A Note on Risk-Limiting Bayesian Polling Audits for Two-Candidate Elections

Sarah Morin[1], Grant McClearn[1], Neal McBurnett[1], Poorvi L. Vora[1(✉)], and Filip Zagórski[2]

[1] Department of Computer Science, The George Washington University, Washington, DC, USA
poorvi@gwu.edu

[2] Department of Computer Science, Wroclaw University of Science and Technology, Wroclaw, Poland

Abstract. This short paper provides a general form for a polling audit that is both Bayesian and risk-limiting: the *Bayesian Risk-Limiting (Polling) Audit*, which enables the use of a Bayesian approach to explore more efficient Risk-Limiting Audits. A numerical example illustrates the implications to practice.

Keywords: Election audits · Election security · Risk-limiting audits · Bayesian audits

1 Introduction

The framework of *risk-limiting audits (RLAs)*, as described by Lindeman and Stark [1], formalizes a rigorous approach to election verification. The purpose of an audit is to require a full hand count if the outcome is wrong; the *risk* is the rate at which it fails to do so, and depends on the (unknown) underlying true election tally. An *RLA* is an audit that guarantees that the worst-case risk—the largest value of the risk over all possible true election tallies—is smaller than a pre-specified bound.

The Bayesian audit, as described by Rivest and Shen [5], begins with an assumed prior probability distribution over the election tally. It guarantees a pre-specified upper bound on the *upset probability*, which is the weighted average of risk values, each risk value corresponding to an election tally inconsistent with the announced outcome and weighted by the corresponding prior probability. As an average of risks, the upset probability could be considerably smaller than the worst-case risk; limiting it does not, in general, limit the worst-case risk. The Bayesian framework is promising as a means of designing efficient audits

This material is based upon work supported in part by NSF Award CNS 1421373. Author was partially supported by Polish National Science Centre contract number DEC-2013/09/D/ST6/03927 and by Wroclaw University of Science and Technology [0401/0052/18].

M. Bernhard et al. (Eds.): FC 2020 Workshops, LNCS 12063, pp. 337–344, 2020.
https://doi.org/10.1007/978-3-030-54455-3_24

(requiring a small average sample size, we make this more precise later), and an important question is whether we can bound the worst-case (maximum) risk of a Bayesian audit given the upper bound on its upset probability.

The BRAVO audit [1] can be reduced to a comparison test as described in the classical work of Wald [7]. The CLIP audit of Rivest [4] is another *RLA* which may also be reduced to such a comparison, though the values are computed using simulations[1]. Before this work, it was not known if Bayesian audits could be reduced to comparison tests; they are generally computed using Pólya urn simulations.

In this short paper, while restricting ourselves to polling audits of two-candidate plurality elections with no invalid ballots, we state the following results without proof:

1. We define a class of Bayesian audits that are most efficient *RLA*s. Most efficient *RLA*s are those that use the smallest expected number of ballots given either hypothesis: a correct election outcome or an incorrect one, if the election is drawn from the assumed prior. The expectation is computed over the randomness of the tally and the sampling process. We describe how the *BRAVO* audit may be viewed as a special case of a generalized Bayesian *RLA*, based on a more general version of the Bayesian audit defined by Rivest and Shen.
2. The Bayesian audit can be reduced to a simple comparison test between the number of votes for the winner in the audit sample and two pre-computed values for this sample size (we denote this size n):
   - a minimum number of votes for the winner, $k_{min}(n)$, above which the election outcome is declared correct, and
   - a maximum number of votes for the winner, $k_{max}(n)$, below which the audit proceeds to a hand count.

We present an illustrative example of $k_{min}(n)$ values computed for various audits. Proofs of results 1 and 2 above and more illustrative examples may be found in [6].

1.1 Organization

This paper is organized as follows. Section 2 describes the model and establishes most of the notation. Section 3 describes *RLA*s [1] and Bayesian audits [5]. Our contributions are to be found in Sect. 4, which states our theoretical results, and Sect. 5, which presents the illustrative example. Section 6 concludes.

2 The Model

We consider a plurality election with two candidates, N voters and no invalid ballots. Once the votes are cast, either the true outcome of the election is a tie, or

[1] Philip Stark has mentioned a CLIP-like audit which does not use simulations as work in progress.

there is a well-defined true winner. In the worst case, however, unless all votes are manually counted, the true winner is generally unknown. In the Bayesian approach, the true winner is modeled as a random variable, which we denote W. We further denote by w an instance of W, by w_a and ℓ_a the announced winner and loser respectively and by x the (true, unknown) number of votes obtained by w_a. Thus $w_a = w$ if and only if $x > \frac{N}{2}$.

A polling audit will estimate whether w_a is the true winner. Consider a sample of n votes drawn uniformly at random: v_1, v_2, ..., v_n, $n < N$, $v_i \in \{w_a, \ell_a\}$. The sample forms the *signal* or the *observation*; the corresponding random variable is denoted $\mathbf{S}_n \in \{w_a, \ell_a\}^n$, the specific value $\mathbf{s}_n = [v_1, v_2, ..., v_n]$. Let k_n denote the number of votes for w_a in the sample; then $n - k_n$ votes are for ℓ_a.

The audit computes a binary-valued estimate of the true winner from $\mathbf{s}_n$:

$$\hat{w}_n : \{w_a, \ell_a\}^n \rightarrow \{w_a, \ell_a\}$$

We will refer to the function $\hat{w}_n$ as the *estimator* and $\hat{w}_n(\mathbf{s}_n)$ as the *estimate*. The audit uses an error measure to compute the quality of the estimate.

- If $\hat{w}_n(\mathbf{s}_n) = w_a$ and the error measure is acceptable we are done (the audit stops) and declare that the election outcome was correctly announced.
- If $\hat{w}_n(\mathbf{s}_n) = \ell_a$ and the error measure is acceptable we stop drawing votes and proceed to perform a complete hand count.
- If the error measure is not acceptable we draw more votes to improve the estimate.

Thus, when we use the term *the audit stops*, we mean that the audit verified the election outcome. When we say the audit proceeds to a *hand count*, we mean that the tentative estimate is $\hat{w}_n(\mathbf{s}_n) = \ell_a$, we stop drawing samples and proceed to a full hand count.

In computing the audit we can make two types of errors:

1. *Miss:* A *miss* occurs when the announced outcome is incorrect, $w \neq w_a$, but the estimator misses this, $\hat{w}_n(\mathbf{s}_n) = w_a$, the error measure is small enough and the audit stops. We denote by P_M the probability of a miss—given that the announced outcome is incorrect, the probability that the audit will miss this:
$$P_M = Pr[audit\ stops\ \ ||\ w \neq w_a]$$
P_M is the *risk* in risk limiting audits. If the audit is viewed as a statistical test, with the null hypothesis being $w = \ell_a$, when it stops, P_M is the Type I error.
2. *Unnecessary Hand Count:* Similarly, if $w = w_a$, but $\hat{w}_n(\mathbf{s}_n) = \ell_a$, acceptance of the estimate would lead to an unnecessary hand count. We denote the probability of an *unnecessary hand count* by P_U:
$$P_U = Pr[hand\ count\ \ ||\ w = w_a]$$
If the audit is viewed as a statistical test, with the null hypothesis being $w = \ell_a$, when the audit stops, P_U is the Type II error.

3 Defining the Audit

In this section, we describe two types of audits. We do not attempt to introduce any new ideas, but try to faithfully represent the existing literature.

3.1 Risk-Limiting Audits (RLAs) [1]

A *risk-limiting audit (RLA)* with *risk limit* α—as described by, for example, Lindeman and Stark [1]—is one for which the risk is smaller than α for all possible (unknown) true tallies in the election (or—equivalently for the two-candidate election—all possible values of x). For convenience when we compare audits, we refer to this audit as an α-*RLA*.

An example of an α-*RLA* for a two-candidate election with no invalid ballots and where ballots are drawn with replacement is the following, which is an instance of Wald's Sequential Probability Ratio Test (*SPRT*):

$$\hat{w}_n = \begin{cases} w_a & \frac{p^{k_n}(1-p)^{n-k_n}}{(\frac{1}{2})^n} > \frac{1-\beta}{\alpha} \\ \ell_a & \frac{p^{k_n}(1-p)^{n-k_n}}{(\frac{1}{2})^n} < \frac{\beta}{1-\alpha} \\ \textit{undetermined (draw more samples)} & \textit{else} \end{cases} \quad (1)$$

We denote the above as the (α, β, p)-*SPRT RLA*. Note that a similar expression may be obtained for sampling without replacement, see, for example, [3].

Proposition:
When the only possible values of the true vote count, x, are pN when w_a wins, and $\frac{N}{2}$ otherwise, the (α, β, p) SPRT RLA *has $P_M < \alpha$ and $P_U < \beta$, and is a most efficient test achieving these bounds.*

Proof: This follows from Wald's argument [7].

Note that, in this case, there is no explicitly-assumed prior over the election tally; hence the term "most efficient test" here means one that requires the smallest expected number of ballots given either hypothesis: a correct election outcome or an incorrect one, if the tallies are pN when w_a wins and $\frac{N}{2}$ otherwise. The expectation is computed over the sampling process. While the test we denote the (α, β, p)-*SPRT RLA* is believed to be an *RLA*, we are not aware of this having been proven in the literature.

Lindeman and Stark recommend the use of the (α, β, p)-*SPRT RLA* with $p = s - t$ where s is the fractional vote count announced for the winner and t is a tolerance used to improve the performance of the audit when the vote tallies are not accurate, but the announced outcome is correct.

The *BRAVO* audit as described in [2] is the $(\alpha, 0, p)$-*SPRT RLA* which we denote the (α, p)-*BRAVO* audit. Note that $\beta = 0$, and p can be modified to be slightly smaller than the announced fractional vote count as described in [1].

Other *RLA*s include the CLIP audit [4] which may be expressed as a simple comparison test between the number of votes for the winner and a pre-computed value that depends on sample size.

3.2 Bayesian Audits [5]

Bayesian audits, defined by Rivest and Shen [5], assume knowledge of a *prior* probability distribution on x; we denote this distribution by f_X. Given the sample $\mathbf{s}_n$, W inherits a *posterior* distribution, $Pr[W \mid \mathbf{S}_n = \mathbf{s}_n]$, also known as the *a posteriori* probability of W. The Bayesian audit estimates the winning candidate that maximizes this probability (that is, the candidate for whom this value is largest), with the constraint that the probability of estimation error is smaller than γ, a pre-determined quantity, $0 < \gamma < \frac{1}{2}$. The election outcome is correct if the estimated winning candidate is w_a and the error smaller than γ. In this case, the estimation error is also termed the *upset probability*.

The *(computational) Bayesian Audit* assumes the audit draws votes without replacement and uses knowledge of f_X to simulate the distribution on the unexamined votes, conditional on $\mathbf{s}_n$, using Pólya urns. The estimated candidate is the one with the largest number of wins in the simulations, provided the fraction of wins is greater than $1 - \gamma$.

We study the general Bayesian audit and do not restrict ourselves to Pólya urn simulations or drawing samples without replacement. We will refer to the general Bayesian audit as the (γ, f_X)-Bayesian audit and specify whether ballots are drawn with or without replacement. Additionally, we assume that $Pr[w = w_a] = Pr[w = \ell_a]$ and denote the probability of error by γ.

4 Our Main Results

In this section we state our main results without proofs. For proofs, see [6].

4.1 Is BRAVO a Bayesian Audit?

Corollary 1. *The* (γ, γ, p)-SPRT RLA *with/without replacement is the* (γ, f_X)-*Bayesian audit with/without replacement for*

$$f_X = \frac{1}{2}\delta_{x,\frac{N}{2}} + \frac{1}{2}\delta_{x,pN}$$

Note that the (α, p)-*BRAVO* audit may not be represented as a special case of the above because the Bayesian audit as defined by Rivest and Shen requires $\alpha = \beta$. However, a more general definition of the Bayesian audit, where the probability of erring when the outcome is correct is zero and not equal to the probability of erring when the outcome is wrong, would correspond to the *BRAVO* audit for f_X as above.

4.2 The Bayesian Audit Is a Comparison Test

We observe (without proof here) that the decision rule for the Bayesian audit is a simple comparison test. In fact, we observe that the *SPRT RLA* and Bayesian audits may be defined in the form:

$$\hat{w}_n(\mathbf{s}_n) = \begin{cases} w_a & k_n \geq k_{min}(n) \\ \ell_a & k_n \leq k_{max}(n) \\ \textit{undetermined} \\ (\textit{draw more samples}) & \textit{else} \end{cases} \tag{2}$$

where $k_{min}(n)$ and $k_{max}(n)$ are determined by the specific audit. This follows from the fact that the likelihood ratio is monotone increasing with k_n for a fixed n.

4.3 Bayesian RLAs

Given a prior f_X of the vote count for election E, define the *risk-maximizing distribution corresponding to* f_X (denoted f_X^*) as follows.

$$f_X^* = \begin{cases} f_X(x) & x > \frac{N}{2} \\ \frac{1}{2} & x = \frac{N}{2} \\ 0 & \textit{else} \end{cases} \tag{3}$$

Note that f_X^* is a valid distribution for the vote count of an election.

Theorem 1. *The* (α, f_X^*)*-Bayesian Audit is an* α-RLA *with* $P_U < \alpha$ *for election* E *with prior* f_X *and is a most efficient audit achieving* $P_M < \alpha$ *and* $P_U < \alpha$ *for the prior* f_X^**.*

The above may be used to show that the (α, α, p)-*SPRT RLA* is an *RLA*. Additionally, a similar approach may be used to show that the (α, β, p)-*SPRT RLA* is an *RLA*. We are not aware of a proof of this in the literature on election audits. Note that, as mentioned in Sect. 1, most efficient *RLA*s are those that use the smallest expected number of ballots given either hypothesis: a correct election outcome or an incorrect one, if the election is drawn from the assumed prior. The expectation is computed over the randomness of the tally and the sampling process.

5 An Illustrative Example

We computed values of $k_{min}(n)$ for an election with $N = 100$ ballots cast, two candidates, no invalid ballots, $\alpha = 0.001$ and audit sample sizes (i.e. values of n) from 9–75.

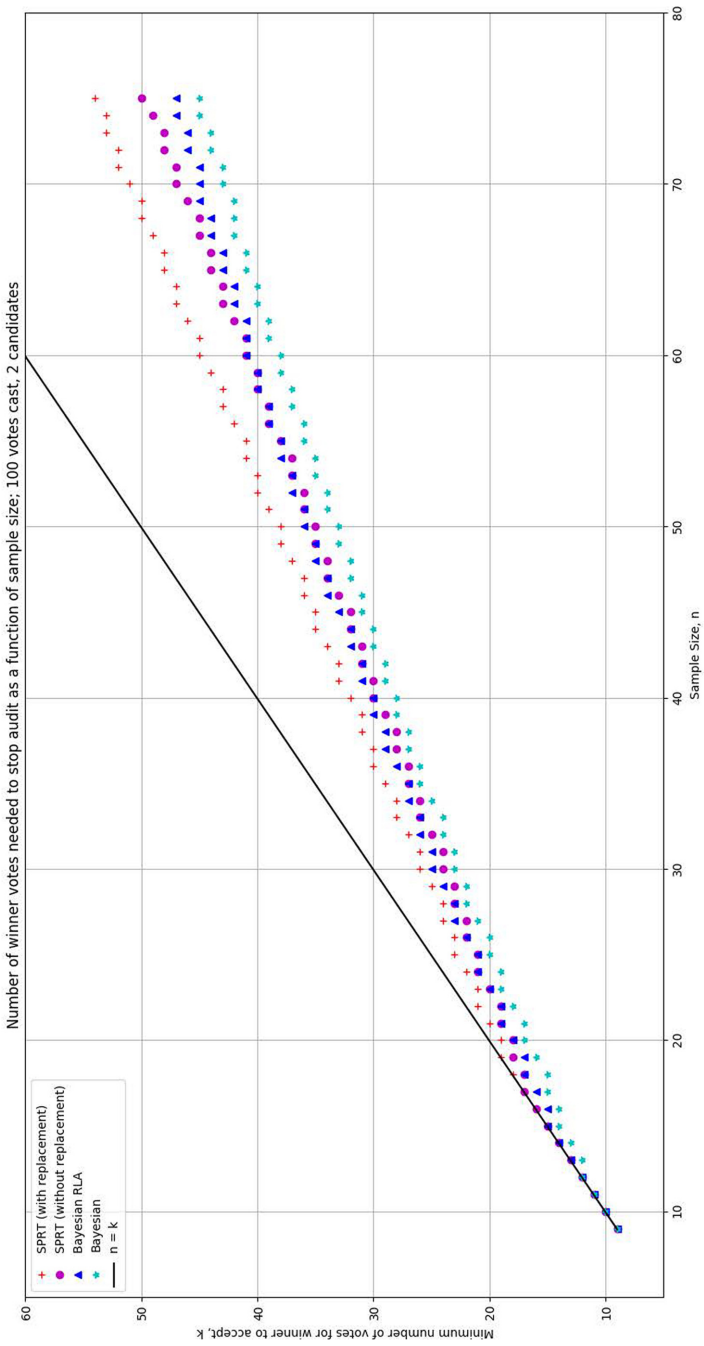

Fig. 1. Minimum number of winner votes as a function of sample size

We compared the following audits:

1. *SPRT RLA* with replacement, $p = 0.75$. That is, if the declared winner has won the election, we assume it is with a fractional vote count of 0.75.
2. *SPRT RLA* without replacement, $p = 0.75$.
3. Bayesian *RLA* corresponding to the uniform distribution. That is, the prior is uniform over all winning tallies, and the only possibility for $w \neq w_a$ is a fractional vote of 0.5 (a tie), with probability 0.5. The fractional vote of 0.75 in the *SPRT RLA* was chosen because the center of mass of the Bayesian prior when $w = w_a$ is a fractional vote of 0.75.
4. The Bayesian audit corresponding to the uniform distribution.

Figure 1 plots the values of $k_{min}(n)$ for samples sizes from 9 through 75. Note that each of (2) and (3) is the most efficient audit for its prior (when viewed as a Bayesian audit), so not much can be made of the number of samples needed. Note further that (1) is an audit with replacement and hence expected to require more samples than (2), which has the same assumed prior. Finally, note that (4) requires the fewest samples as expected because its upset probability is α, and it is not risk-limited. That is, its error bound is an average error bound, and not a worst-case one.

6 Conclusions and Future Work

We describe a risk-limiting Bayesian polling audit for two-candidate elections and describe how a Bayesian polling audit for two-candidate elections is a simple comparison test between the number of votes for the announced winner in a sample and a pre-computed value for that sample size. Open questions include the application of this model to more complex elections.

References

1. Lindeman, M., Stark, P.B.: A gentle introduction to risk-limiting audits. IEEE Secur. Priv. **10**(5), 42–49 (2012)
2. Lindeman, M., Stark, P.B., Yates, V.S.: BRAVO: ballot-polling risk-limiting audits to verify outcomes. In: EVT/WOTE (2012)
3. Ottoboni, K., Stark, P.B., Lindeman, M., McBurnett, N.: Risk-limiting audits by stratified union-intersection tests of elections (SUITE). In: Krimmer, R., et al. (eds.) E-Vote-ID 2018. LNCS, vol. 11143, pp. 174–188. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00419-4_12
4. Rivest, R.L.: ClipAudit–a simple post-election risk-limiting audit. arXiv:1701.08312
5. Rivest, R.L., Shen, E.: A Bayesian method for auditing elections. In: EVT/WOTE (2012)
6. Vora, P.L.: Risk-limiting Bayesian polling audits for two candidate elections. CoRR abs/1902.00999 (2019). http://arxiv.org/abs/1902.00999
7. Wald, A.: Sequential tests of statistical hypotheses. Ann. Math. Stat. **16**(2), 117–186 (1945)