Journal of Discrete Mathematics and Its Applications 10 (2) (2025) 207–221

JDMA

Journal of Discrete Mathematics and Its Applications

Available Online at: http://jdma.sru.ac.ir

Shahid Rajaee Teacher Training University

Research Paper

Risk-limiting audits for score voting systems

Hamid Devisti, Massoud Hadian*

School of Mathematics and Computer Science, Iran University of Science and Technology, Tehran, I. R. Iran.

Academic Editor: Vahid Mohammadi

Abstract. Several voting systems are utilized to allocate both political and non-political positions across countries worldwide. Plurality voting and approval voting are among the most widely implemented electoral systems. Established statistical methods are employed to ensure the accuracy of vote counting and the validation of election results. Two fundamental approaches that significantly enhance the likelihood of identifying errors in election outcomes are Risk-limiting audit (RLA) and Bayesian audit (BA). These audit methods assess the security of elections using statistical tools, based on the random selection of cast votes and their interpretation as evidence supporting or contradicting the reported results. In this paper, we first examine the advantages of a specific form of approval voting, referred to as score voting, and then describe two types of risk-limiting audits to evaluate the accuracy of vote counting and the results. The proposed auditing method for score voting is adapted from the Ballot-polling risk-limiting audits to verify outcomes (BRAVO). Simulation results confirm the effectiveness and accuracy of our approach.

Keywords. BRAVO method, Fair voting, SPRT method.
Mathematics Subject Classification (2020): $62L10$, $62H15$.

1 Introduction

There are two primary objectives in any voting procedure: first, to ensure the selection of the most desirable option among the candidates; and second, to ensure that the election

*Corresponding author (Email address: mhadian@iust.ac.ir).
Received 11 December 2024; Revised 06 January 2025; Accepted 08 March 2025
First Publish Date: 01 June 2025

 ©Shahid Rajaee Teacher Training University

results are accurate and reliable. The focus of this paper is primarily on the second objective, although a brief discussion of the first will also be provided.

A critical examination of traditional voting methods reveals several inherent drawbacks and limitations. Many voting systems emphasize the consensus of a majority of voters in selecting one or more candidates. This majority-focused approach, though widespread, does not account for situations where voters may select candidates not because they are the most preferred, but to eliminate other candidates. This phenomenon, where voters choose to avoid an unfavorable option, highlights the complexity of voter preferences. Some voting systems, such as plurality voting and majority voting, which are based on majority rule, can lead to the tyranny of the majority or plurality rule. The tyranny of the majority refers to a scenario in which the majority imposes its will on minority factions, disregarding their interests. In countries with a presidential system, where the head of government is elected directly by the people, majority rule can result in an autocratic or dictatorial system, particularly when there are no term limits. For instance, in countries such as North Korea, Egypt, and Syria, political power is often concentrated in the hands of a minority group, even though elections are held. Prior to the Twenty-second amendment to the U.S. constitution, there were no presidential term limits, allowing a single individual to hold office for an indefinite period. After the amendment passed in 1947, a president can serve no more than two terms (a maximum of eight years), partly to prevent the tyranny of the majority. Similarly, in parliamentary systems, individuals with a consistent base of support can retain a seat for extended periods, as there are often no term limits on parliamentary elections.

To address the limitations of plurality voting, approval voting has been proposed. Approval voting, first formally introduced in [18], and later revised and published in [3], allows voters to approve multiple candidates. However, approval voting has its own shortcomings, the most notable being that it does not account for the intensity of voters preferences. Voters may approve multiple candidates, but the system does not differentiate the level of support for each candidate. Since individuals' opinions about candidates are not uniform or without intensity, this lack of differentiation can cause problems, particularly when political parties attempt to convince voters to approve all candidates on a given list. This limits the individual's decision-making power when they must decide how strongly they support each candidate. A potential improvement is ranked voting, which allows voters to rank candidates in order of preference. However, ranked voting methods vary in how winners are determined, and some of these methods do not result in a single, clear winner. Additionally, ranked voting does not provide a measure of preference intensity. The evaluation in ranked voting is based on ordinal ranking, where the difference between ranks is not specified. The most optimal system, which avoids the limitations of previous methods, is score voting.

Score voting aggregates voters' preferences in a manner that reflects the intensity of those preferences. By using a scoring system, it becomes possible to more accurately measure the differences between candidates. In score voting, instead of relying on methods like ranked voting, the winner(s) is determined by the sum of the scores assigned by voters.

While scoring systems provide a more nuanced understanding of voter preferences, an important aspect of election security is ensuring the accuracy and reliability of vote counting. The most widely used methods for verifying vote counts are RLAs and BAs. The RLA is an efficient, conditional algorithm that assesses the accuracy of election results through statistical hypothesis testing. The null hypothesis posits that the reported winner did not actually win, while the alternative hypothesis suggests that the reported winner is the true victor. To test this, a sequential sampling process is employed, with the sample size not predetermined but bounded by an upper limit. Sampling continues until the upper limit is reached, or the evidence supports the alternative hypothesis, at which point the results are confirmed. If the alternative hypothesis is not supported, a full manual recount is conducted. Random sampling can be done with or without replacement.

The Bayesian audit, similar to the RLA, also uses random sampling but incorporates prior probabilities of vote proportions for the winner and loser. The evidence in the sample is used to update the prior distribution of the vote proportion in favor of the reported winner. If the sampling and counting conditions are met, both RLA and BA can yield identical results. Bayesian inference methods are commonly applied in election polling and auditing, as discussed in [1,7,11,12].

This paper introduces a ballot polling risk-limiting audit (RLA) adapted from the BRAVO method [10]. The method is designed for k-winner contests and utilizes appropriate statistical tests to audit election results. The paper is organized as follows: Section 2 reviews related work; Section 3 discusses the advantages of score voting; Section 4 explains the context and notation used; Section 5 presents the proposed audit methods; and Section 6 provides a comparison of results and simulation outcomes.

2 Related works

The application of statistical methods in audit processes was first introduced in [9]. The error-count method presented in this paper reduces the sample size required for audits, while still producing results comparable to a statistical recount conducted with a larger sample. In [14], post-election audits based on statistical testing were introduced. The approach proposed by Stark utilizes a P-value for hypothesis testing, where the decision to either perform a full manual recount or continue sampling is based on comparing the P-value with a predefined test statistic value. In [10], a flexible protocol for auditing election results, called BRAVO, was introduced. This protocol also includes a table for estimating the average sample size required to accept or reject the null hypothesis. Many subsequent audit methods have been inspired by the BRAVO framework.

Sarwate et al briefly discussed the use of scored systems in their work [13]. In their method, they consider the difference between the real value and the reported value of each ballot, which they refer to as the ballot error. Using these errors, they construct a test statistic. Rather than relying on the fraction of ballots cast, they use the election margin—the minimum level of error necessary to alter the election outcome. In [16], risk-limiting audits were proposed for proportional representation election systems, where each voter selects a party, and seats are allocated to parties based on their proportion of the total votes.

In [11], A simple risk-limiting post-election audit (CLIP AUDIT) method was introduced, which is based on the difference in the number of ballots for reported winners and losers in the sample. This method is claimed to be independent of the unofficial margin, in contrast to the BRAVO method. [2] developed a risk-limiting audit for Instant Runoff Voting (IRV) based on the BRAVO framework. In this approach, IRV elections are treated as several simultaneous First-Past-The-Post (FPTP) elections. In each round of the audit, one candidate is eliminated, and the audit continues until a winner is determined.

In [15], the sets of half-average nulls generate risk-limiting audit (SHANGRLA) was introduced. The innovation of this method lies in the construction of sets of assertions for contests. Each null hypothesis is formalized as a statement such as *the average value of assorter functions is not greater than 1/2* for a collection of finite lists of nonnegative numbers.

Finally, [7] evaluated several audit methods using simulation. The study compared different methods based on the quantities that need to be specified and their ability to automatically limit risk. According to the findings, the BRAVO method automatically controls risk, though the proportion of winner votes (as a required quantity) must still be predefined.

3 Why the score voting?

The main question is: What are the advantages of score voting? This section seeks to address this question. As mentioned in the introduction, political parties sometimes encourage voters to support an entire list of candidates (i.e., all candidates on the list). This method of voting is often irrational and can lead to several issues. For instance, pairing well-known candidates with lesser-known or less-qualified candidates can increase the likelihood of undeserving individuals being elected to positions of political power. This results in potential abuse of power by the parties. A notable example is the 2016 parliamentary elections in Iran, where one political party won all 30 seats in Tehran by promoting a single list of 30 candidates. The party leaders encouraged voters to select all candidates on the list, leading to the party's sweeping victory. However, the poor performance of these representatives prompted significant criticism of this method. In contrast, the score voting system allows voters to assign a proportional score to each candidate based on their preferences. Higher scores increase a candidate's chance of winning, while lower scores decrease it. Therefore, candidates with the highest scores emerge as the winners.

There are several criteria for evaluating electoral systems, which assess their strengths and weaknesses from various perspectives (see [6]). These criteria, defined as follows, help compare different systems. No voting method satisfies all of these criteria, and there is no ideal system. Score voting is compared with four electoral systems based on primary election criteria in Table 1.

- **Monotonicity Criterion**: A ranked voting system is monotonic if it is impossible to prevent the election of a candidate by ranking them higher on some ballots, nor possible to elect an otherwise unelected candidate by ranking them lower on some ballots, while all other ballots remain unchanged.

- **Participation Criterion**: Voting systems that do not satisfy the participation criterion are described as exhibiting the "no-show paradox," enabling a specific type of tactical voting. In such cases, a voter may increase their preferred candidate's chances of winning by choosing not to vote. The participation criterion in definitive and probabilistic framework is defined as follows:
  - Definitive framework: According to the participation criterion, introducing a ballot that ranks candidate A strictly above candidate B should not result in candidate B winning instead of candidate A.
  - Probabilistic framework: The participation criterion states that adding a ballot where candidate A is strictly preferred to candidate B should not change the winner from candidate A to candidate B.
- **Consistency Criterion**: A voting system satisfies the consistency criterion if, when the electorate is divided arbitrarily into parts and the same result is achieved in each part, the election of the entire electorate yields the same result.
- **Condorcet Criterion**: An electoral system satisfies the Condorcet criterion (or Condorcet winner criterion) if it always selects the Condorcet winner[a] whenever one exists.
- **Resolvability Criterion**: A voting system is said to be resolvable if it has a low likelihood of tie votes. Systems that satisfy the resolvability criterion ensure the uniqueness of the winner.

Table 1. *Yes** : Voting system satisfy criterion in special conditions

Voting system	Criterion			
	Condorcet	Resolvability	Monontonicity	Participation
Approval voting	*Yes*	*Yes**	*Yes**	*Yes**
Instant-runoff voting	*Yes*	*No*	*Yes**	*No*
Plurality voting	*Yes*	*Yes**	*Yes*	*No*
Score-voting	*Yes*	*Yes*	*Yes*	*No*
Two-round voting	*Yes**	*No*	*No*	*Yes*

[a] A Condorcet method is an election method that elects the candidate who wins a majority of votes in every head-to-head election against each of the other candidates, i.e., a candidate preferred by more voters than any other candidate, when such a candidate exists.

4 Context and notation

A variety of score voting systems exist, which are similar to one another to varying degrees. In this discussion, we focus on the most common and straightforward version of score voting. In this system, there is a set C consisting of C candidates and a set $\mathcal{B}$ of N ballots. The set of candidates, $\mathcal{C}$, is divided into two subsets: $\mathcal{W}$, which contains k winners, and $\mathcal{L}$, which contains $C - k$ losers.

Each ballot consists of a list of C candidates, with scores assigned to each candidate. To facilitate the voters and the tallying of scores, the ballots are often presented in a matrix format. The names of the candidates are listed in the first column, and the scores are indicated in the first row. An example of such a ballot is shown in the table below.

Candidate/Score	1	2	3	4
A	□	□	□	■
B	□	□	□	■
C	□	□	□	■
D	□	□	□	■

To implement score voting in an actual election, ballot papers are designed similarly to multiple-choice answer sheets, and the tallying and summing of scores are conducted using a scantron machine. This method helps to reduce errors associated with manual counting. The objective of score voting is to select the winner(s) by computing the total score of each candidate. The candidates with the highest total scores are declared the winners.

In score voting, each voter assigns a score to each candidate based on their preference intensity, with scores ranging from 1 to C. There are two fundamental assumptions for each ballot. First, no two distinct candidates receive the same score. Second, each candidate must receive a score. Thus, each ballot can be viewed as a permutation function with a constant summation, as follows:

$$SV : \{1,2,\ldots,C\} \xrightarrow{(1,2,\ldots,C)\mapsto(i_1,i_2,\ldots,i_C)} \{1,2,\ldots,C\},$$

where the total score of each ballot is given by

$$\sum_{j=1}^{C} i_j = \frac{C(C+1)}{2}.$$

The total score for the entire contest is

$$\frac{NC(C+1)}{2},$$

and the winners (in the case of a multiple-winner election) are the candidates who accumulate the highest total scores. The total score for each candidate is the sum of the scores from all ballots. Let the total score of candidate c_i be denoted by S_i for each $i \in \{1,2,\ldots,C\}$, and let S_{ij} represent the sum of scores of candidate c_i in the ballots where candidate c_i received a score of j for every $i, j \in \{1, 2, \ldots, C\}$.

We have the relationship:

$$S_i = \sum_{j=1}^{C} S_{ij}.$$

where S_i and S_{ij} refer to the total and partial scores, respectively. The true values of S_i and S_{ij} are unknown; however, their reported values are provided, denoted as $\hat{S}_i$ and $\hat{S}_{ij}$, respectively. To present the final election and audit results, we use a tabular format. In the final table, the candidates' names are listed in descending order of their scores in the first column, the partial scores are displayed in the first row, and the total score for each candidate is shown in the last column. For every $i, j \in \{1, 2, \ldots, C\}$, the entry in position (i, j) represents the number of ballots in which candidate ci received a score of j.

An example of such a table for a contest with 7 candidates and 1000 ballots is shown below:

Table 2. Report of candidate's score

Candidate/Score	1	2	3	4	5	total score
c_1	251	201	172	191	185	3142
c_2	232	213	180	190	185	3117
c_3	155	223	293	176	153	3051
c_4	201	192	186	221	200	2973
c_5	161	171	169	222	277	2117

5 Contest and Audit

We aim to conduct an audit of the results of an election involving C candidates, where $k > 1$ candidates have been declared winners, and the remaining $C - k$ candidates have been declared losers. Let $\mathcal{W}$ and $\mathcal{L}$ denote the sets of winners and losers, respectively. Ballots can be categorized as either valid or invalid. Valid ballots contain values that are reported in favor of candidates, while invalid ballots are assigned a fiction candidate 0. Let the true, unknown total scores of the C candidates and fiction candidate be represented by the vector $S_C \equiv (S_i)_{i=0}^{C}$. The reported scores for the candidates are denoted by the vector $\widehat{S_C} \equiv (\widehat{S}_i)_{i=0}^{C}$, and the true unknown proportion of the total score is represented by the vector $(\pi_i)_{i=0}^{C}$.

The reported proportion of the total score for candidate j is defined as

$$P_j = \frac{\hat{S}_j}{\sum_{i=1}^{C} \hat{S}_i}, \qquad j \in \{1, 2, \ldots, C\}.$$

We use the fraction of valid scores obtained from valid votes, denoted by

$$t_j = \frac{P_j}{\sum_{i=1}^{C} P_i}, \qquad j \in \{1,2,\dots,C\}.$$

Since the denominator includes only the scores derived from valid ballots, rather than the total scores, it follows that $P_j < t_j$. Let w and l be pronounced as "winner" and "loser," respectively. Let $t_{wl} = \frac{t_w}{t_w+t_l}$ represent the fraction of valid scores obtained by w and l among all valid scores in the ballots cast during the contest. This fraction serves as the foundation for subsequent computations.

5.1 Audit of total scores

We employ a comprehensive strategy to verify the election results. Each of the k reported winners must defeat all $C-k$ reported losers in a two-candidate contest, meaning the total score of each winner must exceed the total score of each loser. Therefore, $k(C-k)$ hypotheses tests need to be conducted, or equivalently, $k(C-k)$ assertions must be verified:

$$S_w > S_l, \qquad \forall\, w \in \mathcal{W}, l \in \mathcal{L}.$$

To test these assertions, ballots are drawn randomly and uniformly from the cast votes. After observing the scores of candidates on the sampled ballot, the assertions for each pair $(w,l) \in \mathcal{W} \times \mathcal{L}$ are checked simultaneously. Upon rejecting or accepting each hypothesis during the audit, the corresponding candidates are either excluded or a manual recount is performed.

To evaluate the accuracy of the reported results, we apply a risk-limiting audit (RLA). In this process, $k(C-k)$ statistical hypothesis tests will be examined at a significance level α, as follows:

$$H_0 : \pi_w \leq \pi_l \qquad vs \qquad H_1 : \pi_w > \pi_l. \tag{1}$$

The procedure for the RLAs in score voting, based on the aforementioned statistical hypotheses, is outlined as follows:

1. Select the risk limit $\alpha \in (0,1)$, and M, the maximum number of ballots to audit before proceeding to a full hand count.
2. Set $m = 0$ and $T_{wl} = 1\ \forall (w,l) \in \mathcal{W} \times \mathcal{L}$
3. Randomly select a ballot without replacement from those cast in the contest, and then increment m.
4. Based on the observed difference between s_w and s_l in the extracted ballots for each $w \in \mathcal{W}$ and $l \in \mathcal{L}$, the value of T_{wl} is updated as follows:

   (i) $\quad T_{wl} \times \left(\frac{t_{wl}}{0.5}\right)^i \qquad$ if $\quad s_w - s_l = i, \qquad i \in \{1,2,\dots,C-1\}.$

   (ii) $\quad T_{wl} \times \left(\frac{1-t_{wl}}{0.5}\right)^i \qquad$ if $\quad s_w - s_l = -i, \qquad i \in \{1,2,\dots,C-1\},$

   where s_w and s_1 are the scores corresponding to w and l.

5. If $T_{wl} \geq \frac{1}{\alpha}$ for some (w,l), reject the corresponding null hypothesis and stop the operational computation related to those tests.

6. If all null hypotheses are rejected, conclude the audit, confirming that the reported results are valid and the election is verified. Otherwise, if $m < M$, proceed back to Step 3.

7. Once the maximum sample size M is reached, if at least one null hypothesis is accepted or no definitive result is obtained, perform a complete manual count, replacing the reported results with the results from the manual count.

The statistic used in audit method is as follows:

$$T_{wl} = (2t_{wl})^{\delta}(2 - 2t_{wl})^{\lambda}, \qquad \delta = \sum_{k=1}^{m_w} i_k \quad \lambda = \sum_{k=1}^{m_l} i'_k,$$

where $1 \leq m = m_w + m_l \leq M$ and $i_k, i'_k \in \{1, \ldots, C-1\}$. Here, m_w represents the number of ballots where the score of w exceeds the score of l, whereas m_l denotes the number of ballots where the score of l exceeds the score of w. This statistic is based on the difference in scores between w and l on the ballots. As the difference in scores on the drawn ballot approaches $C-1$, the probability of rejecting the null hypothesis increases, and as the difference in scores approaches $1-C$, the probability of accepting the alternative hypothesis decreases. Both of these properties directly impact the limiting risk α.

5.2 Audit of partial scores

The alternative method described here provides additional details about the election results. This approach allows for the examination of partial scores of both winners and losers and can serve as a complementary procedure to the previously discussed method. It is also more efficient for large values of N. As shown in Section 4, the total score of each candidate is the sum of their partial scores from 1 to C. Therefore, the parameters $(S_i, i = 1,2,\ldots,C)$ represent the sum of the parameters $(S_{ij}, j = 1,2,\ldots.,C)$. Since the reported values of S_{wi} and Sli are available, these corresponding parameters can be tested for each pair (w,l) and for each score $i \in \{1,2,\ldots,C\}$.

The parameters to be tested in this method are $\pi_{wi} = \frac{S_{wi}}{S_t}$ and $\pi_{lj} = \frac{S_{lj}}{S_t}$, where S_t represents the total score obtained from all valid votes in the contest.

In this method, we conduct a detailed analysis of the scores of w and l and perform the $Ck(C-k)$ hypothesis test using the statistics T_{wli} and T_{wlj}. It is important to note that scores of i were reported in favor of the winners, while scores of j were reported in favor of the losers.

Let $S_{wli} = \frac{S_{wi}}{S_w i + S_{li}}$ represent the fraction of partial scores of i belonging to w, reported to have been received among the ballots that show score i for either w or l. Similarly, the fraction $S_{wlj} = \frac{S_{lj}}{S_{lj} + S_{wj}}$ is defined in the same way. The values of S_{wli} and S_{wlj} are both greater than

0.5 and are used as the basis for the multiplier. Hypothesis tests are organized based on the $C_k(C-k)$ values of S_{wli} and S_{wlj} as follows:

$$H_0 = \pi_{wi} \leq \pi_{li} \quad vs \quad H_1 = \pi_{wi} > \pi_{li}$$
$$H_0 = \pi_{lj} \leq \pi_{wj} \quad vs \quad H_1 = \pi_{lj} > \pi_{wj}$$

The second risk-limiting audit (RLA) method for score voting, with respect to these statistical hypotheses, is outlined here. The steps for implementing the RLA are as follows:

1. Select the risk limit $\alpha \in (0,1)$, and M, the maximum number of ballots to audit before proceeding to a full hand count.

2. Set $m = 0$ and initialize $\{T_{wli} = 1, T_{wlj} = 1\}$ for all $(w,l) \in \mathcal{W} \times \mathcal{L}, i, j \in \{1,2,\ldots,C\}$.

3. Draw a ballot uniformly at random, without replacement, from the cast ballots and increment m.

4. Based on the ballot's content:

   - If the score i is observed for w, multiply T_{wli} by $\frac{S_{wli}}{0.5}$ for each $l \in \mathcal{L}$, and repeat for all such w.
   - If the score i is observed for l, multiply T_{wli} by $\frac{1-S_{wli}}{0.5}$ for each $w \in \mathcal{W}$, and repeat for all such l.
   - If the score j is observed for w, multiply T_{wlj} by $\frac{S_{wlj}}{0.5}$ for each $w \in \mathcal{W}$, and repeat for all such l.
   - If the score j is observed for l, multiply T_{wlj} by $\frac{1-S_{wlj}}{0.5}$ for each $l \in \mathcal{L}$, and repeat for all such w.

5. If $T_{wli} \geq \frac{1}{\alpha}$ for some (w,l,i), reject the corresponding null hypothesis and stop the operational related to that test.

6. If $T_{wlj} \geq \frac{1}{\alpha}$ for some (w,l,j), reject the corresponding null hypothesis and stop the operational related to those test.

7. Conclude the audit if all null hypotheses are rejected, confirming the reported results as valid and verifying the election. Otherwise, if $m < M$, proceed to Step 3.

8. After reaching the sample size M, if at least one of the null hypotheses is accepted or no definitive result is obtained, perform a complete manual count and replace the reported results with the results of the manual count.

This method ensures that the risk of certifying incorrect results is limited, while allowing for efficient verification of the election outcome based on partial score data.

6 Computational Results

In most multi-winner elections, candidates form coalitions to secure all the available seats. These coalitions are subsets of the candidates who participated in the contest. A portion of voters tend to assign the highest scores to such coalitions, and we refer to this proportion as the "index of directed votes." The true value of this parameter is not directly available, so it must be estimated using the reported election results. We denote this estimate as β and use it to determine the mean sample number in the auditing process for score voting. Our tables are constructed based on the value of β.

The method for calculating β is described as follows:
Consider an election with C candidates, k winners, and N valid votes, where k candidates are ultimately declared winners. The sum of the scores reported for the winners is denoted by S_w, and the total score for the election is $S_t = \frac{N_C(C+1)}{2}$. The score Sw can be decomposed into contributions from two subgroups of voters: the directed subgroup and the neutral subgroup. In the directed subgroup, a proportion β (the index of directed votes) of the highest scores, $(C, C-1, \ldots, C-k+1)$, are allocated to the reported winners. In the neutral subgroup, the scores for the winners are distributed uniformly, such that all candidates receive equal scores. Thus, we can express S_w as:

$$S_w = \beta N(C + C - 1 + \cdots + C - k + 1) + \gamma \frac{kS_t}{C}.$$

where γ and β are positive parameters.

It is straightforward to show that $\gamma = 1 - \beta$. After simplifying the above expression, we obtain the following formula for β:

$$\beta = \frac{S_w - \frac{Nk(C+1)}{2}}{\frac{Nk(C-k)}{2}}.$$

We have simulated the two described risk-limiting audit methods in Section 4 for the case where $C = 10$ and $k = 4$, with various values of N and different risk limits ($\alpha = 1\%$ and $\alpha = 5\%$), as well as for values of $\beta = 2\%, 3\%, 4\%, 5\%$. The simulations were implemented and analyzed using MATLAB. Due to the lack of access to large real data, we used simulated data for our analysis.

In the simulations, ballots are modeled as random vectors, where the arrays represent permutations of integers from 1 to 10. Tables 3 and 4 report estimates of the Average Sample Number (ASN), which represents the expected number of ballots required to either accept or reject the null hypothesis. For each combination of N, α, and β, we ran 100 simulations and computed the average number of ballots checked across these simulations. The notation R indicates cases where our method could not compute the ASN, necessitating a full recount of all ballots.

The results show that ASN is influenced by β, with higher values of β leading to a decrease in ASN. The numbers following R in the tables represent the threshold values, and the first and second methods provide the ASN for each number greater than these thresholds. For example, in the first column of Table 4, the second audit method provides the ASN for $(N \geq 200{,}000, \alpha = 1\%, \beta = 2\%)$.

Table 3. Average of ballots sampled over 100 simulations for the risk-limimting audit (Total scores)

	$\alpha = 1\%$				$\alpha = 5\%$			
N	$\beta = 2\%$	$\beta = 3\%$	$\beta = 4\%$	$\beta = 5\%$	$\beta = 2\%$	$\beta = 3\%$	$\beta = 4\%$	$\beta = 5\%$
10000	R	R	1236	871	R	R	742	566
20000	R	R	1302	791	3124	1312	854	532
30000	R	2123	1226	831	3118	1409	728	496
40000	R	2314	1220	910	3315	1434	795	459
50000	R	2182	1243	819	3275	1410	844	548
60000	5167	2120	1290	796	3148	1338	769	520
70000	5167	2062	1247	835	2904	1355	756	525
80000	5152	2435	1289	865	3170	1307	764	501
90000	5080	2275	1188	804	3404	1569	744	458
100000	5150	2251	1220	792	3419	1262	712	508
200000	5012	2250	1336	819	3127	1357	813	652
300000	5205	2593	1251	890	2922	1466	787	452
400000	5019	2346	1303	824	3132	1378	835	525
500000	4641	2232	1287	793	2947	1388	777	455
600000	4922	2307	1169	757	3217	1388	761	477
700000	4682	2169	1271	781	3263	1311	725	518
800000	4645	2330	1289	807	3074	1428	736	477
900000	4849	2212	1245	790	3088	1296	774	498
1000000	5105	2152	1227	879	3062	1311	684	492

As can be seen, more ballots are needed for auditing in the second method and this is reasonable because of the more number of hypothesis tests.

6.1 Performance and comparison

We employed the sequential probability ratio test (SPRT) method to conduct statistical analyses, following the approach outlined by Wald [17]. In the SPRT framework, the sample size is not predetermined; instead, an upper limit is defined. A key advantage of SPRT lies in its ability to produce valid results with smaller sample sizes compared to methods requiring a fixed sample size, all while preserving equivalent type I and type II error rates. Moreover, SPRT does not rely on assumptions about the underlying probability distribution, making it suitable for populations with unknown or unspecified distributions.

The specific outcomes of a score voting strategy do not necessarily align with those of other votingmethods, such as Condorcet voting, ranked-choice voting, or Instant-runoff voting. In a Condorcet voting system, the winning candidate is determined based on the highest preference rank in head-to-head competitions against other candidates. Ranked voting can be interpreted in various ways. The most common interpretation declares the candidate with the most first-place votes as the winner.

Instant-runoff voting, a subset of ranked voting, combines the Condorcet method with the elimination of lower-ranked candidates to identify the winner. Lindeman et al acknowledged

Table 4. Average of ballots sampled over 100 simulations for the risk-limimting audit (Partial scores)

	$\alpha = 1\%$				$\alpha = 5\%$			
N	$\beta = 2\%$	$\beta = 3\%$	$\beta = 4\%$	$\beta = 5\%$	$\beta = 2\%$	$\beta = 3\%$	$\beta = 4\%$	$\beta = 5\%$
10000	R	R	R	R	R	R	R	R
20000	R	R	R	8917	R	R	7781	5522
30000	R	R	13716	9011	R	R	7984	4940
40000	R	R	14325	9237	R	14525	8967	5375
50000	R	24685	13060	8502	R	16130	8316	5540
60000	R	24169	13642	9107	R	15274	8183	5428
70000	R	24712	14159	8949	31851	14947	8260	5332
80000	R	25133	14520	8946	31851	15693	8673	5501
90000	R	24289	13867	8870	31341	14242	8620	5437
100000	R	24863	13443	8824	30047	14443	8446	5530
200000	54851	25066	13477	9313	34625	15715	8688	5528
300000	59344	24952	15001	8461	37539	14595	8618	5431
400000	58182	26268	13606	8536	33728	16170	8399	5362
500000	62103	24676	15343	8965	35679	14405	8150	5234
600000	59914	25411	14038	9536	35318	15137	8573	5729
700000	56461	25709	14956	9132	33004	15197	8828	5255
800000	57358	25247	14261	9210	33802	15466	8229	5433
900000	55873	24837	14327	8567	34861	14785	9337	5351
1000000	57841	24387	14548	8808	30047	13869	8159	5775

that their audit method is not applicable to ranked voting systems [10]. In our approach, if the average of candidates' ratings on ballots is used as the criterion for victory, the outcomes of ranked voting and score voting coincide. As previously stated, two voting methods are comparable when they yield identical results. Consequently, score voting can only be compared with other score-based systems.

Sarwate and Shacham proposed an audit method specifically designed for score voting. Their method employs a statistic based on the margins between the actual and reported values. Comparative results indicate that the average sample number (ASN) in our method is 5-10% smaller than in their method. Table 5 presents the findings for $N = 100000, \alpha = 5\%, C = 10, k = 4$.

The results indicate that as the difference in votes between winners and losers increases (i.e., when β is larger), the ASN decreases. Generally, in cases where the voting population is large, the ASN values of the two methods exhibit minimal differences. However, there is a slight advantage to our method due to the simplicity of the test statistic.

7 Conclusion

We have defined two risk-limiting audit for score voting systems. These two methods can be used as complement to a real election. Our audit methods have two main weaknesses. The first is that the generating of data and drawing of ballots have been carried out in a simulation setting. This cause that some of the limitations of a real election such as incorrect votes are not considered. Second, it does not work well enough to implement in election with small number of ballots ($N < 10000$) when the amount of β is small. Research could focus on these

Table 5. Estimted sample size, as a nunber of ballots and precentage of the total ballots, requaried to audit the election results based on Sarwate and Shacham audit (Audit 1) and our audit (Audit 2).

β	Audit 1		Audit 2	
	ASN	%	ASN	%
2%	3591	0.03%	3419	0.03%
3%	1327	0.01%	1262	0.01%
4%	782	0.007%	712	0.007%
5%	553	0.005%	508	0.005%
6%	488	0.004%	446	0.004%
7%	341	0.003%	311	0.003%
8%	297	0.002%	273	0.002%
9%	139	0.001%	128	0.001%
10%	106	0.001%	97	0.001%

drawbacks in future.

Funding

This research received no external funding.

Data Availability Statement

Data is contained within the article.

Conflicts of Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

[1] A. Bauer, A. Bender, A. Klima, H. Küchenhoff, Koala: a new paradigm for election coverage: An opinion poll-based ”now-cast” of probabilities of events in multi-party electoral systems, AStA Advances in Statistical Analysis 104 (2020) 101–115. https://doi.org/10.1007/s10182-019-00352-6
[2] M. Blom, P. Stuckey, V. Teague, Raire: Risk-limiting audits for irv elections, arXiv preprint arXiv:1903.08804 (2019). https://doi.org/10.48550/arXiv.1903.08804
[3] S. Brams, P. Fishburn, S. Merrill, The responsiveness of approval voting: Comments on saari and van newenhizen, Public Choice 59(2) (1988)121–131. https://doi.org/10.1007/BF00054448
[4] A. Dawid, M, Musio, Estimation of spatial processes using local scoring rules: Spatial special issue, AStA Advances in Statistical Analysis 97 (2013) 173–179. https://doi.org/10.1007/s10182-012-0191-8

[5] T. Gneiting, A. Raftery, Strictly proper scoring rules, prediction, and estimation, Journal of the American statistical Association 102(477) (2007) 359–378.https://doi.org/ 10.1198/016214506000001437
[6] E. Herron, R. Pekkanen, M. Shugart, The Oxford handbook of electoral systems, Oxford University Press, 2018. https://doi.org/10.1093/oxfordhb/9780190258658.001.0001
[7] Z. Huang, R. Rivest, Ph. Stark, V. Teague, D. Vukcević, A unified evaluation of two-candidate ballot-polling election auditing methods, In Electronic Voting: 5th International Joint Conference, E-Vote-ID 2020, Bregenz, Austria, Proceedings, 5 (2020) 112–128. https://doi.org/10.1007/978-3-030-60347-2_8
[8] A. Hyvärinen, Some extensions of score matching, Computational statistics & data analysis 51(5) (2007) 2499–2512. https://doi.org/10.1016/j.csda.2006.09.003
[9] K. Johnson, Election certification by statistical audit of voter-verified paper ballots, Available at SSRN 640943, (2004). https://doi.org/10.2139/ssrn.640943
[10] M. Lindeman, Ph. Stark, V. Yates, Bravo: Ballot-polling risk-limiting audits to verify outcomes, EVT/WOTE (2012). https://www.usenix.org/system/files/conference/evtwote12/evtwote12-final27.pdf
[11] R. Rivest, Clipaudit: a simple risk-limiting post-election audit, arXiv preprint arXiv:1701.08312 (2017). https://doi.org/10.48550/arXiv.1701.08312
[12] R. Rivest, E. Shen, A bayesian method for auditing elections, EVT/WOTE (2012). https://www.usenix.org/system/files/conference/evtwote12/rivest_bayes_rev_073112.pdf
[13] A. Sarwate, S. Checkoway, H. Shacham, Risk-limiting audits and the margin of victory in non-plurality elections, Statistics, Politics, and Policy 4(1) (2013) 29–64. https://doi.org/10.1515/spp-2012-0003
[14] Ph. Stark, Conservative statistical post-election audits, Ann. Appl. Stat. 2 (2008) 550–581. https://doi.org/10.1214/08-AOAS161
[15] Ph. Stark, Sets of half-average nulls generate risk-limiting audits: Shangrla, Financial Cryptography and Data Security: FC 2020 International Workshops, AsiaUSEC, CoDeFi, VOTING, and WTSC, Kota Kinabalu, Malaysia, Revised Selected Papers 24(2020) 319–336. https://doi.org/10.1007/978-3-030-54455-3_23
[16] Ph. Stark, V. Teague, A. Essex, Verifiable european elections: Risk-limiting audits for {D'Hondt} and its relatives, USENIX Journal of Election Technology and Systems (JETS) 1 (2014) 18–39. https://www.usenix.org/sites/default/files/jets_0301_issue.pdf
[17] A. Wald, Sequential tests of statistical hypotheses, Ann. Math. Stat. 16(2) (1945) 117–186. https://doi.org/10.1214/aoms/1177731118
[18] R. Weber, Approval voting, Journal of Economic Perspectives, 9(1) (1971) 39–49. https://doi.org/10.1257/jep.9.1.39

Citation: H, Devisti, M, Hadian, Risk-limiting audits for score voting systems, J. Disc. Math. Appl. 10(2) (2025) 207–221.
https://doi.org/10.22061/jdma.2025.11534.1108

COPYRIGHTS
©2025 The author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, as long as the original authors and source are cited. No permission is required from the authors or the publishers.