Advances in Secure Electronic Voting, in Association with Financial Crypto*, 2022.

[4] Common Cause, VerifiedVoting, and Brennan Center. Pilot implementation study of risk-limiting audit methods in the state of rhode island. https://www.brennancenter.org/sites/default/files/2019-09/Report-RI-Design-FINAL-WEB4.pdf.

[5] Zhuoqun Huang, Ronald L. Rivest, Philip B. Stark, Vanessa J. Teague, and Damjan Vukcevic. A unified evaluation of two-candidate ballot-polling election auditing methods. In Robert Krimmer, Melanie Volkamer, Bernhard Beckert, Ralf Küsters, Oksana Kulyk, David Duenas-Cid, and Mikhel Solvak, editors, *Electronic Voting - 5th International Joint Conference, E-Vote-ID 2020, Bregenz, Austria, October 6-9, 2020, Proceedings*, volume 12455 of *Lecture Notes in Computer Science*, pages 112–128. Springer, 2020.

[6] Mark Lindeman and Philip B Stark. A gentle introduction to risk-limiting audits. *IEEE Security & Privacy*, 10(5):42–49, 2012.

[7] Mark Lindeman, Philip B Stark, and Vincent S Yates. BRAVO: Ballot-polling risk-limiting audits to verify outcomes. In *EVT/WOTE*, 2012.

[8] Katherine McLaughlin and Philip B. Stark. Simulations of risk-limiting audit techniques and the effects of reducing batch size on the 2008 California House of Representatives elections. 2010.

[9] Katherine McLaughlin and Philip B. Stark. Workload estimates for risk-limiting audits of large contests. 2011.

[10] Sarah Morin and Grant McClearn. The R2B2 (Round-by-Round, Ballot-by-Ballot) library, https://github.com/gwexploratoryaudits/r2b2.

[11] Kellie Ottoboni, Matthew Bernhard, J. Alex Halderman, Ronald L Rivest, and Philip B. Stark. Bernoulli ballot polling: A manifest improvement for risk-limiting audits. *International Conference on Financial Cryptography and Data Security*, pages 226–241, 2019.

[12] Philip B. Stark. Simulating a ballot-polling audit with cards and dice. In *Multidisciplinary Conference on Election Auditing, MIT*, december 2018.

[13] Philip B. Stark and David A. Wagner. Evidence-based elections. *IEEE Secur. Priv.*, 10(5):33–41, 2012.

[14] VotingWorks. Arlo, https://voting.works/risk-limiting-audits/.

[15] Abraham Wald. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2):117–186, 1945.

[16] Filip Zagórski, Grant McClearn, Sarah Morin, Neal McBurnett, and Poorvi L. Vora. The Athena class of risk-limiting ballot polling audits. *CoRR*, abs/2008.02315, 2020.

[17] Filip Zagórski, Grant McClearn, Sarah Morin, Neal McBurnett, and Poorvi L. Vora. Minerva– an efficient risk-limiting ballot polling audit. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 3059–3076. USENIX Association, August 2021.

A Proofs

Lemma 1. *For* $0 < p_0 < p_a < 1$ *and* $n > 0$*, the ratio* $\sigma(k, p_a, p_0, n)$ *is strictly increasing as a function of k for* $0 \leq k \leq n$.

Proof. From $0 < p_0 < p_a < 1$, we get

$$\frac{p_a}{p_0} > 1$$

and

$$1 - p_0 > 1 - p_a \implies \frac{1 - p_0}{1 - p_a} > 1,$$

and thus

$$\frac{p_a(1-p_0)}{p_0(1-p_a)} > 1.$$

Now simply observe that

$$\frac{p_a(1-p_0)}{p_0(1-p_a)} \cdot \sigma(k,p_a,p_0,n) = \frac{p_a(1-p_0)}{p_0(1-p_a)} \cdot \frac{p_a^k(1-p_a)^{n-k}}{p_a^k(1-p_a)^{n-k}}$$

$$= \frac{p_a^{k+1}(1-p_a)^{n-(k+1)}}{p_a^{k+1}(1-p_a)^{n-(k+1)}} = \sigma(k+1,p_a,p_0,n).$$

$\square$

Lemma 2. *Given a monotone increasing sequence:* $\frac{a_1}{b_1}, \frac{a_2}{b_2}, \ldots, \frac{a_n}{b_n}$, *for* $a_i, b_i > 0$, *the sequence:*

$$z_i = \frac{\sum_{j=i}^n a_j}{\sum_{j=i}^n b_j}$$

is also monotone increasing.

Proof. Note that z_i is a weighted average of the values of $\frac{a_j}{b_j}$ for $j \geq i$:

$$z_i = \sum_{j=i}^n y_j \frac{a_j}{b_j}$$

for

$$y_j = \frac{b_j}{\sum_{j=i}^n b_j} > 0.$$

Further, $\sum_{j=i}^n y_j = 1$ and hence $y_j \leq 1$ and $y_j = 1 \Leftrightarrow i = j = n$. Observe that, because $\frac{a_i}{b_i}$ is monotone increasing, $z_i \geq \frac{a_i}{b_i}$ with equality if and only if $i = n$. Suppose $i < n$. Then

$$z_{i+1} \geq \frac{a_{i+1}}{b_{i+1}} > \frac{a_i}{b_i},$$

and

$$z_i = y_i \frac{a_i}{b_i} + (1-y_i)z_{i+1} < z_{i+1}.$$

Thus z_i is also monotone increasing. $\square$

Lemma 3. *For* $0 < p_0 < p_a < 1$ *and* $n > 0$, *the ratio* $\tau_1(k,p_a,p_0,n)$ *is strictly increasing as a function* k *for* $0 \leq k \leq n$.

Proof. Apply Lemmas 1-2. $\square$

Lemma 4. *Given a strictly monotone increasing sequence:* $x_1, x_2, \ldots x_n$ *and some constant A,*

$$A \leq x_i \Leftrightarrow \exists i_{min} \leq i \ s.t. \ x_{i_{min}-1} < A \leq x_{i_{min}} \leq x_i,$$

unless $A \leq x_1$, *in which case* $i_{min} = 1$.

Proof. Evident. $\square$

Lemma 5. *For* $\mathcal{A} = (\alpha, p_a, p_0, k_{j-1}, n_{j-1}, n_j)$-PROVIDENCE, *there exists a* $k_{min,j}(\text{PROVIDENCE}, p_a, p_0, k_{j-1}, n_{j-1}, n_j)$ *such that*

$$\mathcal{A}(X_j) = \text{Correct} \iff k_j \geq k_{min,j}(\text{PROVIDENCE}, \boldsymbol{n_j}, p_a, p_0).$$

Proof. From Definition 4,

$$\mathcal{A}(X_j) = \text{Correct} \iff \omega_j(k_j, k_{j-1}, p_a, p_0, n_j, n_{j-1}) \geq \frac{1}{\alpha}.$$

Now to apply Lemma 4, it suffices to show that ω_j is monotone increasing with respect to k_j. For $j = 1$, we have $\omega_1 = \tau_1$, so ω_1 is strictly increasing by Lemma 3. For $j \geq 2$,

$$\omega_j(k_j, k_{j-1}, p_a, p_0, n_j, n_{j-1}, \alpha) =$$

$$\sigma(k_{j-1}, p_a, p_0, n_{j-1}) \cdot \tau_1(k_j - k_{j-1}, p_a, p_0, n_j - n_{j-1}).$$

As a function of k_j, σ is constant, and thus ω is strictly increasing by Lemma 3. Therefore by Lemma 4, we have the desired property. $\square$

Lemma 6. *For* $j \geq 1$,

$$\frac{Pr[\boldsymbol{K_j} = \boldsymbol{k_j} \mid \boldsymbol{n_j}, H_a]}{Pr[\boldsymbol{K_j} = \boldsymbol{k_j} \mid \boldsymbol{n_j}, H_0]} = \sigma(k_j, p_a, p_0, n_j).$$

Proof. We induct on the number of rounds. For $j = 1$, we have

$$\frac{Pr[\boldsymbol{K_1} = \boldsymbol{k_1} \mid \boldsymbol{n_1}, H_a]}{Pr[\boldsymbol{K_1} = \boldsymbol{k_1} \mid \boldsymbol{n_1}, H_0]} = \frac{Pr[K_1 = k_1 \mid n_1, H_a]}{Pr[K_1 = k_1 \mid n_1, H_0]}$$

$$= \frac{\text{Bin}(k_1, n_1, p_a)}{\text{Bin}(k_1, n_1, p_0)} = \sigma(k_1, p_a, p_0, n_1).$$

Suppose the lemma is true for round $j = m$ with history $\boldsymbol{k_m}$. Observe that

$$\frac{Pr[\boldsymbol{K_{m+1}} = \boldsymbol{k_{m+1}} \mid \boldsymbol{n_{m+1}}, H_a]}{Pr[\boldsymbol{K_{m+1}} = \boldsymbol{k_{m+1}} \mid \boldsymbol{n_{m+1}}, H_0]}$$

$$= \frac{Pr[\boldsymbol{K_m} = \boldsymbol{k_m} \mid \boldsymbol{n_{m+1}}, H_a] \cdot Pr[K'_{m+1} = k'_{m+1} \mid \boldsymbol{k_m}, \boldsymbol{n_{m+1}}, H_a]}{Pr[\boldsymbol{K_m} = \boldsymbol{k_m} \mid \boldsymbol{n_{m+1}}, H_0] \cdot Pr[K'_{m+1} = k'_{m+1} \mid \boldsymbol{k_m}, \boldsymbol{n_{m+1}}, H_0]}$$

$$= \sigma(k_m, p_a, p_0, n_m) \cdot \frac{Pr[K'_{m+1} = k'_{m+1} \mid \boldsymbol{k_m}, \boldsymbol{n_{m+1}}, H_a]}{Pr[K'_{m+1} = k'_{m+1} \mid \boldsymbol{k_m}, \boldsymbol{n_{m+1}}, H_0]}$$

by the induction hypothesis. Then this is simply equal to

$$\sigma(k_m, p_a, p_0, n_m) \cdot \frac{\text{Bin}(k'_{m+1}, n'_{m+1}, p_a)}{\text{Bin}(k'_{m+1}, n'_{m+1}, p_0)}$$

$$= \frac{p_a^{k_m}(1-p_a)^{n_m-k_m}}{p_0^{k_m}(1-p_0)^{n_m-k_m}} \cdot \frac{p_a^{k'_{m+1}}(1-p_a)^{n'_{m+1}-k'_{m+1}}}{p_0^{k'_{m+1}}(1-p_0)^{n'_{m+1}-k'_{m+1}}}$$

$$= \sigma(k_{m+1}, p_a, p_0, n_{m+1})$$

$\square$