Auditing Ranked Voting Elections with Dirichlet-Tree Models: First Steps*

Floyd Everest[1][0000−0002−2726−6736], Michelle Blom[2][0000−0002−0459−9917], Philip B. Stark[3][0000−0002−3771−9604], Peter J. Stuckey[4][0000−0003−2186−0459], Vanessa Teague[5,6][0000−0003−2648−2565], and Damjan Vukcevic[1,7][0000−0001−7780−9586]

[1] School of Mathematics and Statistics, University of Melbourne, Parkville, Australia
[2] School of Computing and Information Systems, University of Melbourne, Parkville, Australia
[3] Department of Statistics, University of California, Berkeley, CA, USA
[4] Department of Data Science and AI, Monash University, Clayton, Australia
[5] Thinking Cybersecurity Pty. Ltd., Melbourne, Australia
[6] The Australian National University, Canberra, Australia
[7] Melbourne Integrative Genomics, University of Melbourne, Parkville, Australia
damjan.vukcevic@unimelb.edu.au

Abstract. Ranked voting systems, such as instant-runoff voting (IRV) and single transferable vote (STV), are used in many places around the world. They are more complex than plurality and scoring rules, presenting a challenge for auditing their outcomes: there is no known risk-limiting audit (RLA) method for STV other than a full hand count.
We present a new approach to auditing ranked systems that uses a statistical model, a Dirichlet-tree, that can cope with high-dimensional parameters in a computationally efficient manner. We demonstrate this approach with a ballot-polling Bayesian audit for IRV elections. Although the technique is not known to be risk-limiting, we suggest some strategies that might allow it to be calibrated to limit risk.

In *ranked voting*, voters rank candidates in order of preference; some elections require a complete ranking, others allow partial rankings. Counting the votes can be complex, e.g. involving potentially long sequences of eliminations of candidates (for IRV), and transfers of weighted votes between candidates (for STV).

Complexity arises in two ways: (i) a very large number of ways to fill out a ballot ($k!$ ways to rank k candidates); (ii) the social choice functions are sensitive, small changes can sometimes drastically alter the outcome. This poses a challenge for auditing: we require statistical inference in a very high-dimensional parameter space, for a function prone to erratic behaviour.

RLAs have been developed for some ranked voting systems: (i) IRV elections [1]; (ii) 2-seat STV elections [2]. Both RLAs project into lower dimensions, where

* We thank Ronald Rivest for many helpful suggestions for improving the paper. This work was supported by the University of Melbourne's Research Computing Services and the Petascale Campus Initiative; and by the Australian Research Council (Discovery Project DP220101012).

statistical testing is tractable. However, their projections typically capture only a subset of elimination sequences that lead to the winner. If the true sequence is not one of those, but leads to the same winner, then the audits will usually (and unnecessarily) escalate to a full count despite the reported winner being correct.

Thus, there is scope for further development for ranked systems. For IRV we seek a method that can work with a more complete set of elimination sequences, and for STV we want to be able to audit elections with more than 2 winners.[8]

We tackle the problem directly as a Bayesian audit [6]. This is challenging in high-dimensions; a previous attempt [3] gave up on fitting a full model and instead used a bootstrap approach (equivalent to a degenerate Bayesian model).

Our contribution is a new specification of the statistical model that works efficiently in high-dimensions, making Bayesian audits possible for ranked voting elections. We demonstrate this with examples of auditing IRV elections.

1 Dirichlet-tree model for ranked voting

An audit involves calculating the evidence in favour of the reported outcome using a sample of ballots and a statistical model. For ranked voting, the natural model is multinomial: each ballot type (ranking of the candidates) occurs with some (fixed but unknown) frequency across all ballots.

A Bayesian audit can work with this model directly, by specifying a prior distribution on the ballot probabilities. Given a sample of ballots, we obtain a posterior distribution for these probabilities, which induces a posterior distribution on the winner(s). If the reported outcome exceeds some desired posterior probability threshold, we stop the audit, otherwise we sample more ballots.

For a multinomial model, a typical choice of prior is a Dirichlet distribution. This is conjugate, allowing convenient and efficient implementation. It is defined by concentration parameters, $a_i > 0$, for each ballot type $i \in \{1, 2, \ldots, K\}$. The posterior is Dirichlet (by conjugacy) with concentration parameters $a_i + n_i$ after observing n_i ballots of type i. To make the prior candidate-agnostic: $\forall i, a_i = a_0$ for some a_0. Setting $a_0 = 1$ gives a uniform density on the space of probabilities.

This model behaves poorly as K grows very large. If we set $a_0 = 1$, the prior becomes very informative: it will swamp the data, making the posterior converge very slowly. If we set a_0 much smaller, for example $a_0 \approx 1/K$, then the posterior will strongly concentrate on the ballot types observed in the sample, approximating a 'bootstrap' method. This will likely understate the uncertainty. It will also be challenging to implement, with values of $1/K$ being smaller than typical machine precision once there are about 30 candidates.

To overcome these issues, we propose using a Dirichlet-tree prior distribution (e.g. [5]).[9] This is a set of nested Dirichlet distributions with the nesting described by a tree structure. It generalises the Dirichlet while retaining conjugacy with the multinomial. The nesting divides up the space, allowing efficient inference in high dimensions.

[8] E.g., Australian Senate elections use STV to elect up to 12 candidates for each state.

[9] Our implementation is available at: https://github.com/fleverest/elections.dtree

The tree structure we propose follows the preference ordering: the first split in the tree has a branch for each possible first preference (one branch per candidate), the next split has a branch for each possible second preference (amongst remaining candidates), etc. Partial ballots are modelled by 'termination' branches. To initialise the prior, we set the concentration parameter for each branch to be equal to a_0; see Figure 1 for an example with no partial ballots.

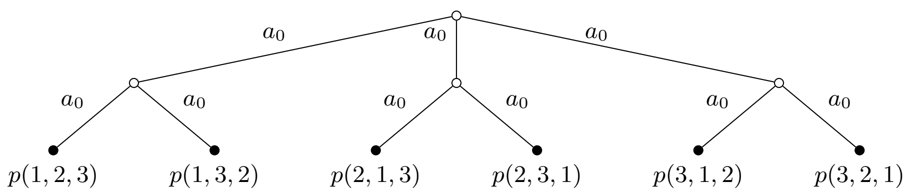

Fig. 1. Dirichlet-tree prior for ranked voting ballots with 3 candidates.

2 Ballot-polling Bayesian audits of IRV elections

We demonstrate our model using data from two elections of different sizes: (i) Seat of Albury, NSW 2015 lower house elections, Australia [5 candidates; 46,357 ballots]; (ii) San Francisco Mayoral election 2007 [18 candidates; 149,465 ballots].[10] The latter has more than $18! \approx 6.4 \times 10^{15}$ possible ballot types.

We used a Dirichlet-tree prior that allows partial ballots and had $a_0 = 0, 1, 10, 100$. We also used a Dirichlet prior with a_0 set such that its prior variance, for an arbitrary complete ballot, matched that of the corresponding Dirichlet-tree prior. Setting $a_0 = 0$ for either prior gives a 'bootstrap' audit [3].

For each election, we simulated 100 audits by randomly permuting the ballots (without introducing any errors). We took samples of up to 200 ballots for Albury and up to 50 for San Francisco, which was sufficient to illustrate the differing behaviour of the priors. At each point in the audit, we estimated posterior probabilities by taking the mean of 100 draws from the posterior.

Figure 2 shows how the posterior probability for the true winner evolved as the samples increased. The Dirichlet-tree model worked for both elections and responded to a_0 as expected: increasing it made the prior more informative and hence respond more slowly to data. The Dirichlet model behaved similarly when we had only a few candidates (Albury) but unstable when we had many (San Francisco), with all choices except the bootstrap ($a_0 = 0$) being too informative.

The bootstrap was erratic at the start (a wide range of posterior values) and stabilised once the sample was big enough. In practice, the poor regularisation at the start would lead to increased risk. Whether this can be curbed by simply specifying a minimum sample size is worth investigating in general.

[10] Data source: https://github.com/michelleblom/margin-irv

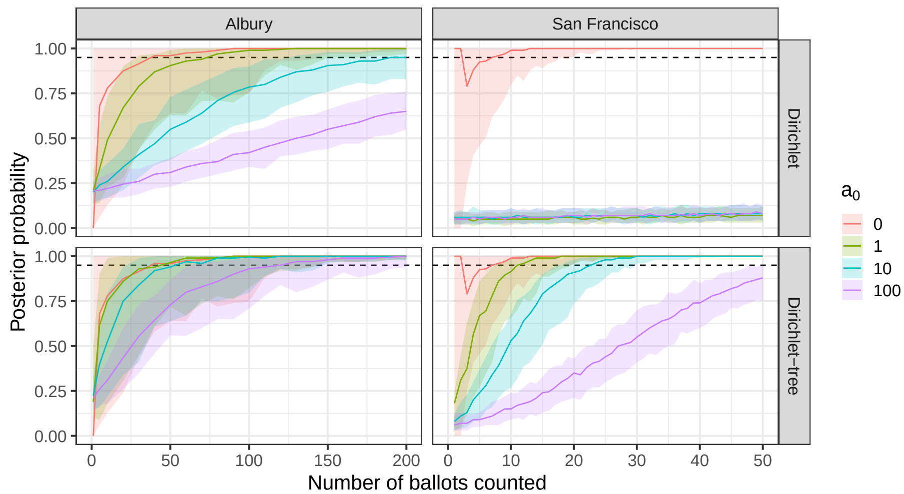

Fig. 2. Distribution of the posterior probability for the winner, vs sample size. The lines show the median across 100 simulated audits, the corresponding bands shade the values between the 5% and 95% quantiles. The dashed line shows a posterior probability of 0.95, for reference. The a_0 values refer to the Dirichlet-tree prior; for the Dirichlet a 'corresponding' value was chosen (see main text).

3 Discussion

We have demonstrated a statistical model that allows efficient ballot-polling Bayesian audits of ranked voting elections. While our example was specifically for IRV, the model can be applied to any ranked voting election by simply changing the social choice function in the calculation of the posterior distribution. Furthermore, the tree structure can be adapted to better suit specific features of particular elections, which should improve efficiency.

A current limitation of our approach is that it cannot be used to run an RLA. This requires an easy way to compute or impose a risk limit. We propose two ways to overcome this: (i) determine the maximum possible risk by deriving the worst-case configuration of true ballots, such as was done for 2-candidate elections [4]; (ii) use a prior-posterior ratio (PPR) martingale [7] to make an RLA using the Dirichlet-tree model. Another limitation is that our approach currently only supports ballot-polling audits. Adapting it to allow other types of audits, such as comparison audits, is another important avenue for future work.

References

1. Blom, M., Stuckey, P.J., Teague, V.: RAIRE: Risk-limiting audits for IRV elections. arXiv:1903.08804 (2019), Preliminary version appeared in Electronic Voting (E-Vote-ID 2018), Springer LNCS 11143.
2. Blom, M., Stuckey, P.J., Teague, V., Vukcevic, D.: A first approach to risk-limiting audits for single transferable vote elections. arXiv:2112.09921 (2021)

3. Chilingirian, B., Perumal, Z., Rivest, R.L., Bowland, G., Conway, A., Stark, P.B., Blom, M., Culnane, C., Teague, V.: Auditing Australian Senate ballots. arXiv:1610.00127 (2016)
4. Huang, Z., Rivest, R.L., Stark, P.B., Teague, V.J., Vukcevic, D.: A unified evaluation of two-candidate ballot-polling election auditing methods. In: Electronic Voting. Lecture Notes in Computer Science, vol. 12455, pp. 112–128. Springer, Cham (Sep 2020). https://doi.org/10.1007/978-3-030-60347-2_8, Preprint: arXiv:2008.08536
5. Minka, T.: The Dirichlet-tree distribution. Tech. rep., Justsystem Pittsburgh Research Center (July 1999), https://www.microsoft.com/en-us/research/publication/dirichlet-tree-distribution/
6. Rivest, R.L., Shen, E.: A Bayesian method for auditing elections. In: 2012 Electronic Voting Technology/Workshop on Trustworthy Elections (EVT/WOTE '12) (2012)
7. Waudby-Smith, I., Ramdas, A.: Confidence sequences for sampling without replacement. In: Advances in Neural Information Processing Systems. vol. 33, pp. 20204–20214. Curran Associates, Inc. (2020), https://proceedings.neurips.cc/paper/2020/file/e96c7de8f6390b1e6c71556e4e0a4959-Paper.pdf, arXiv:2006.04347