Notes on Ballott Problems and Counting Principles

September 12, 2018

Abstract

These are some notes on Ballot Problems and some Principles of Counting, covering some material from the second lecture.

0.1 The Bertrand Ballot Problem

An election is held with paper ballots. At the end of voting, the ballots are withdrawn from the ballot box in a random order, one at a time, and the votes are tallied. Candidate A receives N votes and candidate B receives M votes with $N > M$. *What is the probability at candidate A has a lead at every stage of the tally?*

Before deriving the answer, let's carefully set this up as a problem in probability theory. The *sample space* S is the set of all sequences of length $M + N$ in which N of the entries are A (these correspond to votes for candidate A) and M of the entries are B (these correspond to votes for candidate A). For example, if $N = 5$ and $M = 3$ one such sequence would be

$$AAABBBAA\ . \tag{0.1}$$

The jth entry in the sequence (from the left) represents the jth vote tallied, so in this example, B has tied A when the sixth vote is tallied, so for this outcome, A does not have a lead at every stage.

The total number of outcomes in the sample space is $\binom{N+M}{M} = \binom{N+M}{N}$. Each sequence has $M + N$ terms and to specify it, we must choose which N of them are to be filled with votes for A, or equivalently, which M of them are to be filled with votes for B. If the ballots are well-mixed in the box, then all outcomes are equally likely, and we assign each individual outcome a probability of $\binom{N+M}{M}^{-1}$. If $G \subset S$ is any event and $\#(G)$ denotes the number of outcomes in G,

$$P(G) = \#(G)\binom{N+M}{M}^{-1}\ . \tag{0.2}$$

Now let us break the same space up into 3 mutually disjoint events – one of which is the event of interest: The events are defined as follows

$$\begin{aligned} E_1 &= \text{the event that the first vote is for B} \\ E_2 &= \text{the event that the first vote is for A and at some point there is a tie} \\ E_3 &= \text{the event that A has the lead at every stage} \end{aligned}$$

Note: We do not count 0 to 0 before the votes begin to be counted as at tie, only ties after the tally has started are counted.

It is easy to see that E_1, E_2 and E_3 are mutually disjoint:

$$E_1 \cap E_2 = \emptyset \ , \quad E_2 \cap E_3 = \emptyset \ , \quad E_3 \cap E_1 = \emptyset \ .$$

Also, A gets the first vote, and there is never a tie i and only if A has the lead at every point. Therefore, $E_2 \cup E_3$ is the event that the first vote is for A. Since the first vote is either for A or for B,

$$S = E_1 \cup E_2 \cup E_3 \ .$$

It follows that

$$1 = P(S) = P(E_1) + P(E_2) + P(E_3) \ ,$$

and therefore that

$$P(E_3) = 1 - P(E_1) - P(E_2) \ . \tag{0.3}$$

We will solve the problem by computing $P(E_1)$ and $P(E_2)$. It is easy to compute $P(E_1)$: If the first vote is for B, there are $N + M - 1$ votes remaining, N of them for A, and the outcomes are determined by choosing the N places to put the votes for A. There are

$$\binom{N+M-1}{N} = \frac{(N+M-1)!}{N!(M-)!} = \frac{M}{N+M}\binom{N+M}{N}$$

ways to do this. Therefore, by (0.2),

$$P(E_1) = \frac{M}{N+M} \ . \tag{0.4}$$

Another way to see this is the there are $N + M$ ballots from which the first is to be chosen, and M of these are votes for B.

Next we will show that $P(E_2) = P(E_1)$, or, what is the same thing, that $\#(E_2) = \#(E_1)$. We do this by defining a function f on E_2 such that for each outcome $\omega \in E_2$, $f(\omega) \in E_1$, and moreover, the function is one-to-one and maps E_2 onto E_2. It therefore sets up a one-to-one correspondence between the elements of E_1 and E_2, and this will show that $\#(E_2) = \#(E_1)$.

Here is the definition of the function f. Let ω be any outcome in E_2, and suppose that the *first* tie occurs when $2j$ votes are counted, $j = 1, \ldots, N - 1$. There may be several ties, so to get a well-defined function, we choose the first tie. Define

$$f(\omega) = \omega^*$$

where ω^* is obtained from ω by swapping all of the A and B votes up to the first tie, and leaving the rest unchanged. Since votes were swapped at a tie, the total number of votes for each of A and B is not changed; they still have, respectively, N and M votes each. For example, consider the outcome in (0.1), which has a tie at the sixth vote tallied

$$\omega = AAABBBAA \quad \text{and} \quad \omega^* = BBBAAAAA\ . \tag{0.5}$$

Notice that for ω^*, there is still a tie at the same place, and it is still the *first* tie – not only in this example, but in general. So if we apply the same transformation rule to ω^* to produce ω^{**}, we simply undo the first swap by swapping back. That is,

$$\omega^{**} = \omega\ . \tag{0.6}$$

Since the input of the function $f(\omega) = \omega^*$ can be recovered from the output, the function is invertible, and in fact is self-inverting.

Finally note that if $\omega \in E_2$, by definition and hence the first entry of ω is A, the first entry of ω^* is B. Hence for $\omega \in E_2$, $\omega^* \in E_1$. This shows that the image of E_2 under the function $f(\omega) = \omega^*$ lies in E_1.

On the other hand, if $\omega \in E_1$, then the sequence starts with B in the lead, and since A wins in the end, there must be at least one tie before the end, so ω^* is well defined, and ω^* starts with A and has a tie. Hence $\omega^* \in E_2$. Since $\omega = (\omega^*)^*$, and $\omega^* \in E_2$, there is an outcome ω' in E_2, namely $\omega' = \omega^*$, such that $f(\omega'') = \omega$. This shows that the image of E_2 under the function $f(\omega) = \omega^*$ is *exactly* E_1.

This, we have in invertible function f that maps E_2 *onto* E_1, and it sets up a one-to one correspondence between outcomes in E_2 and E_1. Therefore, $\#(E_2) = \#(E_1)$. This completes the justification of (0.4).

0.2 Catalan numbers

Next, let us consider the case in which candidates A and B receive the same number of votes, N. The sample space S consists of all sequences of $2N$ letters, N of them being A, and N of them being B. Any particular outcome ω is determined once we have chosen where to put the N votes for A (or the N votes for B), so that the cardinality of S is $\binom{2N}{N}$. If $E \subset S$, we put

$$P(E) = \#(E)\binom{2N}{N}^{-1}\ , \tag{0.7}$$

again assuming all outcomes are equally likely.

We ask: *What is the probability that B never has more votes than A at any stage in the tally of the votes? What is the probability that A maintains a strict lead until the final vote is tallied?*

We first reformulate these questions in terms of events. Define the event $E \subset S$ to consist of all outcomes in which for each $j = 1, \ldots, 2N-1$, there are never *more* votes for B than A to the left of j. Define the event $F \subset S$ to consist of all outcomes in which for each $j = 1, \ldots, 2N-1$, there are always *fewer* votes for B than A to the left of j. The answer to the first question is $P(E)$, and the answer to the second question is $P(F)$.

Notice that $F \subset E$, so one thing we can say right away is that $P(F) \leq P(E)$. In fact, F is *strictly* contained in E, and every outcome has positive probability, so that $P(F) < P(E)$. It is less clear by how much, or what either probability is.

We answer the first question by considering the complement E^c of E; i.e., the set of outcomes in S that do not belong to E. Obviously, $S = E \cup E^c$ and $E \cap E^c = \emptyset$, and hence

$$1 = P(S) = P(E) + P(E^c) \ . \tag{0.8}$$

Hence, to answer the first question, it suffices to compute $\#(E^c)$. We observe that $\omega \in E^c$ if and only if B is in the lead at some point, necessarily before the end, when there must be a tie. And if B is ever in the lead, there is a first time that B is ahead by one vote. Hence if $\omega \in E^c$, there is some $j \in \{0, 1, \dots, N-1\}$ so that when $2j+1$ votes are tallied, B has $j+1$ votes, and A has j votes, and B has not had the lead earlier.

At this point there remain in the ballot box $2N - 2j - 1$ ballots, and $j-1$ are for B, and j are for A. Now define ω^* to be the new sequence obtained by swapping the votes for A and B on the *remaining ballots only.* The ballot box now contains j remaining votes for B, and $j-1$ remaining votes for A. At the end of the tally, there are $(j+1) + j = N+1$ votes for B, and $j + (j-1) = N-1$ votes for A. That is, ω^* is a sequence of votes with which B wins by 2 votes.

Define a set G of sequences to be those that are $2N$ terms long, with $N+1$ terms being B, and $N-1$ terms being A. Note that G is *not* a subset of S; every outcome in S is a tie, and with every outcome in G, B wins by 2 votes. However, we have just seen that the function $f(\omega) = \omega^*$, using this new definition of ω^*, maps the set E^c into the set G.

Moreover, this function is one-to-one: Note that for the sequence ω^*, B again takes the lead for the first time at the exact same point as B did form ω, since the transformation does not alter the votes until *after* B has taken the lead. Thus, we may apply the same transformation to ω^*, swapping the votes after B has taken the lead for the first time, and we have that $\omega^{**} = \omega$. This proves that we can recover ω from ω^*, and hence that the function $f(\omega) = \omega^*$ is a one-to-one map of E^c into G.

We now show that the function $f(\omega) = \omega^*$ maps E onto G, this setting up a one-to-one correspondence between outcomes in E^c and outcomes in G.

Consider any outcome $\omega \in G$. Since for ω, B wins by 2 votes, there is some point *strictly* before the end when B is ahead by one vote. Therefore, ω^* is a well defined sequence. We claim that $\omega^* \in E$.

To see this, suppose B *first* takes the lead by 1 vote when $2j+1$ votes are counted, and then necessarily $j \in \{0, \dots, N-1\}$. To obtain ω^*, we now swap all of the votes after this point. Then, when $2j+1$ votes have been counted, before the swap, there are

$$(N+1) - (j+1) = N - j$$

votes for B remaining in the ballot box, and there are

$$(N-1) - j = N - j - 1$$

votes for B remaining in the ballot box. After the swap there are $N-j$ remaining votes for A, and $N-j-1$ remaining votes for B. After the swap, the total for A is $j+N-j=N$, and the total for B is then N as well. Hence $\omega^* \in S$, and moreover, $\omega^* \in E^c$ since B still takes the lead by 1 when $2j+1$ votes are counted. Since $\omega = (\omega^*)^*$, and $\omega^* \in E^c$, this proves that the function maps E^c *onto* G, and together with the fsct that is is one-to-one, that $\#(E^c) = \#(G)$.

Now everything is easy. Any sequence in G is determined by choosing where the $N+1$ votes for B occur, or, equivalently, choosing where the $N-1$ votes for A occur. Therefore

$$\#(E^c) = \#(G) = \binom{2N}{N+1} = \binom{2N}{N+1} .$$

Note that

$$\binom{2N}{N+1} = \frac{(2N)I}{(N+1)!(N-1)!} = \frac{N}{N+1}\frac{(2N)I}{(N!)^2} ,$$

An therefore, by (0.7),

$$P(E^c) = \frac{N}{N+1} .$$

Then by (0.8),

$$P(E) = \frac{1}{N+1} .$$

This answers the first question, and also shows that $\#(E) = \frac{1}{N+1}\binom{2N}{N}$. These numbers arise in the answers to many problems, and they have a name:

0.1 DEFINITION (Catalan numbers)**.** For $N \in \mathbb{N}$, the natural numbers,

$$C_N := \frac{1}{N+1}\binom{2N}{N} . \tag{0.9}$$

For $N = 0$, we define $C_0 = 1$.

To answer the second question, suppose $\omega \in F$. Then necessarily the first vote is for A, and B only catches up with the very last vote, so the sequence begins with an A and ends with a B. If we delete the first and the last votes, we get a sequence of $2(N-1)$ votes with a tie. What kind of sequence do we get? If we just tally those votes, B never takes the lead, or else there would be a tie in the original sequence. We know how many sequences of $2(N-1)$ votes, $N-1$ for A and B each, in which B never takes the lead. We have computed it to be C_{N-1}. Therefore, with F denoting the event corresponding to the second question,

$$\#(F) = C_{N-1} = \frac{1}{N}\binom{2N-2}{N-1} = \frac{1}{N}\frac{N^2}{2N(2N-1)}\binom{2N}{N} = \frac{1}{4N-2}\binom{2N}{N} .$$

Then by (0.7) again,

$$P(F) = \frac{1}{4N-2} .$$

The Catalan numbers satisfy a useful recurrence relation. To see this, note that any outcome in our the even E, defined as above, has a tie when j votes are tallied for some $j \in \{1, \ldots, N\}$. It may come only at the very end, and there may be more than one tie, but there is always a *first* tie.

For each $j = 1, \ldots, N$ define the event E_j to consist of those outcomes for which the first vote is for A, and the first tie occurs when $2j$ votes are tallied, and B never takes the lead. Since the first tie can only occur in one place, the different $E_j's$ are mutually disjoint, and by what we have noted above

$$E = \cup_{j=1}^{N} E_j \ .$$

Therefore

$$C_N = \#(E) = \sum_{j=1}^{N} \#(E_j) \ . \tag{0.10}$$

Now, what sort of outcomes belong to E_j? Up to the point that $2j$ votes are drawn, A has held the lead, but then there is a tie. By our answer to the second question above (and our definition for C_0), there are C_{j-1} such sequences of $2j$ terms. Continuing from here, there are $N - j$ votes for A and B each, and for this part of the sequence, B must never take the lead, though a tie is allowed. By the answer to the first question above, there are C_{N-j} such sequences. Altogether, there are $C_{j-1}C_{N-j}$ outcomes in E_j. that is, $\#(E_j) = C_{j-1}C_{N-j}$, and then (0.10) gives us

$$C_N = \sum_{j=1}^{N} C_{j-1}C_{N-j} \ ,$$

It is customary to write this with N replaced by $N + 1$, and $j = k + 1$ so k ranges from 0 to N, and we have

$$C_{N+1} = \sum_{k=0}^{N} C_k C_{N-k} \ . \tag{0.11}$$

Given that $C_0 = 1$, this recursion formula determines C_N for all N. For example $C_1 = C_0^2 = 1$. Then $C_2 = C_0C_1 + C_1C_0 = 2$, and so forth. One readily computes $C_3 = 5$, $C_4 = 14$, $C_5 = 42$, and so forth.

We have seen that the numbers C_N define in (0.9), together with $C_0 = 1$, do satisfy (0.9), and we have just seen that (0.9), together with $C_0 = 1$, specify a unique sequence of numbers: the Catalan sequence.

Hence if we run into any problem for counting the numbers of elements in a sequence of sets A_N, $N = 1, 2 \ldots$, and if we can determine that

$$\#(A_{N+1}) = \sum_{k=0}^{N} \#(A_k)\#(A_{N-k}) \ ,$$

with the convention that $\#(A_0) = 1$, then we know that

$$\#(A_N) = C_N \ .$$

An example of this occurs with random binary graphs, as shown in class, and there are many, many other examples, some of which we will meet later.

0.3 Runs

Another sort of problem that comes of frequently is natural to introduce in the context of balloting. Suppose 14 votes are cast, 8 are cast for A, and 6 for B. One possible sequence that could occur in the tally of votes is

$$AABAAABBABBAAB$$

In this sequences, there are 4 runs of A's, of length 2, 3, 1 and 2, left to right. There are also 4 runs of B's, of lengths 1, 2, 2, and 1, left to right. A *run* is just an uninterrupted block of votes for the same candidate that is not a proper subset of any other uninterrupted block of votes for the same candidate. More generally, for a sequence of m types of entries, a run is an uninterrupted subsequence of terms that is not a proper subsequence of any other uninterrupted subsequence of like terms.

If A receives N votes and B receives M votes, with $N > M$, there can be no more than $M+1$ runs of votes for A since each such run must be separated from the next by at least one vote for B in the middle. So 2 runs for A must be separated by 1 vote for B, 3 runs for A must be separated by 2 votes for B, and so forth. At the other extreme, the minimill number of runs for A is 1: Simply put all of the A votes together.

Now consider any natural number R with $1 \le R \le M+1$. How many of the $\binom{N+M}{N}$ balloting sequences have exactly R runs of votes for A? To answer this, for each $j = 1, \dots R$, let n_j denote then number of votes for A in the jth run. Then each $n_j \ge 1$, and

$$\sum_{j=1}^{R} n_j = N \ . \tag{0.12}$$

0.2 PROPOSITION. *For any natural number R and any integer $N \ge R$, there are exactly $\binom{N-1}{R-1}$ R-tuples of natural number $(n_1, \dots, n_R)$ satisfying* (0.12)

Proof. Consider the case $N = 8$ and $R = 4$. To do the counting, arrange 8 dots in a row:

$$\bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet$$

There are 7 spaces between the dots. Choose 3 of the 7, and put vertical bars in those places:

$$\bullet \quad \bullet \ | \ \bullet \ | \ \bullet \quad \bullet \quad \bullet \ | \ \bullet \quad \bullet$$

Let n_1 be the number of does to the left of the first bar, n_2 the number between the first and second, n_3 the number between the second and third, and n_4 the number after the third. This yield a sequence (n_1, n_2, n_3, n_4) that satisfies $n_1 + n_2 + n_3 + n_4 = 8$.

Conversely, given any (n_1, n_2, n_3, n_4) that satisfies $n_1 + n_2 + n_3 + n_4 = 8$, put n_1 dots in a row, followed by a bar. Then put in n_2 dots followed by a bard. Then put in n_3 dots followed by a bar. Finally, put in n_4 dots. The result is a patter of the type introduced above, and clearly there is a one-to-one correspondence between the set of possible (n_1, n_2, n_3, n_4), and the set of dot and bar diagrams. The latter is easy to count: There are $\binom{7}{3}$ way to choose 3 place to insert bars in the 7 spaces between the 8 dots. On a little reflection, the general case will now be clear. □

Now let us apply this to the balloting problem, again with N votes for A, M votes for B, and for $R \leq M+1$, and of course, $R \leq N$. We wish to count the sequences that have exactly R runs of votes for A. We do not need the assumption that $N > M$ at this point, but we do assume $M, N \geq 1$.

By Proposition 0.2, there are $\binom{N-1}{R-1}$ ways to assign the numbers $n_1, \ldots n_R$ of votes for A to each of the R runs. It remains to count the number of ways to put the votes for B into the sequence. There are $M+1$ places to put them: Into the $R-1$ places between the runs of votes for A, and at the beginning and the end of the sequence. Let $m_1, \ldots, m_{R+1}$ denote then number of B votes assigned to each of these places. We must have $m_j \geq 1$ when $2 \leq j \leq R$ in order to separate the runs, but we could have $m_1 = 0$ or $m_{R+1} = 0$. In order to apply Proposition 0.2, define a new sequence $(q_1, \ldots, q_{R+1}))$ by $q_j = m_j$ for $2 \leq j \leq R$, $q_1 = m_1 + 1$, and $q_{R+1} = m_{R+1}$. Then $(q_1, \ldots, q_{R+1})$ is a sequence of natural numbers satisfying $\sum_{j=1}^{R+1} q_j = M+2$, and we can recover the original sequence $(m_1, \ldots m_{R+1})$ from $(q_1, \ldots, q_{R+1})$ by subtracting 1 both q_1 and q_{R+1}. Hence the number of ways to place the B terms in the sequence is the number of ways there are to choose $R+1$ natural numbers that sum to $M+2$. By Proposition 0.2, there are $\binom{M+1}{R}$ such sequences.

Hence the number of balloting sequence with exactly R runs of votes for A is

$$\binom{N-1}{R-1}\binom{M+1}{R} .$$

Then if E_R denotes the even that there are exactly R runs of votes for A,

$$P(E_R) = \frac{\binom{N-1}{R-1}\binom{M+1}{R}}{\binom{N+M}{N}} .$$

Notice the events E_R are disjoint, If we now suppose $N > M$, then the condition $R < M+1$ implies the condition $R \leq N$, and so we need only suppose that $R \leq M+1$. The possible values of R are then $1, \ldots, M+1$, and $\cup_{R=1}^{M+1} E_R$ is the whole sample space S of possible sequences. Therefore

$$\sum_{R=1}^{M+1} \frac{\binom{N-1}{R-1}\binom{M+1}{R}}{\binom{N+M}{N}} = 1 ,$$

which yields the identity

$$\sum_{R=1}^{M+1} \binom{N-1}{R-1}\binom{M+1}{R} = \binom{N+M}{N} .$$

We are often, as in this example, interested in counting sequences of *non-negative* integers $(n_1, \ldots, n_R)$ that sum to some specified natural number N. To do so, simply add 1 to each of the terms, yielding a sequence of natural numbers $(q_1, \ldots, q_R)$ that sums to $N+R$. The original sequence may be recovered from this one, so there is a one-to one correspondence between the two classes of sequences, and we know how to count the one consisting of natural numbers. Thus we have:

0.3 PROPOSITION. *For any natural numbers R, N, there are exactly $\binom{N+R-1}{R-1}$ R-tuples $(n_1, \ldots, n_R)$ of non-negative integers satisfying*

$$\sum_{j=1}^{R} n_j = N \ .$$