Topics in
Probability Theory and Stochastic Processes
Steven R. Dunbar

The Ballot Theorem and the Reflection Principle

Rating
Mathematicians Only: prolonged scenes of intense rigor.
**Section Starter Question**

Suppose that in an election candidate A receives 4 votes and candidate B receives 3 votes. Explicitly enumerate all the ways that the votes can be ordered. In how many ways does A maintain more votes than B throughout the counting of the ballots?

---

**Key Concepts**

1. Suppose that in an election, candidate A receives $a$ votes and candidate B receives $b$ votes, where $a \geq kb$ for some positive integer $k$. The number of ways the ballots can be ordered so that A maintains more than $k$ times as many votes as B throughout the counting of the ballots is

   $$\frac{a - kb}{a + b} \binom{a + b}{a}.$$

2. The Reflection Method is a popular proof of the solution of the Ballot Problem because it is relatively easy, visual and it has generalizations beyond ballot counting to random walks and Brownian Motion.

3. The Cycle Lemma proves that for any ballot sequence of $a$ votes for A and $b$ votes for B, exactly $a - kb$ of the $a + b$ cyclic permutations of the sequence are good. Consequently, a fraction $(a - kb)/(a + b)$ of all ballot permutations are good.
Vocabulary

1. **Bertrand’s Ballot Problem** is the following combinatorial problem: Suppose that in an election, candidate A receives $a$ votes and candidate B receives $b$ votes, where $a \geq kb$ for some positive integer $k$. In how many ways can the ballots be ordered so that A maintains more than $k$ times as many votes as B throughout the counting of the ballots?

2. The **Reflection Principle** is that the set of paths from $(1,1)$ to $(a+b,a-b)$ that do touch the $x$-axis somewhere is in one-to-one correspondence with the set of all paths from $(1,-1)$ to $(a+b,a-b)$.

3. A partition of a set into subsets of equal size is called a **uniform partition**.

---

Mathematical Ideas

This section introduces the Ballot Problem which is a purely combinatorial problem. The solution to the problem results in the Ballot Theorem. This section has several proofs of this combinatorial result, including the Reflection Principle. Later sections have probabilistic interpretations of Bertrand’s Ballot Theorem for positive lattice paths, and hitting time results for a Bernoulli random walk. This interpretation is equivalent to the Reflection Principle for Bernoulli random walks. In later sections the Reflection Principle is a tool in the proof of the Arcsine Law for binomial random variables.

Introduction and History

*Definition. Bertrand’s Ballot Problem* is the following combinatorial problem: Suppose that in an election, candidate A receives $a$ votes and candidate
B receives $b$ votes, where $a \geq kb$ for some positive integer $k$. In how many ways can the ballots be ordered so that A maintains more than $k$ times as many votes as B throughout the counting of the ballots?

**Definition.** More strictly, the Ballot Theorem when $k = 1$ is called the **strict Ballot Problem** and for $k \geq 1$ called the **generalized Ballot Problem**.

The solution of both problems is:

**Theorem 1** (Ballot Theorem). Suppose that in an election, candidate A receives $a$ votes and candidate B receives $b$ votes, where $a \geq kb$ for some positive integer $k$. The number of ways the ballots can be ordered so that A maintains more than $k$ times as many votes as B throughout the counting of the ballots is

$$\frac{a - kb}{a + b} \binom{a + b}{a}.$$


According to Renault [18] many sources cite André [4] claiming that he used the “Reflection Principle” to solve Bertrand’s Ballot Problem. It is not strictly accurate to say that André employed the Reflection Method in his proof. The “Reflection Principle” is a variation of André’s method of proof. Bertrand’s textbook on probability [6], first published in 1888, solves the Ballot Problem exactly as André did. Comtet [7] formulated the “Reflection Principle”, stating that it is essentially “due to André”. In the following years, many other minor variations appeared (see [9] for references).
In 1923 Aebly [1] conceived of ballot permutations as paths starting from a corner on a rectangular grid, and he observes a certain symmetry of bad paths across the diagonal (compare [24] pages 172–175). Mirimanoff [16] (the very next article in the same journal) builds on Aebly’s observations and relates this symmetry to ballot permutations of A’s and B’s by applying a transposition to an initial segment of a bad ballot permutation. Viewed geometrically, Mirimanoff’s method is exactly the reflection method we know today. When did authors start claiming that André actually used the reflection method in his solution? This is more difficult to ascertain. In 1947 Dvoretzky and Motzkin [9] introduced a new method of proof to the Ballot Problem and its generalization. They accurately describe André’s contribution, making no mention of the Reflection Method, and write “André’s proof or variations of it may be found in most of the classical treatises on the theory of probability.” In 1957, citing this paper of Dvoretzky and Motzkin, Feller [11, page 66] not quite accurately writes “As these authors point out, most of the formally different proofs in reality use the reflection principle . . .” suggesting, perhaps, that the reflection method was the original method of solution. Feller, four pages later, states the reflection principle and makes the footnote, “The probability literature attributes this method to D. André, (1887).” The first edition of Feller’s book (1950) mentions neither André nor the Reflection Principle. The earliest source that Renault could find linking André and the Reflection Method is J. L. Doob [8, page 393], who writes while describing Brownian motion processes, “...similar exact evaluations are easily made ... using what is known as the reflection principle of Désiré André.” These early sources do not explicitly state that André used the reflection method in his own proof, but this could be inferred, and other writers since then have naturally assumed that André did indeed use the reflection method.

**Proof by Induction**

*Proof (Generalized Ballot Theorem by Induction).* Let $N_k(a, b)$ denote the number of ways the $a + b$ ballots ($a \geq kb$) can be ordered so that candidate A maintains more than $k$ times as many votes as B throughout the counting of the ballots. The conditions $N_k(a, 0) = 1$ for all $a > 0$ and $N_k(kb, b) = 0$ for all $b > 0$ are easily verified by considering the statement of the ballot
problem, and they both satisfy
\[
N_k(a, b) = \frac{a - kb}{a + b} \left( \frac{a + b}{a} \right).
\]

For \( b > 0 \) and \( a > kb \), note that \( N_k(a, b) = N_k(a, b - 1) + N_k(a - 1, b) \) by considering the last vote in a ballot permutation. By induction, this quantity is
\[
\frac{a - k(b - 1)}{a + b - 1} \left( \frac{a + b - 1}{a} \right) + \frac{a - 1 - kb}{a + b - 1} \left( \frac{a + b - 1}{a - 1} \right)
\]
which simplifies to
\[
\frac{a - kb}{a + b} \left( \frac{a + b}{a} \right)
\]

Remark. This proof is easy and short, but it gives no geometric intuition or indication of the symmetries in the Ballot Problem.

Remark. See also [24, pages 182-184] for a detailed induction proof.

The Reflection Principle

The Reflection Principle is a popular proof because it is easy, visual, and it has generalizations beyond ballot counting to random walks and Brownian Motion. Feller [11] gives an account of the Reflection Principle.

Visualize ballot permutations as lattice paths in the Euclidean plane with votes for A as upsteps with vector \((1, 1)\) and votes for B as downsteps with vector \((1, -1)\). See Figure 1. Call ballot permutations (or paths) that satisfy the ballot problem “good”, and call those that do not “bad.”

Definition. The Reflection Principle is that the set of paths from \((1, 1)\) to \((a + b, a - b)\) that do touch the x-axis somewhere is in one-to-one correspondence with the set of all paths from \((1, -1)\) to \((a + b, a - b)\).

Proof (Strict Ballot Theorem by the Reflection Principle). The terminal point of the path is \((a + b, a - b)\). Since every good path must start with an upstep, there are as many good paths as there are paths from \((1, 1)\) to \((a + b, a - b)\) that never touch the x-axis.

The claim is that the set of paths from \((1, 1)\) to \((a + b, a - b)\) that do touch the x-axis somewhere is in one-to-one correspondence with the set of
all paths from \((1, -1)\) to \((a + b, a - b)\). The claim is established by reflecting across the \(x\)-axis the initial segment of the path that ends with the step that first touches the \(x\)-axis. Subtracting the number of these paths from the number of all paths from \((1, 1)\) to \((a + b, a - b)\) produces the number of good paths:

\[
\left( \frac{a + b - 1}{a - 1} \right) - 2 \left( \frac{a + b - 1}{b - 1} \right) = \frac{a - b}{a + b} \left( \frac{a + b}{a} \right)
\]

\(\square\)

**André’s Original Proof**

André was the first to solve the ballot problem by subtracting the number of bad ballot permutations from the number of all possible ballot permutations. This is the approach taken by the Reflection Principle. However, whereas the Reflection Principle modifies an initial segment of a lattice path (equivalently,
a ballot permutation), André uses no geometric reasoning and he interchanges two portions of a ballot permutation.

André supposes that $a$ ballots are marked “A” and $b$ ballots are marked “B”. He first notes that every ballot permutation starting with a “B” is bad and there are $\binom{a+b-1}{a}$ of these.

**Proof (André’s Original Proof of the Strict Ballot Theorem).** This proof is a paraphrase of a translation of André’s original proof, adapted from Renault [18].

“The number of possible outcomes is obviously the number of permutations one can form with $a$ letters A and $b$ letters B. Let $Q(a, b)$ be the number of unfavorable outcomes. The permutations corresponding to them are of two kinds: those that start with B, and those that start with A. The number of unfavorable permutations starting with B equals the number of all permutations which one can form with $a$ letters A and $b-1$ letters B, because it is obviously enough to suppress the initial letter B to obtain the remaining letters. The number of unfavorable permutations starting with A is the same as above, because one can, by a simple rule, make a one-to-one correspondence with the permutations formed with $a$ letters A and $b-1$ letters B. This rule is composed of two parts:

1. Given an unfavorable permutation starting with A, one removes the first occurrence of B that violates the law of the problem [i.e., causes the number of B’s to equal the number of A’s], then one exchanges the two groups separated by this letter: one obtains thus a permutation, uniquely determined, of $a$ letters A and $b-1$ letters B. Consider, for example, the unfavorable permutation AABBABAA, of five letters A and three letters B; by removing the first B that violates the law, one separates two groups AAB, ABAA; by exchanging these groups, one obtains the permutation ABAAAB, formed of five letters A and two letters B. (See Figure 2.)

2. Given an arbitrary permutation of $a$ letters A and $b-1$ letters B, one traverses it from right to left until one obtains a group where the number of A’s exceeds [by one] the number of B’s; one considers this group and that which the letters
Figure 2: Example of André’s one-to-one correspondence with AABBABAA
placed at its left form; one exchanges these two groups, while placing between them a letter B: one thus forms an unfavorable permutation starting with A that is uniquely given. Consider, for example, the permutation ABAAAAAB; while operating as described, one divides it in two groups ABAA, AAB; by exchanging these groups and placing the letter B between them, one forms the unfavorable permutation AAB-BABAA.

It results from all the above that the total number of unfavorable outcomes is twice the number of permutations one can form with \( a \) letters A and \( b - 1 \) letters B."

Thus André establishes a one-to-one correspondence between the bad ballot permutations starting with “A” and all permutations consisting of \( a \) “A’s” and \( b - 1 \) “B’s”. There are \( \binom{a+b-1}{a} \) of these. The number of bad ballot permutations is \( 2\binom{a+b-1}{a} \). The ballot theorem then follows by simplifying

\[
\binom{a+b}{a} - 2\binom{a+b-1}{a} = \frac{a-b}{a+b} \binom{a+b}{a}.
\]

Example. Consider an example with 6 votes, 4 for A and 2 for B. There are \( \binom{6}{4} = 15 \) possible vote count patterns, of which \( 2\binom{5}{4} = 10 \) are bad. There are \( \binom{5}{4} = 5 \) unfavorable permutations starting with B.
<table>
<thead>
<tr>
<th>Ballots</th>
<th>Type</th>
<th>Start</th>
<th>Partition</th>
<th>Rearrangement</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAAABB</td>
<td>Good</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAABAB</td>
<td>Good</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AABAAB</td>
<td>Good</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABAAAB</td>
<td>Bad</td>
<td>Starts with A</td>
<td>A B AAAB</td>
<td>AAABA</td>
</tr>
<tr>
<td>BAAAAB</td>
<td>Bad</td>
<td>Suppress initial B</td>
<td>AA AAB</td>
<td>AABBA</td>
</tr>
<tr>
<td>AAABBA</td>
<td>Good</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AABABA</td>
<td>Good</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABAABA</td>
<td>Bad</td>
<td>Starts with A</td>
<td>A B AABA</td>
<td>AABAA</td>
</tr>
<tr>
<td>BAAABA</td>
<td>Bad</td>
<td>Suppress initial B</td>
<td>AA ABA</td>
<td>ABABAA</td>
</tr>
<tr>
<td>AABBA</td>
<td>Bad</td>
<td>Starts with A</td>
<td>A B B AA</td>
<td>AAAAB</td>
</tr>
<tr>
<td>ABABAA</td>
<td>Bad</td>
<td>Starts with A</td>
<td>A B ABAA</td>
<td>ABAAA</td>
</tr>
<tr>
<td>BABAA</td>
<td>Bad</td>
<td>Suppress initial B</td>
<td>AA BAA</td>
<td>BAABAA</td>
</tr>
<tr>
<td>ABAAA</td>
<td>Bad</td>
<td>Starts with A</td>
<td>A B BAAA</td>
<td>BAAA</td>
</tr>
<tr>
<td>BABAA</td>
<td>Bad</td>
<td>Starts with B</td>
<td>ABA AA</td>
<td>AABABA</td>
</tr>
<tr>
<td>BBAAA</td>
<td>Bad</td>
<td>Starts with B</td>
<td>BAA AA</td>
<td>AABBA</td>
</tr>
</tbody>
</table>

Proof for the Generalized Ballot Problem

This proof for the Generalized Ballot Problem uses the lattice path idea from the Reflection Principle proof. In the Generalized Ballot Problem, votes for A are upsteps $(1, 1)$, votes for B are downsteps $(1, -k)$. We wish to find the number of lattice paths starting at the origin and consisting of $a$ upsteps $(1, 1)$ and $b$ downsteps $(1, -k)$ such that no step ends on or below the $x$-axis. The difficulty in generalizing the Reflection Principle is that when one reflects such a lattice path, the result does not have the required types of upsteps and downsteps. Goulden and Serrano [12] also note the literature has many solutions to the generalized ballot problem but “there appears to be no solution which is in the spirit of the reflection principle.” They go on to provide a proof that rotates the initial path segment by $180^\circ$; even so, they acknowledge that their proof does not specialize to the Reflection Principle when $k = 1$. (As a side note, rotating a path is equivalent to writing a ballot permutation in reverse order. Uspensky [20] used this principle to solve the Ballot Problem in 1937.)

Proof (Generalized Ballot Theorem). For $0 \leq i \leq k$, let $B_i$ denote the set of bad paths whose first bad step ends $i$ units below the $x$-axis. Note that the paths in $B_k$ necessarily start with a downstep. Clearly these $k + 1$ sets are
disjoint and their union is the set of all bad paths. Let $\mathcal{A}$ be the set of all
paths consisting of $a$ upsteps and $b - 1$ downsteps, without regard to location
in the plane; $|\mathcal{A}| = \binom{a+b-1}{a}$. We prove that $|\mathcal{B}_i| = |\mathcal{A}|$ for each $i$ in the range $0 \leq i \leq k$ by providing a one-to-one correspondence between $\mathcal{B}_i$ and $\mathcal{A}$.

Given a path $P \in \mathcal{B}_i$ write $P = XDY$ where $D$ is the first downstep that
ends on or below the $x$-axis (note that $X$ is empty if $i = k$). The path $YX$
is then uniquely determined and is an element of $\mathcal{A}$. Given a path $Q \in \mathcal{A}$,
scan the path from right to left until a vertex is found lying $k - i$ units below
the terminal vertex of $Q$ (note that this vertex is the terminal vertex of $Q$
itself if $i = k$). Such a vertex must exist since the initial vertex of $Q$ lies
more than $k$ units below the terminal vertex of $Q$ because $a - kb \geq 0$ so
$a - k(b - 1) \geq k$. Write $Q = YX$, where $Y$ and $X$ are the paths joined
at that vertex. Construct path $P = XDY$ by interchanging $X$ and $Y$ and
inserting a downstep $D$ between them. Translating $P$ to start at $(0,0)$, the
path touches the $x$-axis only at the origin, and $XD$ ends $i$ units below the

Figure 3: Example with $k = 3$, $XDY \in \mathcal{B}_1$, and $YX \in \mathcal{A}$.
$x$-axis; hence $P \in B_i$. Therefore, the number of good paths is

\[
\binom{a+b}{a} - (k+1)\binom{a+b-1}{a} = \frac{a-kb}{a+b}\binom{a+b}{a}.
\]

\[\square\]

Proof (Alternative Proof to Generalized Ballot Theorem). An alternate formulation of the proof: For $0 \leq i \leq k$, let $B_i$ denote the set of bad paths whose first bad step ends $i$ units below the $x$-axis. Clearly these $k+1$ sets are disjoint and their union is the set of all bad paths. Notice that the paths in $B_k$ are exactly those paths that start with a downstep, and so $|B_k| = \binom{a+b-1}{a}$.

We now show that for any $i \neq k$ we actually have $|B_i| = |B_k|$. Let $P$ be a path in $B_i$, $(i \neq k)$, and identify the first step of $P$ that ends $i$ units below the $x$-axis. Let $X$ be the initial segment of $P$ that ends with that step and write $P = XY$. Let $\tilde{X}$ denote the path that results from rotating $X$ by $180^\circ$, exchanging its endpoints; see Figure ???. Since $X$ ends with a downstep, $\tilde{X}$ starts with a downstep, and consequently $\tilde{X}Y \in B_k$.

The same process converts a path in $B_k$ into a path in $B_i$, $(i \neq k)$. If $P \in B_k$, then identify the first step that ends $i$ units below the $x$-axis. Let $X$ denote the initial segment of $P$ that ends with that step and write $P = XY$. Since $X$ necessarily ends with an upstep, we have $\tilde{X}Y \in B_i$.

Therefore, the number of good paths is

\[
\binom{a+b}{a} - (k+1)\binom{a+b-1}{a} = \frac{a-kb}{a+b}\binom{a+b}{a}.
\]

\[\square\]

Proof by the Cycle Lemma

Proof (Generalized Ballot Theorem by the Cycle Lemma). A ballot permutation is a sequence of $a+b$ terms where each term is either 1 or $-k$; votes for A correspond to the 1’s and votes for B correspond to the $-k$’s. A sequence is called good if every partial sum is positive, and bad otherwise. Observe that the sum of a sequence is $a-kb \geq 0$. Let $C$ be any circular arrangement of $a$ 1’s and $b-k$’s. Now prove the Cycle Lemma: of the $a+b$ terms in $C$, exactly $a-kb$ start good sequences when $C$ is read once around clockwise.

The claim is now that there must exist a sequence $X = 1, 1, \ldots, 1, -k$ in $C$ with $k$ consecutive 1’s. Suppose not, that is, there is a placement of
-k’s around the circle with at most k - 1 1’s between each -k. Then the total number of 1’s around the circular arrangement is at most (k - 1)b. But (k - 1)b < kb ≤ a which is not possible, so the claim is true.

No term of X can start a good sequence, because the partial sum would be less than or equal to zero at the -k term.

Let C' be the circular arrangement created from C by removing X. Then C' has a' = a - k terms of 1’s and b' = b - 1 terms of -k's. Note that a' - kb' = a - k - k(b - 1) = a - kb ≥ 0 so C' satisfies the same hypotheses as C. Furthermore, since the sequence X has sum 0, it has no “net effect” on good sequences. Thus, a term of C' starts a good sequence if and only if the corresponding term in C starts a good sequence. Consequently, C' and C have exactly the same number of terms that start good sequences. See Figure 5. Continuing in this manner, one removes sequences of the form 1, 1, 1, . . . , 1, -k until a circular arrangement consisting only of 1’s remains. At this stage, there are a - kb 1’s, and every term starts a good sequence.
Figure 5: Example of $C'$ and $C''$ with $a = 8$, $b = 2$, $k = 3$. 
Hence there are exactly $a - kb$ good sequences in $C$, and the Cycle Lemma is proved. If there is periodicity in $C$ then not all $a - kb$ good sequences will be distinct. However, we can conclude that the ratio of good sequences to all sequences is $a - kb$ to $a + b$. Therefore, the number of good paths is

$$\frac{a - kb}{a + b} \left( \frac{a + b}{a} \right).$$

**Remark.** The Cycle Lemma provides a surprising result: for any ballot sequence of $a$ votes for A and $b$ votes for B, exactly $a - kb$ of the $a + b$ cyclic permutations of the sequence are good. Consequently, a fraction $(a - kb)/(a + b)$ of all ballot permutations are good.

**Proof by Uniform Partition**

**Remark.** As an overview, the following proof considers a set with $(a - kb)\binom{a + b}{a}$ elements, and partitions this set into $a + b$ subsets of equal size. This is called a uniform partition into $a + b$ subsets. One of the subsets corresponds to the set of good ballot permutations, and from this the ballot theorem follows. The proof is from [17] and extends the proofs in [23].

**Proof (Generalized Ballot Theorem by Uniform Partition).** Consider lattice paths starting from the origin and consisting of $a$ upsteps with vector $(1, 1)$ and $b$ downsteps with vector $(1, -k)$, and assume the strict inequality $a > kb$. Let $\mathcal{A} = \mathcal{A}(a, b, k)$ be the set of all such paths. Given path $P \in \mathcal{A}$ we let $L(P)$ denote the set of $x$-values of the $a - kb$ “rightmost lowest” vertices of $P$; see Figure 6. More precisely, given path $P \in \mathcal{A}$, let $y_0$ denote the least $y$-value of all the vertices of $P$, and let $r(t)$ denote the $x$-value of the rightmost vertex of $P$ along the line $y = t$; then $L(P) = \{r(t) | t \in \mathbb{Z}, y_0 \leq t \leq y_0 + (a - kb) - 1\}$.

Let $\Psi = \{(P, j) | P \in \mathcal{A}, j \in L(P)\}$ and note that $|\Psi| = (a - kb)|\mathcal{A}|$. Let $\Omega_i = \{(P, i) \in \Psi | i \in L(P)\}$, defined for $0 \leq i \leq a + b - 1$. The sets $\Omega_i$ partition $\Psi$ into $a + b$ disjoint subsets.

**Claim 1** There is a one-to-one correspondence between $\Omega_0$ and the set of good paths. If $P \in \mathcal{A}$ is good, then $(0, 0)$ is the lowest vertex in $P$ and it is the only vertex on the $x$-axis, so $(P, 0) \in \Omega_0$. Conversely, if $(P, 0)$ is in $\Omega_0$, then no vertex of $P$ can lie on the $x$-axis to the right of the origin, and so $P$ is good.
Figure 6: Example of uniform partition with $k = 2$, $a = 6$, $b = 2$. For each path $P \in \mathcal{A}$, the path $P$ and the set $L(P)$. Observe that among all the sets $L(P)$, each number from 0 to 7 occurs exactly 7 times.
Claim 2 The sets \( \Omega_i \) uniformly partition \( \Psi \). We show this by providing a one-to-one correspondence between \( \Omega_i \) and \( \Omega_0 \). If \((P, i) \in \Omega_i\), then write \( P = XY \) where \( X \) is the initial path of \( P \) consisting of the first \( i \) steps, and \( Y \) consists of the remaining steps. Since \( i \in L(P) \), we can observe that \( Y \) stays above the height of its initial vertex, and \( X \) never descends \( a - kb \) or more units below the height of its terminal vertex. Consequently the path \( YX \) is good and \((YX, 0) \in \Omega_0\). Conversely, if \((Q, 0) \in \Omega_0\), then write \( Q = YX \) where \( X \) consists of the final \( i \) steps of \( Q \). The same qualities of \( X \) and \( Y \) hold as noted above, and the pair \((XY, i) \in \Omega_i\).

The two claims above imply that the number of good paths in \( A \) is
\[
|\Omega_0| = \left| \Psi \right| = \frac{(a - kb)|A|}{a + b} = \frac{a - kb}{a + b} \left( \frac{a + b}{a} \right)
\]

Remark. Suppose we let \( A_i \) denote the set of paths for which \( i \) is among the \( x \)-values of the \( a - kb \) rightmost vertices, i.e., \( A_i = \{P \in A|i \in L(P)\} \). When \( a - kb = 1 \), the sets \( A_0, A_1, \ldots, A_{a+b-1} \) are disjoint and all have the same cardinality. In other words, partitioning \( A \) according to the \( x \)-value of a path’s rightmost lowest vertex creates a uniform partition of \( A \). When \( a - kb \geq 1 \), the sets \( A_i \) continue to have the same cardinality. However the \( A_i \) are no longer disjoint. To the contrary, each path in \( A \) will be a member of precisely \( a - ykb \) of these sets.

Proof by Rotation and Shadowing

Remark. The following proof is adapted from the “third proof” of Yaglom and Yaglom, [24, pages 172-175]. Their proof uses an “up-or-right” walk on an integer lattice to represent the ballot count instead of the “up-and-down” random walk orientation used in Figures 1, 2, 3, ?? and 6. The adaptation modifies their proof to this “up-and-down” random walk representation and shortens the explanation.

Rotation and Shadowing. Consider any particular arrangement of the \( a + b \) ballots in a list. From this list we can obtain \( a + b - 1 \) new lists as follows: move the first ballot from the front of the list to the last place, thus creating
a new arrangement in which the second, third, ..., and \(a + b\)th ballot of
the original list are now one place forward. Then repeat the process. By
repeating the process, we obtain a total of \(a + b - 1\) new lists, adding to this
the original list, we have a total of \(a + b\) arrangements. We will show that
\(a - b\) of these arrangements are good. Then the total number of arrangements is

\[
\frac{a - b}{a + b} \left( \frac{a + b}{a} \right).
\]

To represent all \(a + b\) rotated arrangements obtained from a particular
arrangement, adjoin to the end \((a + b, A_{a+b})\) of the path a replica of the
particular path. See Figure 7. For convenience label the vertices of the ballot
count path and its appended copy as \(A_0, A_1, A_2, \ldots, A_{a+b}, A_{a+b+1}, \ldots, A_{2a+2b}\).
Now the \(a + b\) rotated arrangements correspond to the paths which consist of
the subpaths that start respectively at \(A_0, A_1, A_2, \ldots, A_{a+b-1}\) and consequently end at the points \(A_{a+b}, A_{a+b+1}, \ldots, A_{2a+2b}\). Good arrangements correspond
to paths of length \(a + b\) in which every vertex other than the first is preceded
by more upsteps with vector \((1, 1)\) than downsteps with vector \((1, -1)\), that is, paths which lie above the horizontal. Now illuminate the extended path
from the right with a beam of parallel rays. Then the good arrangements
correspond to paths in which the light falls upon the base point \(A_k\). That is,
the base point does not lie in the shadow thrown by some downsteps. (Note
that if the \(A_k\) does not lie in the shade of the path \(A_k A_{k+1} A_{k+2} \ldots A_{k+a+b}\),
then \(A_k\) cannot lie in the shade thrown by the steps from \(A_{a+b+k}\) to \(A_{2a+2b}\)
either. This follows from the fact that the path \(A_{a+b+k}\) to \(A_{2a+2b}\) is a duplicate
of \(A_k A_{k+1} \ldots A_{a+b}\). Thus we only have to count how many of the points
\(A_0, A_1, A_2, \ldots, A_{a+b}\) are illuminated by the parallel rays.

Of the points, \(A_0, A_1, A_2, \ldots, A_{a+b-1}\) the only ones which have a chance
of being illuminated are the \(a\) points which correspond to an upstep with
vector \((1, 1)\). But even such a point may lie in the shade cast by a later
upstep. For example, in Figure 7, \(A_5\) and \(A_7\) lie in the shade cast by \(A_9\).

Denote by \(d\) the number of downsteps of \(A_0, A_1, A_2, \ldots, A_{a+b}\) which cast
their shadows to the left, i.e. segments like \(A_0 A_1\) and \(A_1 A_2\) in Figure 7.
Then the other \(b - d\) downsteps cast their shadows on upsteps. Since \(A_{a+b},
A_{a+b+1}, A_{a+b+2}, \ldots, A_{2a+2b}\) is a replica of \(A_0, A_1, A_2, \ldots, A_{a+b-1}\), exactly \(d\)
downsteps of \(A_{a+b}, A_{a+b+1}, A_{a+b+2}, \ldots, A_{2a+2b}\) cast their shadows to the
left of \(A_{a+b}\). These shadows fall on upsteps to the left of \(A_{a+b}\). These
shadows fall on upsteps of \(A_0, A_1, A_2, \ldots, A_{a+b}\); hence there are altogether
\(b - d + d = b\) upsteps in the shade. Thus, of the \(a\) points \(A_i\) which have a
Figure 7: Example of the rotation and shadowing proof with $a = 7$ and $b = 4$. A copy of the path is adjoined to the original path.
chance of being illuminated, exactly \( b \) are removed by virtue of the shadow cast on the segment \( A_i A_{i+1} \). Hence there remain \( a - b \) illuminated points of \( A_0, A_1, A_2, \ldots, A_{a+b-1} \). For example, in Figure 7 \( a = 7, b = 4 \) and exactly \( 7 - 4 = 3 \) are illuminated. This proves that for \( a > b \) exactly \( a - b \) of the \( a + b \) arrangements of ballot counts obtained from any one arrangement by rotation satisfy the requirements of the Ballot Theorem.

Note that the \( a + b \) arrangements obtained from a single one by rotations are not necessarily all distinct. If \( a \) and \( b \) are not relatively prime, then it may happen that an arrangement may consist of parts which are repetitions. For example, with \( a = 6 \) and \( b = 3 \), an arrangement may be \( ABAABAABA \). In such a case, the rotations will result in arrangements which are repeated the same number of times. In the example, among the 9 arrangements from rotations, there are 3 arrangements, each repeated 3 times.

In any case, the ratio of the number of good arrangements to the total number of arrangements is

\[
\frac{a-b}{a+b}.
\]

Therefore, the number of possible good arrangements is

\[
\frac{a-b}{a+b} \binom{a+b}{a}.
\]

\[\square\]

Remark. The proof by rotation and shadowing is another proof of the Cycle Lemma.

Remark. This rotation and shadowing proof can be modified for the generalized ballot theorem. The parallel rays are not horizontal, but come at an angle to the axis. See [24, pages 179-182].

Weak Ballot Problem and Catalan Numbers

The ballot problem is often stated in a “weak” version: suppose that candidate A receives \( m \) votes and candidate B receives \( n \) votes, where \( m \geq kn \) for some positive integer \( k \), and compute the number of ways the ballots can be ordered so that A always has at least \( k \) times as many votes as B throughout the counting of the ballots.

Any ballot permutation in which A maintains at least \( k \) times the number of votes for B can be converted into one in which A maintains more than \( k \)
times the number of votes for B by simply appending a vote for A to the beginning of the permutation. Clearly this process is reversible, and hence the solution to the weak version is the same as the “strict” version when A receives $m + 1$ votes and B receives $n$ votes:

$$\frac{(m + 1) - kn}{(m + 1) + n} \binom{(m + 1) + n}{m + 1} = \frac{m + 1 - kn}{m + 1} \binom{m + n}{m}.$$ 

Putting $k = 1$ and $m = n$ produces the **Catalan numbers**:

$$C_n = \frac{1}{n + 1} \binom{2n}{n}.$$

**Sources**

This section is adapted from Renault, [17] and [18]. The Reflection Principle is adapted from [15]. Some the history is also adapted from [14]. The rotation and shadowing proof is adapted from [24]. Some of the problems are from [24].

**Algorithms, Scripts, Simulations**

**Algorithm**

For the generalized ballot problem, set the size of a downstep $k$, the number $a$ of ballots for A and the number $b$ of ballots for B. Then the total number of ballots is the length $l = a + b$. There are then $\binom{a + b}{a - b}$ places to have the downsteps corresponding to ballots for B. In a sequence of nested loops systematically place the downsteps representing ballots for B in an $(a + b) \times (a + b)$ matrix of ones. Then cumulatively sum this matrix down the columns to get all possible running ballot counts and save the matrix for later plotting or analysis.

Compute $a - kb$. For each of the $\binom{a + b}{a - b}$ ballot counts, find the minimum value $y_0$. Then create a block matrix of size $(a - kb) \times (a - kb) \binom{a + b}{a - b}$ with blocks
of size \((a-\frac{kb}{2}) \times (a-\frac{kb}{2})\). Fill each block with the coordinates of the points corresponding to the values \(L(P) = \{r(t)|t \in Z, y_0 \leq t \leq y_0 + (a-\frac{kb}{2})-1\}\). Save the matrix for later plotting or analysis.

Note that this algorithm does not scale well because of the inherent growth of the binomial coefficients.

**Scripts**

**Octave**

Octave script for uniform partitions

```octave
k = 2;
a = 6;
b = 2;
length = a+b;

ballots = ones(length, length*(length-1)/2);

# Systematically place the downsteps from ballots for B.
c = 0;
for i = 1:length-1
    for j = (i+1):length
        c = c + 1;
        ballots(i,c) = -k;
        ballots(j,c) = -k;
    endfor;
endfor;

ballot_count = cumsum(ballots);

# Put the ballot count in a matrix whose
# first column is x-coordinates
# and first row is 0 to start at origin
plot_ballot_count = zeros(length+1, length*(length-1)/2 + 1);
plot_ballot_count(2:length+1, 2:length*(length-1)/2 + 1) = ballot_count;
plot_ballot_count(:,1) = 0:length;
```

23
save uniform_partition_plots.dat plot_ballot_count

# Build an block matrix with length*(length−1)/2 blocks
# of size (a−k*b)−by−2. Blocks correspond to each plot.
# Rows of each block contain the (a−k*b) points of L(P),
# where for path (or plot) P in good paths \mathcal(A)
y0 = \text{min}(plot\_ballot\_count);
amkb = a - k*b;
LP = zeros(amkb,amkb*length*(length−1)/2);
# Octave find treats a whole matrix as a single column,
# but I need to treat the uniform partition column by column, so here
# where for clarity, better to loop over columns, using find on each
for p=1:length*(length−1)/2   # amkb−by−2 block for each ballot count
    for i = 1:amkb  # rows for y0, y0+1, ..., y0+(amkb−1)
        LP(i,2*p−1) = \text{max}( \text{find}(plot\_ballot\_count(:,p+1) == y0(p+1) +
            # first column of block p is x-coord where L(P) member (i−1) is
            LP(i,2*p) = y0(p+1) + (i−1);
            # second column of block p is y-coord of the L(p) member (i−1)
        endfor;
    endfor;

save -append uniform_partition_plots.dat LP

GnuPlot

GnuPlot script for uniform partitions

set term wxt size 640,768  # seems to be a pleasing ratio of width to height
set multiplot layout 7,4
#
set border 0  # no border (mathematical, not eng style)
set xtics nomirror  # no tic marks on top
set ytics nomirror  # ... or right
set xzeroaxis linetype −1
set xtics axis
set yzeroaxis linetype −1
unset key
# 1s 7 is black filled circles, with black lines
do for [i=2:29] {
j = 2*i - 3 # gnuplot doesn’t seem to like arithmetic
k = 2*i - 2 # ” using (2*i - 3):(2*i - 4) so do it here
plot ’uniform_partition_plots.dat’ index 0 \\
  using 1:i:xticlabel(“”):yticlabel(“”): \\
  with linespoints ls 7, \\
  index 1 \\
  using j:k \\
  with points ls 7 lc rgb ”red” \\
}
unset multiplot

Problems to Work for Understanding

1. Do the calculations to show

\[
\left( \frac{a + b - 1}{a - 1} \right) - 2 \left( \frac{a + b - 1}{b - 1} \right) = \frac{a - b}{a + b} \left( \frac{a + b}{a} \right)
\]

2. Do the calculations to show

\[
\left( \frac{a + b}{a} \right) - 2 \left( \frac{a + b - 1}{a} \right) = \frac{a - b}{a + b} \left( \frac{a + b}{a} \right).
\]

Explain why this is the same as the previous problem.

3. Do the calculations to show

\[
\left( \frac{a + b}{a} \right) - (k + 1) \left( \frac{a + b - 1}{a} \right) = \frac{a - kb}{a + b} \left( \frac{a + b}{a} \right).
\]

4. Do the calculations to show

\[
\frac{a - k(b - 1)}{a + b - 1} \left( \frac{a + b - 1}{a} \right) + \frac{a - 1 - kb}{a + b - 1} \left( \frac{a + b - 1}{a - 1} \right)
\]
simplifies to
\[
\frac{a - kb}{a + b} \left( \frac{a + b}{a} \right)
\]

5. Consider the case of 7 votes with 4 votes for A and 3 votes for B. Make a chart of “good” and “bad” ballot counting permutations and rearrangements as in the example of André’s proof.

6. Suppose \( m + n \) people are waiting in line at a box office; \( n \) of them have only five-dollar bills and the other \( m \) have only ten-dollar bills with nothing smaller. The tickets cost $5 each. When the box-office opens, the till has no money. If each customer buys just one ticket, what is the probability that none of them will have to wait for change? (From [24], problem 83a.)

7. Solve the same problem under the assumption that initially there were \( p \) five-dollar bills in the till.

8. For the purposes of this problem assume there exist three-dollar bills. Suppose \( m + n \) people are waiting in line at a box office; \( n \) of them have only one-dollar bills and the other \( m \) have only three-dollar bills and nothing smaller. The tickets cost $1 each and each person wants one ticket. When the box-office opens, the till has no money. If each customer buys just one ticket, what is the probability that none of them will have to wait for change? (From [24], problem 83c.)

9. Modify the scripts to display the uniform partition of the ballot counts for 4 votes for A and 3 votes for B.
Reading Suggestion:

References


________

**Outside Readings and Links:**

1.
2.
3.
4.

I check all the information on each page for correctness and typographical errors. Nevertheless, some errors may occur and I would be grateful if you would alert me to such errors. I make every reasonable effort to present current and accurate information for public use, however I do not guarantee the accuracy or timeliness of information on this website. Your use of the information from this website is strictly voluntary and at your risk.

I have checked the links to external sites for usefulness. Links to external websites are provided as a convenience. I do not endorse, control, monitor, or guarantee the information contained in any external website. I don’t guarantee that the links are active at all times. Use the links here with the same caution as you would all information on the Internet. This website reflects the thoughts, interests and opinions of its author. They do not explicitly represent official positions or policies of my employer.

Information on this website is subject to change without notice.
Steve Dunbar’s Home Page, [http://www.math.unl.edu/~sdunbar1](http://www.math.unl.edu/~sdunbar1)
Email to Steve Dunbar, sdunbar1 at unl dot edu
Last modified: Processed from \LaTeX{} source on December 31, 2013