Instant Runoffs with Self-Apportionment

Dave Robinson, Jul 2019

Contents:

Abstract:

The "first past the post" or plurality election methods commonly used in the United States, United Kingdom, and Canada for election of a governing body (council, board or legislature) routinely leave large fractions of constituents without satisfactory representation, and large imbalances in mandate among winners. They usually require complex redistricting schemes that are easily manipulated by pitting arbitrary subsets of voters against each other. An objective method is proposed here that allows voters to naturally group themselves, leaving nearly everyone with equal and satisfactory representation. It usually arranges that even a narrow majority of elected legislators accountably represents a majority of voters, which is rarely true with the commonly used methods. While more complex than a plurality vote, it is less so than the current methods to define plurality voting districts. It is a refinement of the well-known Instant Runoff and Single Transferable Vote methods. This version more clearly defines constituencies, is easier to count and audit, and takes full advantage of the mathematical insights behind them.

Overview:

This method allows voters to choose representatives from a pool of candidates numbering at least as many as are to be chosen (the number of seats to be filled). Voters rank candidates in order of preference. Each voter has several votes. The procedure casts each of these toward one's highest usable ranking. Some candidates may be eliminated, and candidates with more votes than needed to win may allow voters to apply some of their votes to lower choices, according to prescribed rules. Adjustments are made in every round until the number of remaining candidates equals the number of seats to be filled, and the candidates have nearly equal numbers of votes.

Procedure:

0. Voting

Voters rank candidates in order of preference, with the understanding that if circumstances prevent any of their votes from counting for a higher choice, they will count for a lower choice.

1. Counting votes

Each ballot has a number of votes equal to P (the "precision"), which is 1 for a single winner contest, or otherwise the number of seats multiplied by one more than the number of seats. P can optionally be a whole-number multiple of this. Equivalently, each ballot could be described as having one vote that can be split into fractions with P as the denominator. As described here, a candidate can be a remaining candidate or an eliminated candidate. Remaining candidates include winners and unelected candidates who have not been eliminated.

Votes from each ballot are assigned to its highest-ranked remaining candidate, constrained as follows:

A. Each remaining candidate has a prescribed maximum number of votes per ballot, initially equal to P. For example, if the maximum of one's first choice is P-1, then the ballot would count P-1 for the first choice, and 1 for the second choice.

B. If there are leftover votes after applying constraint A, they are reassigned to their lowest-ranked winner if one exists, regardless of the maximum.

C. Ballots that cannot count votes to any remaining candidate are counted for "none of these".

For each winner, a record (histogram) is kept of how many transferable ballots count a given number of votes for each winner. A ballot is transferable if there is a lower-ranked candidate who has not been eliminated. The histogram is kept sorted from highest to lowest number of votes per ballot.

2. Identification of winners

A. The winning threshold is calculated as the total number of votes for continuing candidates divided by one more than the number of seats.

B. Any candidate with more votes than the winning threshold is declared a winner. (Equivalently, the threshold could also be defined as being one vote more, and a candidate exactly meeting the threshold is also declared a winner.)

C. If there are not more remaining candidates than seats to be filled, all remaining candidates are declared winners.

3. Maximum reduction

A. If at least one unfilled seat remains: the histogram of each winner is consulted to determine the lowest maximum that the winner could have while staying above the winning threshold, and that reduction is applied. A proposed maximum can be checked by multiplying the number of ballots in each histogram bin by the difference between its number of votes and the proposed maximum, discarding any negative values, and summing these. The correct maximum can be quickly found by proposing it to be the number of votes per ballot in successive histogram bins starting with the second highest (or zero if the end of the histogram is reached). If this proposed maximum is too low, a proportional (interpolated) and rounded-up value between this and the previous proposed maximum is the correct value.

B. If only winners remain, and all were declared winners earlier than this round, the "ideal" threshold is calculated as the total number of votes for the winners divided by the number of seats. The histogram of each winner is consulted to determine the lowest maximum that the winner could have while staying above the ideal threshold, and that reduction is applied, following the same procedure as in 3A.

4. Elimination

If no maximum reductions are possible in step 3, the unelected remaining candidate with the fewest votes is eliminated. The candidate's maximum is set to zero. A tie in this result is resolved by attempting to identify which tied candidate had the fewest votes in successively previous rounds that elected or eliminated a candidate, or randomly if this is not successful.

5. If a candidate was newly elected in step 2, or a maximum reduced in steps 3 or 4, return to step 1. Otherwise, the count is complete.

Sometimes this procedure results in several rounds of maximum reductions before a candidate is elected or eliminated. It can be easier to fathom the results if these are lumped into a single round. More specifically, a new round would be reported only for those with a new winner; those with a new elimination; rounds just after an elimination (because the threshold may have been reduced); when step 3B is first used; or in a final round summarizing all of the winners.

For step 1, a noteworthy choice of the whole-number multiple for P is to specify that the minimum number of votes per ballot must be 55. This way, contests with 2 to 5 seats have 60 votes per ballot and those with 6 to 9 seats range from 56 to 90. This makes the completeness of surplus transfers (as discussed below) less dependent on the number of seats.

If desired for convenience, mid-term vacancies due to resignations or disqualifications can be refilled by recounting the election, considering only the unelected candidates, and using only the votes counting toward the vacated seats along with the "none of these" ballots. (Cambridge, MA uses this method without including “none of these” ballots.)

If a few ballots are to be added or removed, this can often be done without a full recount. In each round, the ballots are added or subtracted from the tally. The number of votes needed to win is corrected to account for the added ballots. The new totals are checked to see if there are changes to who is newly elected in the round, if there are any changes to maxima, and if the order of maximum reductions and eliminations has changed. If there are no such changes, a full recount is not necessary.

Rationale behind this method:

This voting method uses lists of preferences made by voters from among a meaningful variety of viable candidates to elect a board, council, or legislative body in which:
- all members have a nearly equal mandate
- nearly all voters have satisfactory representation, a fact documented by the election results
- a clear and traceable definition of constituencies exists to connect voters to representatives
- a majority of the legislative body traceably represents a majority of voters.

This is achieved without the need for a separate redistricting process, or at least a less complicated one that divides a large legislature into geographic groups of perhaps 5 to 15 seats. It is a variation of the "single transferable vote" methods, especially similar to the Warren version. (See the first issue of the online journal "Voting Matters" for details of that method). The main refinements of Warren’s method are the keeping of ballots that rank winners out of the “none of these” pile; inspecting lower choices for remaining candidates to help keep votes out of that pile (a practice used in Cambridge, MA STV elections), the use of a discrete number of votes rather than high-precision fractional votes or rounding, and clearer definitions of the method of maximum reductions and tie breaking (borrowing from other standard STV methods for ties, such as those of the UK Electoral Reform Society and the Proportional Representation Society of Australia).

STV has been around for more than 100 years, so why do we need another variation? Because it still has room for improvement, and it is still not used as widely as many of us think it should. A main reason for this is that it is difficult to implement and understand, and the advantages are not always clear to people. Improvements to the method that build on past successes and failures can reduce the magnitude of these obstacles.

This proposed method has the following advantages over older Single Transferable Vote methods:

1. It is reasonably easy to hand count and audit, using whole numbers of votes. There are no worries about floating point errors or rounding precision. The reliance of the Meek and Warren methods on taking many infinitesimal slices off each ballot is absent. However, the principles of the Meek and Warren methods are still satisfied while keeping the plausibility of hand counts.

2. As with the Meek and Warren methods, it is simple and deterministic: who one's ballot counts for can be determined at any point knowing only one's rankings and the maximum number of votes per ballot countable toward a candidate. You do not need to know who the ballot counted for previously. Randomness is used only to break ties, and this is usually unnecessary. Independence on detailed ballot history can simplify audits.

3. With most other versions, if a single ballot needs to be added or removed from the count, or if a mistake was made in an early round, it is necessary to recount the entire election in order to have an exact result. With IRSA, this is rarely the case, because surplus transfers occur in coarser steps.

4. It scales well down to a small number of ballots, making the method usable by small organizations or communities. A single ballot with a sufficient number of rankings can produce a meaningful outcome. Many methods sacrifice this feature by using election thresholds rounded to the nearest ballot, a facade for the complexity of floating-point ballot fractions. When there are few votes, this causes strange results where not enough candidates can meet the threshold. Discrete votes allow whole-number

thresholds while mostly avoiding this problem.

5. It approaches the ideal outcome: the winners average, and are close to, (100/seats)% of the vote, except for a small pile of ballots that rank no winners. In contrast, the Meek method is very good at giving winners a mandate that looks equal, but does so by placing many countable votes into an "exhausted" pile similar to the "none of these" pile.

6. It captures or closely approximates desirable properties of traditional STV methods. Voters in a constituency with enough votes for two seats can all rank the same favorite and second favorite and elect both. Surplus transfers are fair and deterministic, and strategic voting is not worthwhile in practice.

7. As with the original Warren method, it shows moderation and equity in the division of ballots among candidates, helping to provide a clear connection between voters and their representatives. Nearly everyone ends up with one or two representatives, both receiving a large fraction of their votes. In the process, it does a good job of honoring higher rankings.

8. It provides a clear measure of unsatisfied voters. Only ballots that do not rank any winning candidates appear in the "None of these" pile.

Disadvantages in perspective:

1. With the limited precision of discrete votes, surplus transfers are not always complete, and the ability of a candidate to transfer can depend on how many seats are up for election and how many other candidates remain. In practice, the risk of an objectionable result becomes very low well before unfathomably fine precision must be used. In communities that consider this a major concern, the method still works well if a large whole-number multiplier is used as described in Step 1. Most elected candidates have more than one chance to transfer a surplus, greatly improving completeness.

2. Fault can be found with any solution to the problem of dividing a winner's ballots into some that transfer and some that do not. Most other approaches either involve either randomness or more dividing. In others that prioritize votes by which round they first counted for a candidate, most of the ballots can be left undivided, at the expense of a small number that become divided finely. The use of maxima, as proposed by Warren, gives transfer priority to previously undivided ballots. This way, every ballot has an equal opportunity to be divided, but is unlikely to be divided more than once. Such equal treatment removes opportunities for strategic voting, as Meek and Warren realized. Limiting repeated divisions more clearly establishes a small number of specific winning candidates as one's representatives, and makes it easier to audit the election and determine who one's vote counted for.

3. Another dilemma is how to transfer ballots that do not list a lower choice. Resolving it requires a standardized interpretation of a voter's intent. If a voter only lists one ranking, a common-sense interpretation is that the voter wants the ballot to count unconditionally for that candidate. Some people strongly feel that these voters should have their surpluses sent to the "none of these" pile and their vote should count less than others; I. D. Hill, an advocate of the Meek method, claimed that this is fairer to remaining candidates. However, the original candidate is a remaining candidate, and it is by no means fair that he or she should be singled out for exclusion. Every voter should have the right to direct a surplus to any remaining candidate. It is unlikely that the voter prefers that the ballot be counted for less than its full value, and doing so contradicts our goal of maximizing the number of votes counted toward winners. Hill's preferred approach causes the "enough" threshold to decrease more than necessary, which can elect candidates who have less of a mandate, and induce surplus transfers from candidates the voter does not support.

4. The use of discrete votes makes it difficult to accommodate voters who want to give the same rank to several candidates. One could imagine doing this by splitting votes evenly among candidates given the same rank, constrained by maxima. When using discrete votes, there can be a remainder that is must be counted somehow, but the remainder could change later. This would cause further complications, and is probably not worthwhile. Spreading votes among many candidates is contrary to this method's goal of creating a clearly defined relationship between a voter and a small number of representatives.

5. If there are two winners A and B, there may be votes counting for B from ballots that list A followed by B, and votes counting for A that list B followed by A. Economics professors would argue that the voters should trade representatives. This could be arranged after the winners are determined, or even prevented during the count, but it would add complexity. No well-known STV method address this problem, and plurality methods cannot even draw attention to the problem because they do not collect enough information from voters. Methods based on the Warren approach help to minimize this problem by keeping the splitting of votes among candidates to a minimum.

7. All STV methods are subject to versions of the "nonmonotonicity" scenario. Here, ranking a candidate causes one candidate to be eliminated, whose voters' lower choices allow an undesired candidate to win, whereas not ranking the candidate changes the elimination order in a way that the desired candidate wins. Similar scenarios pertain to surplus transfers rather than eliminations. These scenarios are not commonly observed in real elections, and are impractical voting strategies unless the voter has knowledge of exactly how everyone else will vote. These can be seen as a symptom of the way Single Transferable Vote methods interpret your vote: they drive a hard bargain to keep your vote with your higher choices, which many voters would want to do if manually voting in multi-round runoff elections. A family of methods known as "pairwise" or "Condorcet" avoids the nonmonotonicity problems, at the cost of quickly compromising on lower-ranked candidates.

4. In general, the rule set is not as simple as possible, but it is not more complicated than many prevailing STV procedures. Some rules make actual counting easier, make results easier to interpret, create a clearer connection between winners and constituents (fewer repeatedly sliced ballots), and allow a clearer measure of unsatisfied voters.

Other design details, and possible variations:

Number of actions per round – The original Meek and Warren methods perform transfers from all winners simultaneously, an approach that is honored here. If transfers from winners are handled one at a time instead, it can be easier to keep track of which votes are going where. Ties can occur more frequently, but they are more easily resolved by considering previous rounds.

The total number of rounds may appear disturbingly larger than some other counting methods if the consolidation steps described above are not used. Meek elections typically do not show the many intermediate steps used to determine candidate weights (the equivalent of maxima), and methods like ERS97 (by the UK Electoral Reform Society) include substages. Another possible consolidation method is to combine elimination of losers whose votes could not make another candidate win or whose combined votes are still lower than any other candidate. The latter could also reduce the amount of counting at the expense of creating additional rules that are hard to explain. It could be useful to distinguish between rules for counting versus rules for reporting results, to emphasize that reporting rules do not affect the outcome of an election.

How to estimate completeness of transfers – when votes are recycled to winners, without a method to look ahead to see which votes will recycle, it can be difficult to identify how much a maximum can be reduced, creating the possibility that it is not reduced as much as possible. Such look-ahead procedures can make counting more tedious. The method described here provides a reasonable balance between complexity and completeness, making it improbable that an incomplete transfer could change an election outcome.

More generally, the major features that distinguish IRSA from earlier STV versions enable each other, or provide added benefits by their combination. Multiple whole-number votes enable exact thresholds, and simplify surplus transfers that minimize repeated slicing and simplify incremental ballot additions. Keeping ballots with winners out of the "none of these" pile enables convergence on equal shares of the vote for each winner, helps make surplus transfers more complete without being too complicated, and is also necessary for exact thresholds. Half-baked versions of IRSA are conceivable, but they would have less than half the benefits.

Closing remarks:

Most election reformers are ironically conservative about the methods they support, resisting new proposals such as this one and favoring those that have been around longer. This is sensible, given the plethora of ideas and approaches out there. The fact remains, though, that STV does not have the balance between performance and complexity to get across the "valley of death" that technologies must bridge to get from a few specialty uses to widespread adoption, and significant conceptual changes are needed to achieve this. A voting population must have both a good understanding of the method and confidence in its performance, or the voters will be easily swayed by arguments made to reject or repeal the method, even if the arguments are untruthful or made by those who do not have the best interest of all voters at heart. I believe that the approach proposed here offers major improvements to both performance and simplicity of implementation that will allow obstacles to adoption to be overcome.

Some references:

The Meek and Warren methods are described in detail in the online journal Voting Matters.
http://www.votingmatters.org.uk/
Issue 1 introduces the Meek and Warren methods, with a proof of the existence and uniqueness of the surplus transfer results by Douglas R. Woodall.
Jeff O'Neill addresses tie breaking and Issue 18, and tree-based counting in Issue 21.
David Hill advocates Meek in numerous issues.

I.D. Hill, B.A. Wichmann, D.R. Woodall. Algorithm 123 – Single Transferable Vote by Meek's Method. The Computer Journal, Volume 30 Issue 3, June 1987, pages 276-281, Oxford University Press.

- Provides a computer program for counting by the Meek method, and a mathematical argument that the method works as designed.

ERS Newland-Britton: http://www.rosenstiel.co.uk/stvrules/intro.htm

Australian PR society: http://www.prsa.org.au/rule1977.htm

- Avoids exhausted ballots during surplus transfer
- Prioritizes transfers by largest surplus, breaking ties by previous round or lot
- Lots of rounding
- Elected candidates can't accept votes

Cambridge: Massachusetts General Laws, Chapter 54A

- Vacancies filled using vacating candidate's ballots
- Avoid exhausted ballots during transfer (by city rule, not state law)

New Zealand Meek STV: http://www.stv.govt.nz/stv/electing.htm

Maltese STV: http://www.maltadata.com/

Irish STV:
http://aceproject.org/regions-en/countries-and-territories/IE/case-studies/ireland-the-archetypal-single-transferable-vote-system-1997/