AN INVESTIGATION OF THE CHUNG-FELLER THEOREM AND SIMILAR COMBINATORIAL IDENTITIES

ELI A. WOLFHAGEN

Abstract. In this paper, we shall prove the Chung-Feller Theorem in multiple ways as well as extend its results to more general cases, similar to that of the ballot problem.

1. Introduction

The main focus of this paper is to prove the following result:

The Chung Feller Theorem. *The number of paths from $(0,0)$ to (n,n) with exactly $2k$ steps above the line $x = y$ is independent of k, for every k such that $0 \leq k \leq n$. In fact, it is equal to the nth Catalan number.*

2. Preliminary Definitions and an Inductive Proof

Generalizing Dyck paths, we can study the number of paths indicated in the above Chung-Feller theorem in a simpler form.

Definition 1. A *k-negative path* of length $2n$ is a path from $(0,0)$ to $(2n,0)$, with steps $(1,1)$ and $(1,-1)$, such that exactly $2k$ of these steps are below the horizontal axis.

The set of all Dyck paths forms a free monoid, so it is possible to uniquely factor any Dyck path into what are called prime paths.

Definition 2. A *prime Dyck path* is a Dyck path that returns to the x-axis only once at the end of the path. A *negative prime Dyck path* is a a prime Dyck path reflected about the x-axis.

Note. In general, any *negative* Dyck path is the reflection about the x-axis of a Dyck path.

Any k-negative path either starts out with some sort of prime path, either a prime Dyck path or a negative prime Dyck path. In the first case, the k-negative path starts with an up-step followed by a Dyck path of some length $2p-2$ which is then followed by a down-step and a k-negative path of length $2n-2p$, for some $p \leq n-k$. On the other hand, the second case deals with paths that start out with a down-step, followed by a $(q-1)$-negative path of length $2q-2$ which is followed by an up-step and a $(k-q)$-negative path of length $2n-2q$, for some $q \leq k$.

Before our proof, we should introduce some preliminary notation to ease the computation:

Notation 1. Let $\mathcal{N}_{up}$ be the number of k-negative paths of length $2n$ that start with an up-step. Let $\mathcal{N}_{down}$, similarly, be the number of k-negative paths of length $2n$ that start with a down-step. Let $\mathcal{N}$ be the total number of k-negative paths of length $2n$.

An Inductive Proof of the Chung-Feller Theorem. As a base case, for $n = 0$, there is only one path with 0 steps. So clearly there is only $C_0 = 1$ path with 0 steps below the x-axis.

Now, assume that for all i such that $0 \leq i < n$, the number of k-negative paths of length $2i$ is equal to C_n for all $k \leq i$. So for the case of k-negative paths of length $2n$ we will deal with both cases described above.

If a path starts out with an up-step and a Dyck path of length $2p - 2$, then the number of such paths is the total number of Dyck paths of length $2p-2$ multiplied by the number of k-negative paths of length $2n - 2p < 2n$. By the inductive hypothesis, this means that the number of such paths $N_p^+ = C_{p-1}C_{n-p}$. For the total number of k-negative paths that start off with an up-step we need to take the sum of all the N_p^+, for all $1 \leq p \leq n - p$. Therefore, we have that the total number of k-negative paths of length $2n$ that start with an up-step is

$$\mathcal{N}_{up} = \sum_{p=1}^{n-k} N_p^+ = \sum_{p=1}^{n-k} C_{p-1}C_{n-p}.$$

Similarly we define N_q^- as the number of k-negative paths that start out with a down-step and negative Dyck path of length $2q - 2$. So N_q^- is equal to the number of negative Dyck paths of length $2q - 2$ multiplied by the number of $(k - q)$-negative paths of length $2n - 2q < 2n$. Obviously, since the negative Dyck paths are reflections of (positive) Dyck paths, the number of negative Dyck paths of length $2q - 2$ is equal to the number of Dyck paths of length $2q - 2$.Therefore, we again have the summation of products of two Catalan numbers, and the total number of k-negative paths of length $2n$ that start with a down-step is

$$\mathcal{N}_{down} = \sum_{q=1}^{k} C_{q-1}C_{n-q} = \sum_{q=1}^{k} C_{n-q}C_{q-1}.$$

Therefore the total number of k-negative paths of length $2n$ is equal to

$$\mathcal{N} = \mathcal{N}_{up} + \mathcal{N}_{down} = \sum_{p=1}^{n-k} C_{p-1}C_{n-p} + \sum_{q=1}^{k} C_{n-q}C_{q-1}, \tag{1}$$

that is

$$\mathcal{N} = (C_0C_{n-1} + C_1C_{n-2} + \cdots + C_{n-k-1}C_k) + (C_{n-1}C_0 + C_{n-2}C_1 + \cdots + C_{n-k}C_{k-1}).$$

Reversing the order of the summands in the second parentheses clearly gives

$$\begin{aligned}\mathcal{N} &= (C_0C_{n-1} + \cdots + C_{n-k-1}C_k) + (C_{n-k}C_{k-1} \cdots + C_{n-1}C_0)\\ &= \sum_{i=0}^{n-1} C_iC_{n-i-1} = C_n.\end{aligned}$$

Therefore, the total number of k-negative paths of length $2n$ is equal to C_n. □

3. A Bijective Proof

Notation 2. Let S_k denote the set of all k-negative paths of length $2n$.

The inductive proof in section 2 reveals the inner workings of a bijection $\phi : S_k \to S_{k+1}$. We will define such a bijection recursively.

For $s \in S_k$, find the last positive prime Dyck path q. Next factor s into $s = pqr$, where p is the arbitrary path up to the last positive prime Dyck path and r is the rest of the path after q. Since q is the last positive prime Dyck path, the remainder of the path r must be a negative Dyck path.

Like any other prime Dyck path, q is composed of an up-step followed by an arbitrary Dyck path Q and a down-step. So q can be rewritten as $q = uQd$, where u denotes an up-step and d denotes a down-step. Define the bijection ϕ by

$$\phi(s) = \phi(puQdr) = pdruQ, \quad \text{for any} s \in S_k. \tag{2}$$

Claim 1. *$\phi : s \mapsto \phi(s) = pdruQ$ is a bijection between S_k and S_{k+1}.*

Proof. Given an $s \in S_k$, $\phi(s) = pdruD$ by (2). Since p begins and ends on the the x-axis, the step d is below the x-axis. As noted earlier r is an arbitrary negative Dyck path, so in the path $\phi(s)$, r starts and ends at height -1, without going above it. Therefore the step u is negative, starting at height -1 and ends on the x-axis. Since Q is a Dyck path it contains no negative steps, so altogether $\phi(s)$ has 2 steps below the x-axis in addition to the number of negative steps in path segments p and r. Since the only negative steps in the path $s \in S_k$ occur in p and r, the number of negative steps in path segments p and r is equal to $2k$. Thus $\phi(s)$ has $2k + 2$ steps below the x-axis and so is a member of S_{k+1}.

Now, it is possible to think of ϕ as a cyclic shift on only the qr portion of s. The cyclic shift occurs at the step d, sending $uQdr$ to $druQ$. Since cyclic shifts are bijective and the identity function is clearly bijective, ϕ is bijective. □

Plugging in k and $k + 1$ into equation (1) we get

$$\mathcal{N}_k = \sum_{p=0}^{n-k} C_{p-1}C_{n-p} + \sum_{q=0}^{k} C_{q-1}C_{n-q}, \quad \text{and}$$
$$\mathcal{N}_{k+1} = \sum_{p=0}^{n-k-1} C_{p-1}C_{n-p} + \sum_{q=0}^{k+1} C_{q-1}C_{n-q}.$$

A careful inspection of these two equations reveals that number $C_{n-k-1}C_k$ moves from the first summation in $\mathcal{N}_k$ to the second summation in $\mathcal{N}_{k+1}$. That is, the set of paths starting with a prime Dyck path of length $2n - 2k$ followed by a negative Dyck path of length $2k$, is sent to the set of paths starting with a negative prime Dyck path of length $2k-2$ followed by an arbitrary Dyck path of length $2n-2k-2$. So any possible bijection between S_k and S_{k+1} must have this property, sending the path $s' = qr \in S_k$ to the path $\hat{s} = vw \in S_{k+1}$ where q and v are the positive and negative prime Dyck paths, respectively, described above.

Claim 2. *The ϕ described by equation (2) fulfills the above property.*

Proof. $s' \in S_k$ has only one positive prime $q = uDd$ of length $2n - 2k$, therefore D has length $2n - 2k - 2$. Additionally, r is the remaining negative Dyck path of length $2k$. By the definition of our bijection, $\phi(s') = druD$, which starts with a negative prime Dyck path of length $2k + 2$ and ends with D which is a Dyck path of length $2n - 2k - 2$. Therefore, if we call $v = dru$ and $w = D$ it is easy to see that our ϕ satisfies the desired property. $\square$

Therefore, we can now bijectively prove the Chung-Feller theorem.

Bijective Proof. Since ϕ exhibits the property enumerated above and is bijective, the cardinality of S_k and S_{k+1} are equal, for all nonnegative $k \leq n-1$. Since there are $n+1$ equal sets of paths of length $2n$ each with the same cardinality, and the total number of paths of length $2n$ is $\binom{2n}{n}$, we have the equation

$$(n+1)\mathcal{K} = \binom{2n}{n},$$

where $\mathcal{K}$ is the number of k-negative paths of length $2n$ for some k such that $0 \leq k \leq n$.

Therefore $\mathcal{K} = \frac{1}{n+1}\binom{2n}{n} = C_n$. $\square$

4. A Generating Function Approach to the Theorem

The generating function for the Catalan numbers is given by

$$c(x) = \sum_{n=0}^{\infty} C_n x^n = \frac{1-\sqrt{1-4x}}{2x}. \tag{3}$$

Let us construct a generating function for k-negative paths by weighting each step by $\sqrt{x}$ and each step below the x-axis by $\sqrt{t}$. As described in section 2, the primes of a k-negative path are either positive prime Dyck paths or negative prime Dyck paths. Let P_n denote the number of (positive) prime Dyck paths of length $2n$.

The common decomposition of prime Dyck paths breaks such a path of length $2n$ into an arbitrary Dyck path of length $2n-2$ sandwiched between an up-step and down-step. So the number of prime Dyck paths of length $2n$ is given by the $(n-1)$th Catalan number, that is $P_n = C_{n-1}$. Therefore the generating function for the prime Dyck paths is given by

$$p_+(x) = \sum_{n=1}^{\infty} P_n x^n = \sum_{n=1}^{\infty} C_{n-1} x^n = x \sum_{n=0}^{\infty} C_n x^n = xc(x). \tag{4}$$

Similarly, for negative prime Dyck paths the generating function, now weighted by $\sqrt{xt}$ because each step in a negative prime Dyck path is below the x-axis is given by

$$p_-(x,t) = \sum_{n=1}^{\infty} P_n (tx)^n = tx \sum_{n=0}^{\infty} C_n (tx)^n = txc(tx). \tag{5}$$

So since arbitrary paths can be factored into l primes (either positive or negative), we can construct a generating function out of the geometric sum of these primes,

$$N(x,t) = \sum_{l=0}^{\infty} (p_- + p_+)^l = \frac{1}{1-(p_- + p_+)}.$$

In this case the coefficient of $t^i x^j$ is the number paths of total length $2j$ and with $2i$ steps below the x-axis.

By the definition of $c(x)$ and equations (4) and (5) we get that

$$N(t,x) = \frac{1}{1 - xc(x) - txc(tx)}$$
$$= \frac{1}{1 - (x\frac{1-\sqrt{1-4x}}{2x} + tx\frac{1-\sqrt{1-4tx}}{2tx})}$$
$$= \frac{2}{\sqrt{1-4x} + \sqrt{1-4tx}}.$$

Rationalizing the denominator we see that

$$N(t,x) = \frac{\sqrt{1-4x} - \sqrt{1-4tx}}{2x(t-1)}$$
$$= \frac{1 - \sqrt{1-4x} - 1 + \sqrt{1-4xt}}{2x(1-t)}$$
$$= \frac{1}{1-t} \cdot \left(\frac{1-\sqrt{1-4x}}{2x} - \frac{1-\sqrt{1-4xt}}{2x} \right)$$
$$= \frac{1}{1-t} \left(c(x) - tc(tx) \right)$$
$$= \frac{1}{1-t} \left(\sum_{n=0}^{\infty} C_n x^n - \sum_{n=0}^{\infty} C_n t^{n+1} x^n \right)$$

$$N(t,x) = \sum_{n=0}^{\infty} \frac{1-t^{n+1}}{1-t} C_n x^n. \tag{6}$$

Using $N(t,x)$ and a simple algebraic identity we can prove the Chung-Feller Theorem.

Proof. Because

$$\frac{1-t^{n+1}}{1-t} = 1 + t + t^2 + \cdots + t^n,$$

equation (6) can be rewritten as

$$k(t,x) = \sum_{n=0}^{\infty} C_n \left(\sum_{k=0}^{n} t^k \right) x^n. \tag{7}$$

It is easy to see from (7) that given an n thecoefficient of $t^k x^n$ in $N(t,x)$ is equal to C_n for all k such that $0 \le k \le n$. Based on the construction of $N(t,x)$, this means that the number of k-negative paths of length $2n$ is equal to C_n for all k such that $0 \le k \le n$. □

5. Generalized Ballot Numbers

Definition 3. A *k-bad ballot path* from $(0,0)$ to (m,n), with $m \ge n$, is a path with steps $(1,0)$ and $(0,1)$ such that exactly $2k$ steps lie above the main diagonal $x = y$.

Definition 4. The *k-bad ballot number* $B_k(m,n)$ denotes the number of k-bad paths from $(0,0)$ to (m,n), with $m \ge n$.

As formulated above the k-bad ballot numbers differ from the standard ballot numbers in the type of election the numbers model. The classical ballot problem asks for the number of ways to count ballots in a two-person election such that candidate A is always ahead in the tally over candidate B. That is, the ballot problem deals with the number of paths from $(0,0)$ to (m,n), with $m > n$, such that the path never touches the main diagonal $x = y$.

However, with our new understanding of paths that jump over the main diagonal it is possible to model a more realistic ballot count, one in which there are exactly k votes for candidate B such that after the tallying of the vote he is ahead of candidate A. We shall call this type of election a *k-bad election.* Now, to model this event we need to know the number of paths from $(0,0)$ to (m,n), with $m \geq n$, which have $2k$ steps above the diagonal, that is k-bad ballot paths.

For any such path, find the last time that it intersects the diagonal and cut it there; we know that the point at which the path is severed will be (i,i), for some i with $k \leq i \leq n$. Now we have to count two different types of paths:

1. paths from $(0,0)$ to (i,i) with $2k$ steps above the diagonal, and
2. paths from (i,i) to (m,n) that stay strictly below the diagonal,

for some i such that $k \leq i \leq n$. Clearly, by the Chung-Feller Theorem, because $0 \leq k \leq i$, the total number of paths from $(0,0)$ from (i,i) with $2k$ steps above the diagonal is just C_i.

If $m = n$, then the last point at which any path hits the main diagonal is (n,n), so there is no remaining path to enumerate. However, if $m > n$ then we have to count the number of possibilities.

Lemma 1. *The number of paths from (i,i) to (m,n), with $m > n$, that stay strictly below the diagonal is equal to $B(m-i, n-i)$, where*

$$B(m,n) = \binom{m+n-1}{n} - \binom{m+n-1}{m} \tag{8}$$

Proof. If we let $M = m - i$ and $N = n - i$, then it is sufficient to just prove that the number of paths that never touch the main diagonal from $(0,0)$ to (M,N) with $M > N$ is $B(M,N)$. Clearly in order for a path from $(0,0)$ to (M,N) to stay below the main diagonal it must start out with a unit east step, thus since this is the only way to proceed from $(0,0)$ it is sufficient to count the number of paths from $(1,0)$ to (M,N) that do not touch the main diagonal. Thus of the total number of paths from $(1,0)$ to (M,N), there are paths that do not touch the main diagonal and those that do, which we shall call "bad" paths.

Now for each bad path, if one takes the path up to the first time it hits the line $x = y$, and reflect just this portion of the path about the main diagonal, a path from $(0,1)$ to (M,N) is obtained. Also if, one takes any path from $(0,1)$ to (M,N), finds the first time it touches the main diagonal and reflects the path up to this point about the line $x = y$, a bad path is obtained. Thus there exists a bijection between the set of all bad paths and the set of all path from $(0,1)$ to (M,N).

Since we do not want the bad paths, we take the total number of paths from $(1,0)$ to (M,N) and subtract the number of bad paths, that is the total number of paths from $(0,1)$ to (M,N). Therefore, the desired number of paths is

$$\binom{M+N-1}{N} - \binom{M+N-1}{M},$$

which by (8) we know to be equal to $B(M,N)$. □

This means that for any i such that $k \leq i \leq n$, the number of paths from $(0,0)$ to (m,n), such that $m > n$, with $2k$ steps above the diagonal which last intersects the diagonal at (i,i) is

$$C_iB(m-i,n-i).$$

Therefore, summing over all possible i such that $k \le i \le n$, the final formulation for the number of ways to count the ballots of a two-person k-bad election is

$$\sum_{i=k}^{n} C_iB(m-i,n-i).$$

Adding in the case in which candidates A and B tie overall, $(m = n)$, we get

$$B_k(m,n) = \begin{cases} \sum_{i=k}^{n} C_iB(m-i,n-i), & \text{if } m > n; \\ C_n, & \text{if } m = n \ge k; \\ 0, & \text{if } n < k. \end{cases} \tag{9}$$

Unfortunately there is no closed form for this sum for all k or for all points (m,n), with $m \ge n$.

The k-bad ballot numbers can also be defined by a recurrence equation, extended from the original ballot numbers $B(m,n)$. Clearly, paths ending at (m,n) either go through $(m-1,n)$ or $(m,n-1)$, so $B(m,n) = B(m-1,n) + B(m,n-1)$. This recurrence relation can be shown to generalize to our new k-bad ballot numbers.

Claim 3. *For $m > n \ge 0$, $B_k(m,n) = B_k(m-1,n) + B_k(m,n-1)$, with the initial conditions that*

$$B_k(n,n) = \begin{cases} C_n, & \text{for } n \ge k; \\ 0, & \text{for } n < k \end{cases}$$

and $B_k(m,-1) = 0$ for all nonnegative integers m.

Proof. Since k-bad paths have steps of $(0,1)$ and $(1,0)$ in order for a k-bad path to reach (m,n) it must either go through $(m,n-1)$ or $(m-1,n)$. Therefore, we get that the number of k-bad paths from $(0,0)$ to (m,n) is equal to the sum of the number of k-bad paths from $(0,0)$ to $(m-1,n)$ and from $(0,0)$ to $(m,n-1)$. Thus by the definition of k-bad ballot numbers, we get the result

$$B_k(m,n) = B_k(m-1,n) + B_k(m,n-1). \tag{10}$$

The initial conditions for $B_k(n,n)$ and $B_k(m,-1)$ come about by a similar common sense inspection of k-bad paths themselves. Obviously there is no way for a path with steps (1,0) and (0,1) to reach a point $(m,-1)$ for any m. Also, if $n < k$ then there is no way to fit $2k$ steps above the main diagonal if the total path length is only $2n$. Furthermore, the Chung-Feller theorem states that if $n \ge k$, then the number of k-bad paths from (0,0) to (n,n) is exactly C_n. So our recurrence and initial conditions accurately describe the k-bad ballot numbers. □

6. Generating functions for $B_k(n+l,n)$

If we focus our attention on just the set of points $(n+l,n)$ for some $l \ge 0$, then we can formulate the generating function for the k-bad ballot numbers easily. This situation is equivalent to enumerating the number of ways ballots can be tallied in a k-bad election where the loser receives n votes, and the winner has a margin of victory of l.

The formulation of a generating function is easier is because it is commonly known that the generating function for the standard ballot numbers $B(n+l,n)$ for some nonnegative integer l is given by

$$b_l(x) = \sum_{n=0}^{\infty} B(n+l,n)x^n = c(x)^l, \tag{11}$$

where $c(x)$ is the generating function for the Catalan numbers given by (3) in section 4.

Plugging in to equation (9) we get that for any nonnegative integer k,

$$B_k(n+l,n) = \begin{cases} \sum_{i=k}^{n} C_i B(n+l-i,n-i), & \text{if } l > 0; \\ C_n, & \text{if } l = 0; \\ 0, & \text{if } l < 0. \end{cases} \tag{12}$$

Claim 4. *The generating function for 0-bad ballot numbers for any nonnegative integer margin of victory l is equal to $b_{0,l}(x) = \sum_{n=0}^{\infty} B_0(n+l,n)x^n = c(x)^{l+1}$.*

Proof. By definition $B_0(n+l,n) = \sum_{i=0}^{n} C_i B(n+l-i,n-i)$. So by (3) and (11) we get $B_0(n+l,n) = [x^n]c(x)c(x)^l$. So therefore, if $B_0(n+l,n)$ is equal to the coefficient of x^n in the generating function $c(x)^{l+1}$, then since $b_{0,l}(x) = \sum_{n=0}^{\infty} B_0(n+l,n)x^n$, we get that $b_{0,l}(x) = c(x)^{l+1}$. □

This result is different from the (11) because the original ballot numbers count paths that stay strictly beneath the main diagonal, while the 0-bad ballot numbers count paths that do not go above the main diagonal.

We can generalize this result, by finding a generating function for k-bad ballot numbers for any nonnegative k.

Theorem 1. *For any nonnegative integers l and k, the generating function for the k-bad ballot numbers with a margin of victory of l is*

$$b_{k,l}(x) = \sum_{n=0}^{\infty} B_k(n+l,n)x^n = c(x)^l(c(x) - \sum_{i=0}^{k-1} C_i x^i). \tag{13}$$

Proof. Fix a nonnegative integer l. Then for any nonnegative integer k, we want a formula for $\sum_{n=0}^{\infty} B_k(n,l)x^n$, where $B_k(n,l) = \sum_{i=k}^{n} C_i B(n+l-i,n-i)$.

Now, $c(x)^l(c(x) - \sum_{i=0}^{k-1} C_i x^i)$ is the generating function for some infinite sequence of numbers. It is also known that $c(x)^l$ is the generating function for the ballot numbers $B(j+l,j)$, that is $\sum_{j=0}^{\infty} B(j+l,j)x^j = c(x)^l$. Additionally, it is obvious that $c(x) - \sum_{i=0}^{k-1} C_i x^i = \sum_{i=0}^{\infty} C_i x^i - \sum_{i=0}^{k-1} C_i x^i = \sum_{i=k}^{\infty} C_i x^i$. So we get that

$$\begin{aligned} c(x)^l\left(c(x) - \sum_{i=0}^{k-1} C_i x^i\right) &= \left(\sum_{j=0}^{\infty} B(j+l,j)x^j\right) \cdot \left(\sum_{i=k}^{n} C_i x^i\right) \\ &= \sum_{i \geq k, j \geq 0} C_i B(j+l,j)x^{i+j}. \end{aligned} \tag{14}$$

Now, if we set $i+j = n$ then $j = n-i \geq 0$ and equation (14) becomes

$$\sum_{n \geq i \geq k} C_i B(n+l-i,n-i)x^n,$$

which can be rewritten as

$$\sum_{n=0}^{\infty} x^n \Bigg(\sum_{i=k}^{n} C_i B(n+l-i, n-i)\Bigg).$$

By equation (12) we get that

$$c(x)^l \Bigg(c(x) - \sum_{i=0}^{k-1} C_i x^i\Bigg) = \sum_{n=0}^{\infty} B_k(n+l, n)x^n = b_{k,\,l}(x).$$

□

Therefore we have one way to represent the generating function of $B_k(n+l, n)$ for any nonnegative k and l. Let us look at one particular case of this general formula.

Corollary 1. *For any nonnegative integer l, the generating function for $B_1(n, l)$ is*

$$b_{1,\,l}(x) = \sum_{n=0}^{\infty} B_1(n, l)x^n = c(x)^{l+1}.$$

Proof. So, given some nonnegative integer l, we can plug in to the formula derived in the previous theorem we get that

$$\begin{aligned} b_{1,\,l}(x) &= \sum_{n=0}^{\infty} B_1(n, l)x^n \\ &= c(x)^l (c(x) - \sum_{i=0}^{1-1} C_i x^i) \\ &= c(x)^l (c(x) - C_0 x^0) = c(x)^l (c(x) - 1). \end{aligned}$$

Now, we know that $c(x) = 1 + xc(x)^2$, so we get that $c(x) - 1 = xc(x)^2$, and finally that

$$b_{1,\,l}(x) = c(x)^l (c(x) - 1) = xc(x)^{l+2}. \tag{15}$$

□

This looks to be a much easier way of representing the generating function for $B_1(n+l, n)$. Combinatorially, this formulation of the generating function is much more illuminating than the generating function given in equation (13). Without loss of generality, it is sufficient to provide a combinatorial interpretation for $l = 0$. Any 1-bad ballot path with $l = 0$ is composed of an ordered pair of arbitrary Dyck paths separated by two steps above the main diagonal. Each arbitrary Dyck path is represented by $c(x)$ and the two steps above the diagonal are weighted by an x, so altogether there the generating function for 1-bad ballot paths with $l = 0$ is given by $xc(x)^2$. This clearly agrees with equation (15).

The form of (15) also generalizes to another way to write the generating function for arbitrary nonnegative k using the standard ballot numbers.

Theorem 2. *Given a nonnegative l, for any positive k, the generating function for the k-bad ballot numbers with a margin of victory l can be written as*

$$b_{k,\,l}(x) = x^k c(x)^{l+1} (\sum_{j=0}^{k-1} B(k, j) c(x)^{k-j}).$$

Note. We will denote the polynomial in x and $c(x)$, $x^k c(x)^{l+1}\left(\sum_{j=0}^{k-1} B(k,j)c(x)^{k-j}\right)$ as P_k, because it is understood that l is a fixed nonnegative integer.

Proof. Fix a nonnegative integer l as a margin of victory. As a base case for 1-bad paths the generating function

$$\begin{aligned} P_1 &= x^1 c(x)^{l+1}\left(\sum_{j=0}^{1-1} B(1,j)c(x)^{1-j}\right) \\ &= xc(x)^{l+1}(B(1,0)c(x)) = xc(x)^{l+2} = b_{1,l}(x). \end{aligned}$$

Assume that for some integer $m > 1$, the generating function for the k-bad ballot numbers with margin of victory l is given by

$$b_{m,l}(x) = P_m = x^m c(x)^{l+1}\left(\sum_{j=0}^{m-1} B(m,j)c(x)^{m-j}\right). \tag{16}$$

Now let us look at

$$P_{m+1} = x^{m+1}c(x)^{l+1}\left(\sum_{j=0}^{m} B(m+1,j)c(x)^{m+1-j}\right).$$

By the recurrence equation for standard ballot numbers $B(m+1,j) = B(m,j) + B(m+1,j-1)$, so we get that

$$\begin{aligned} P_{m+1} &= x^{m+1}c(x)^{l+1}\left(\sum_{j=0}^{m} B(m+1,j)c(x)^{m-j}\right) \\ &= x^{m+1}c(x)^{l+1}\left(\sum_{j=0}^{m}(B(m,j) + B(m+1,j-1))c(x)^{m+1-j}\right). \end{aligned}$$

Now splitting this up into two summations we get

$$\begin{aligned} S_1 &= x^{m+1}c(x)^{l+1}\left(\sum_{j=0}^{m} B(m,j)c(x)^{m+1-j}\right), \quad \text{and} \\ S_2 &= x^{m+1}c(x)^{l+1}\left(\sum_{j=0}^{m} B(m+1,j-1)c(x)^{m+1-j}\right). \end{aligned}$$

Since $B(m,m) = 0$, S_1 can be written as

$$S_1 = xc(x)(x^m c(x)^{l+2}\left(\sum_{j=0}^{m-1} B(m,j)c(x)^{m-j}\right) = xc(x)P^l.$$

Similarly, since $B(m+1,-1) = 0$, S_2 can be rewritten as

$$\begin{aligned} S_2 &= x^{m+1}c(x)^{l+1}\left(\sum_{j=1}^{m+1} B(m+1,j-1)c(x)^{m+1-j}\right) \\ &= \frac{1}{c(x)} \cdot x^{m+1}c(x)^{l+1}\left(\sum_{j=1}^{m+1} B(m+1,j-1)c(x)^{m+1-(j-1)} - B(m+1,m)c(x)\right) \\ &= \frac{1}{c(x)}P_{m+1} - x^{m+1}c(x)^{l+1}B(m+1,m). \end{aligned}$$

So we now have the equation

$$P_{m+1} = S_1 + S_2 = xc(x)P_m + \frac{1}{c(x)}P_{m+1} - x^{m+1}c(x)^{l+1}B(m+1,m),$$

which can be solved for P_{m+1} yielding the equation

$$\left(1 - \frac{1}{c(x)}\right)P_{m+1} = xc(x)P_m - xc(x)(x^m c(x)^l B(m+1,m)). \tag{17}$$

Since $B(m+1,m) = C_m$, and since $c(x) = 1 + xc(x)^2$, we can simplify the above equation into the form

$$xc(x)P^l = xc(x)\left(P_m - c(x)^l C_m x^m\right),$$

which upon cancelling the $xc(x)$ from both sides gives

$$P_{m+1} = P_m - c(x)^l C_m x^m. \tag{18}$$

By (16),

$$P_m = b_{m,\,l} = c(x)^l\left(c(x) - \sum_{i=0}^{k-1} C_i x^i\right),$$

so therefore equation (18) becomes

$$\begin{aligned} P_{m+1} &= c(x)^l\left(c(x) - \sum_{i=0}^{m-1} C_i x^i\right) - c(x)^l C_m x^m \\ &= c(x)^l\left(c(x) - \sum_{i=0}^{m-1} C_i x^i - C_m x^m\right) \\ &= c(x)^l\left(c(x) - \sum_{i=0}^{m} C_i x^i\right) = b_{m+1,\,l}(x). \end{aligned} \tag{19}$$

Therefore by the principle of induction, given a nonnegative integer l, the generating function for the k-bad ballot numbers with a margin of victory of l is given by

$$b_{k,\,l}(x) = P_k = x^k c(x)^{l+1}\left(\sum_{j=0}^{k-1} B(k,j)c(x)^{k-j}\right),$$

for any positive integer k. □