Bidirectional Ballot Sequences, Random Walks, and a New Construction of MSTD Sets

Yufei Zhao

Department of Mathematics, Massachusetts Institute of Technology

77 Massachusetts Ave., Cambridge, MA 02139, USA

yufeiz@mit.edu

August 17, 2009

Abstract

We introduce *bidirectional ballot sequences*, which are 0-1 sequences where every prefix and every suffix contains more 1's than 0's. The probability that a random 0-1 string of length n is a bidirectional ballot sequence is shown to be asymptotically $1/4n$. We use bidirectional ballot sequences to construct a new family of MSTD (more sums than differences) subsets of $\{0, 1, \ldots, n-1\}$. Our construction gives $\Theta(2^n/n)$ MSTD sets, improving the previous best construction with $\Omega(2^n/n^4)$ MSTD sets by Miller, Orosz, and Scheinerman.

We also consider a higher dimensional generalization of bidirectional ballot sequences. For an n-step random walk on the d-dimensional lattice, what is the probability that the starting point is the coordinate-wise minimum and the ending point is the coordinate-wise maximum? We show that the answer is asymptotically $(d/n)^d$, where d is fixed and $n \to \infty$.

Keywords: bidirectional ballot sequence, random walk

2000 Mathematics Subject Classification: 05A16, 60G50

Prelude: a gambling probability

Suppose that we play the following game. We start with \$1 and then make n successive bets, each with $1/2$ probability of gaining \$1 and $1/2$ probability of losing \$1. What is the the probability that our balance never reaches zero, and our final balance at the end of the n bets is the maximum balance that we have ever had in the game?

The exact answer to this question is quite complicated. However, if we are only interested in an approximation, then the answer becomes surprisingly simple: it is about $1/n$.

1 Introduction

In this paper, we introduce a new combinatorial object.

Definition 1.1. A 0-1 sequence of length n is a *bidirectional ballot sequence* if every prefix and suffix contains strictly more 1's than 0's. The number of bidirectional ballot sequences of length n is denoted B_n.

Table 1 gives some values of B_n. At the time of this writing, the sequence B_n was not found on the Sloane On-Line Encyclopedia of Integer Sequences [9].

Table 1: Number of bidirectional ballot sequences of length n.

n	1	2	3	4	5	6	7	8	9	10	11	12
B_n	1	1	1	1	2	3	5	9	15	28	49	91
n	13	14	15	16	17	18	19	20	21	22	23	24
B_n	166	307	574	1065	2016	3769	7176	13532	25842	49113	93995	179775

Recall that a classical *ballot sequence* is a 0-1 sequence where we only require that every prefix has more 1's than 0's. A bidirectional ballot sequence is then a ballot sequence whose reverse is also a ballot sequence.

We can interpret 0-1 sequences in terms of lattice walks, where we start at the origin and take steps of the form $(1,1)$ and $(1,-1)$, corresponding to the terms 1 and 0 in the sequence, respectively. Let a *ballot walk* (resp. *bidirectional ballot walk*) be such a lattice walk corresponding to a ballot sequence (resp. bidirectional ballot sequence). So, a ballot walk is a lattice walk with the property that the starting point is the unique lowest point, and a bidirectional ballot walk has the additional property that the ending point is the unique highest point. See Figure 1 for an example.

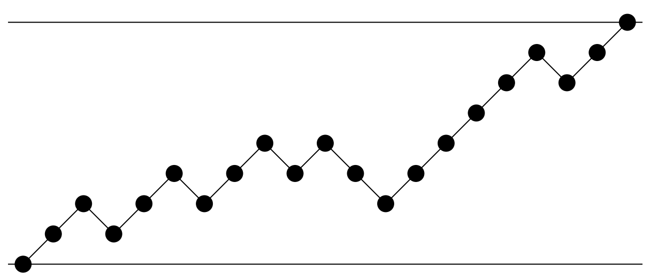

Figure 1: A bidirectional ballot walk corresponding to the sequence 1101101101001111011.

Note that the question to the question asked in the prelude is $B_{n+2}/2^n$, since there we only require that every prefix and every suffix has at least as many 1's as 0's.

Our motivation for studying bidirectional ballot sequences comes from our attempt to construct a new family of *more sums than differences* (MSTD) sets. An MSTD set is a finite set S of integers with $|S+S| > |S-S|$, where the sum set $S+S$ and the difference set $S-S$ are defined as

$$S + S = \{s_1 + s_2 : s_1, s_2 \in S\}$$
$$S - S = \{s_1 - s_2 : s_1, s_2 \in S\}.$$

Since addition is commutative while subtraction is not, two distinct integers s_1 and s_2 generate one sum but two differences. This suggests that $S+S$ should "usually" be smaller than $S-S$. Thus we expect MSTD sets to be rare.

The first example of an MSTD was found by Conway in the 1960's: $\{0,2,3,4,7,11,12,14\}$. The name MSTD was later given by Nathanson [7]. MSTD sets have recently become a popular research topic [2, 3, 4, 5, 6, 7, 10, 11].

Let $\rho_{n-1}2^n$ be the number of MSTD subsets of $\{0,1,2,\ldots,n-1\}$. We refer to ρ_n informally as the *density* of the family of MSTD sets. This quantity was first studied by Martin and O'Bryant [4], who showed that $\rho_n \geq 2\times 10^{-7}$ for $n \geq 14$. However, this bound is far from optimal. Recently, the author [11] showed that ρ_n converges to a limit greater than 4×10^{-4}. From Monte Carlo experiments, we expect $\lim \rho_n$ to be about 4.5×10^{-4} [4].

The proofs of the lower bounds on ρ_n are non-constructive. On the other hand, infinite families of MSTD sets were constructed by Hegarty [2], Nathanson [7], and Miller, Orosz, and Scheinerman [5]. In particular, Miller et al. gave the densest construction in terms of the number of subsets of $\{0,1,\ldots,n-1\}$; their construction has density $\Omega(1/n^4)$.

In this paper, we offer a different construction of an infinite family of MSTD sets using bidirectional ballot sequences. Our construction, described in Section 2, has density $\Theta(1/n)$, improving the previous result of Miller et al. [5]. The MSTD sets in the new family are in bijective correspondence with bidirectional ballot sequences of a certain length. In Section 3, we prove that our construction has density $\Theta(1/n)$ by showing the equivalent result that

$$B_n = \Theta\left(\frac{2^n}{n}\right). \tag{1}$$

However, (1) seems rather unsatisfying, since the data in Table 2 suggests that perhaps $\lim_{n\to\infty} nB_n/2^{n-2} = 1$. This guess is correct, and it is the main result of this paper.

Table 2: Some values of $nB_n/2^{n-2}$.

n	$nB_n/2^{n-2}$
100	1.0067268...
1000	1.00066729...
10000	1.0000666729...

Theorem 1.2. *The number of bidirectional ballot sequences satisfies* $B_n \sim \dfrac{2^{n-2}}{n}$.

Theorem 1.2 answers the question posed at the beginning of this paper. Our proof of Theorem 1.2, found in Section 4, is independent of the proof of the weaker result $B_n = \Theta(2^n/n)$ that we give in Section 3. The reason that we give both proofs is that our proof of weaker result is simple and revealing, while our proof of Theorem 1.2 is somewhat technical.

In Section 5 we consider a higher dimensional generalization of bidirectional ballot sequences that we call *d-confined walks*. For a fixed positive integer d and an n-step random walk with unit steps on the d-dimensional lattice $\mathbb{Z}^d$, we show that the probability that the starting point is the coordinate-wise minimum and the ending point is coordinate-wise maximum is asymptotically $(d/n)^d$ as $n \to \infty$.

Finally, in Section 6 we state some some conjectures and further questions.

2 Construction of MSTD sets

We use $[a, b]$ to denote the set $\{a, a+1, \ldots, b\}$. In this section we describe our construction of a new family of MSTD subsets of $[0, n-1]$. Our construction has density $\Theta(1/n)$, improving the previous best density of $\Omega(1/n^4)$ by Miller, Orosz, and Scheinerman. [5].

The first idea used in our construction is similar to the techniques used in both [4] and [5]; namely we look for sets of the form

$$S = L \cup M \cup R,$$

where

$$\begin{aligned} L &= S \cap [0, \ell - 1], \\ M &= S \cap [\ell, n - r - 1], \\ R &= S \cap [n - r, n - 1]. \end{aligned}$$

We will fix L and R to be sets with certain desirable properties and let M vary.

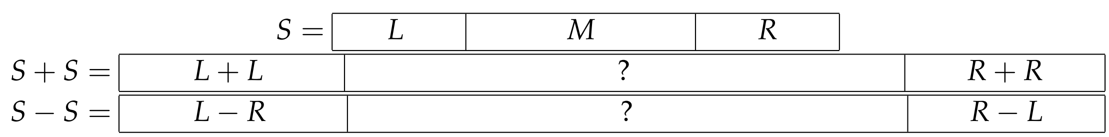

Figure 2: Illustration of the construction of S.

For instance, adapting the construction from [4] and taking $\ell = r = 11$ and

$$L = \{0, 2, 3, 7, 8, 9, 10\}, \tag{2}$$

$$R = \{n-11, n-10, n-9, n-8, n-6, n-3, n-2, n-1\}, \tag{3}$$

we have

$$L + L = [0, 20] \setminus \{1\}, \qquad R + R = [2n - 22, 2n - 2].$$

On the other hand, $S - S$ is missing at least two differences, namely $\pm(n-7)$, so $|S - S| \le 2n - 3$. If we can get $S + S$ to contain $[21, 2n - 23]$ (i.e., all the middle sums not yet covered by $L + L$ or $R + R$), then $S + S$ is only missing the sum 1, and thus $|S + S| = 2n - 2$, thereby making S an MSTD set.

So our goal is to choose M so that $S + S$ is not missing any sums in the middle segment, i.e., $[21, 2n - 23]$. From the probabilistic argument of [4], we know that the set of all M's with this property occupies a positive lower density of all subsets of $[11, n-12]$. However, that proof is non-constructive.

Note that if $M+M$ is not missing any sums (i.e., $M+M = [2 \cdot 11, 2(n-12)]$), then S has the desired properties. This condition forces $11, n-12 \in M$, so that $21, 2n-23 \in S+S$ as well. Let us temporarily do some re-indexing so that the problem becomes finding subsets M of $[1, m]$ such that $M+M = [2, 2m]$. Note that the probabilistic argument of [4] also shows that the set of such M's has at least positive constant density.

The construction of [5] is as follows: let M contain all k elements on each of its two ends (i.e., $[1, k] \cup [m-k+1, n] \subset M$), and furthermore let M have the property that it does not have a run of more than k consecutive missing elements. Here k is allowed to vary. This construction gives a density of $\Omega(1/n^4)$.

We use a different approach to construct M. The property of M that we seek is the following: for every prefix and suffix of $[1, m]$, more than half of the elements are in M. This is equivalent to saying that M is the set of positions of 1's in a bidirectional ballot sequence. The following lemma proves that this constraint is sufficient for our purposes.

Lemma 2.1. *If $M \subset [1, m]$ satisfies*

$$|M \cap [1, k]| > \frac{k}{2}, \qquad \text{and} \qquad |M \cap [m-k+1, m]| > \frac{k}{2}$$

for every $0 < k \leq m$, then $M+M = [2, 2m]$.

Proof. Let $2 \leq x \leq 2m$. If $x \leq m$, then since M contains more than half of the elements in $[1, x-1]$, by the pigeonhole principle, there is some y so that $y, x-y \in M$, so that $x \in M+M$. Similarly, if $x > m$, then since M contains more than half of the elements in $[x-m, m]$, we can find some $x-y, y \in M$ so that $m \in M+M$ as well. □

The construction of this new family of MSTD sets is summarized in the theorem below.

Theorem 2.2. *Let $n \geq 24$. Let M be a subset of $[11, n-12]$ with the property that every prefix and every suffix of the interval $[11, n-12]$ has more than half of its elements in M. Then $S = L \cup M \cup R$ is an MSTD set, where L and R are given in (2) and (3). The number of MSTD sets of $\{0, 1, \ldots, n-1\}$ in this family is $\Theta(2^n/n)$.*

Note that the choices of M are in bijection with bidirectional ballot sequences of length $n-22$. The second statement of Theorem 2.2 follows from $B_n = \Theta(2^n/n)$, which we prove in the next section.

3 Proof of $B_n = \Theta(2^n/n)$

In this section we prove the following result.

Proposition 3.1. *The number of bidirectional ballot sequences satisfies $B_n = \Theta\left(2^n/n\right)$.*

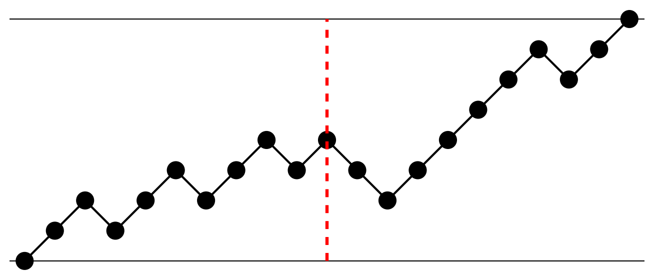

Figure 3: Dividing a bidirectional ballot walk into two halves.

The key idea in the proof is to divide a bidirectional ballot walk into two halves, as in Figure 3. The second half should be "reversed," i.e., viewed with a 180° rotation. For the upper bound, we notice that each half is necessarily a ballot walk. For the lower bound, we need some sufficient condition on the two halves so that neither "overshoots" the other when the two halves are glued together.

Let us recall the following classic theorem about ballot sequences (e.g., see [8]).

Theorem 3.2 (Ballot Theorem)**.** *Let $p > q$. The number of ballot sequences with p 1's and q 0's, or equivalently the number of ballot walks with p steps of the form $(1,1)$ and q steps of the form $(1,-1)$, is equal to*

$$\binom{p+q-1}{p-1} - \binom{p+q-1}{p} = \frac{p-q}{p+q}\binom{p+q}{p}.$$

Corollary 3.3. *Let $0 \le a < b$ be real numbers. The number of ballot walks with n steps and whose final height is inclusively between a and b is*

$$\binom{n-1}{\lceil \frac{1}{2}(a+n) \rceil - 1} - \binom{n-1}{\lfloor \frac{1}{2}(b+n) \rfloor}.$$

Proof. We use the Ballot Theorem and sum over all (p,q) with $p+q=n$ and $a \le 2p-n \le b$ to find that the desired quantity is

$$\sum_{a \le 2p-n \le b} \left(\binom{n-1}{p-1} - \binom{n-1}{p} \right) = \binom{n-1}{\lceil \frac{1}{2}(a+n) \rceil - 1} - \binom{n-1}{\lfloor \frac{1}{2}(b+n) \rfloor}. \qquad \square$$

We will also use the following well-known fact about the normal approximation of binomial coefficients. It can be proved using either Stirling's formula or the Central Limit Theorem.

Proposition 3.4. *For any real number t,*

$$\lim_{n\to\infty} \frac{\sqrt{n}}{2^n} \binom{n}{\frac{1}{2}(n+t\sqrt{n})} = \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}t^2}. \tag{4}$$

3.1 Upper Bound

Lemma 3.5. *The number of ballot walks with n steps is* $\binom{n-1}{\lceil n/2 \rceil - 1} \sim \frac{2^n}{\sqrt{2\pi n}}$.

Proof. This follows directly from Corollary 3.3 and Proposition 3.4. □

Let $n_0 = \lfloor n/2 \rfloor$ and $n_1 = \lceil n/2 \rceil$. A bidirectional ballot walk is necessarily a ballot walk of length n_0 followed by the reverse of a ballot walk of length n_1 (though not every bidirectional ballot walk has this form). Therefore, the number of bidirectional ballot walks with n steps is at most

$$O\left(\frac{2^{n_0}}{\sqrt{n_0}}\right) O\left(\frac{2^{n_1}}{\sqrt{n_1}}\right) = O\left(\frac{2^n}{n}\right).$$

Thus we have proven the following upper bound on B_n.

Proposition 3.6. $B_n = O(2^n/n)$.

3.2 Lower Bound

We know that the first half and the reverse of the second half are both ballot walks, but this alone is not enough to guarantee that the overall walk is a bidirectional ballot walk. So we place additional constraints on each half of the walk.

Definition 3.7. Let b be a positive integer. A b-bounded walk is a ballot walk that never goes into the region $y > 2b$ and ends in the region $y > b$.

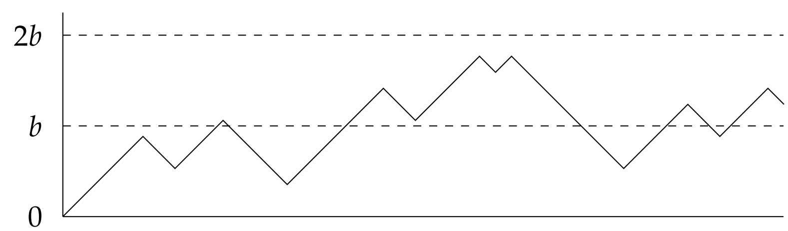

Figure 4: An example of a b-bounded walk.

Lemma 3.8. *The concatenation of a b-bounded walk followed by the reverse of another b-bounded walk is necessarily a bidirectional ballot walk.*

Figure 5 is a "proof by picture" of the lemma. The b-boundedness ensures that neither half overshoots the other.

Lemma 3.9. *The number of* $\lfloor \sqrt{n} \rfloor$*-bounded walks of n steps is* $\Omega(2^n/\sqrt{n})$.

Proof. We see that b-bounded walks of n steps are precisely ballot walks that end in the region $b+1 \le y \le 2b$ and never go into the region $y > 2b$. Using Corollary 3.3, we see that the number of ballot walks with n steps that end in $b+1 \le y \le 2b$ is equal to

$$\binom{n-1}{\lceil \frac{1}{2}(n+b-1) \rceil} - \binom{n-1}{\lfloor n/2 \rfloor + b}.$$

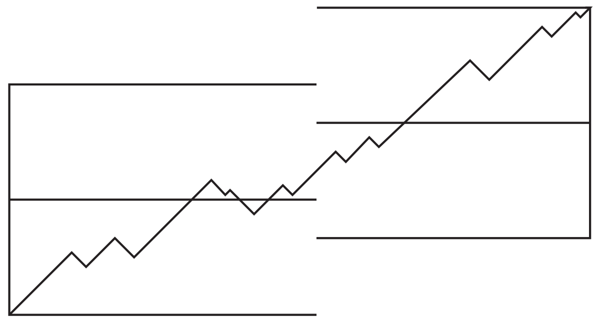

Figure 5: “Proof by picture” of Lemma 3.8.

Now we need to consider those ballot walks that end in $b < y \le 2b$ but go into $y > 2b$ at some point in the walk. Let $(t, 2b+1)$ be the last point in walk that is in the region $y > 2b$. We can reflect the portion of the walk after that point to get a ballot walk that ends in $y > 2b+1$. See Figure 6 for an illustration. This map is injective since we can always get back to the original walk, but it is not necessarily onto. Then, we know that the number of ballot walks that end in $b < y \le 2b$ but go into $y > 2b$ at some point is at most the number of ballot walks that end in $y \ge 2b+2$. By Corollary 3.3, the number of ballot walks that end in $y \ge 2b+2$ is equal to $\binom{n-1}{\lceil n/2 \rceil + b}$.

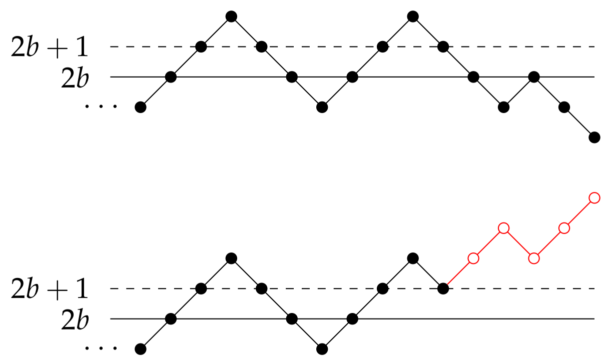

Figure 6: Reflecting the last segment of a walk.

Therefore, the number of b-bounded walks is at least

$$\binom{n-1}{\lceil \frac{1}{2}(n+b-1) \rceil} - \binom{n-1}{\lfloor n/2 \rfloor + b} - \binom{n-1}{\lceil n/2 \rceil + b}.$$

Let $b = \lfloor \sqrt{n} \rfloor$. Using Proposition 3.4, we have

$$\lim_{n\to\infty} \frac{\sqrt{n}}{2^n} \left(\binom{n-1}{\lceil \frac{1}{2}(n+b-1) \rceil} - \binom{n-1}{\lfloor n/2 \rfloor + b} - \binom{n-1}{\lceil n/2 \rceil + b} \right) = \frac{1}{\sqrt{2\pi}} (e^{-1/2} - 2e^{-2}) > 0.$$

It follows that the number of $\lfloor \sqrt{n} \rfloor$-bounded walks is $\Omega(2^n / \sqrt{n})$. □

As before, we can form bidirectional ballot walks by concatenating two b-bounded walks, where the second half is reversed. Let $n_0 = \lfloor n/2 \rfloor$ and $n_1 = \lfloor n/2 \rfloor$. Then, the number of bidirectional ballot walks is at least

$$\Omega\left(\frac{2^{n_0}}{\sqrt{n_0}}\right)\Omega\left(\frac{2^{n_1}}{\sqrt{n_1}}\right)=\Omega(2^n/n).$$

Thus we have proven the following.

Proposition 3.10. $B_n=\Omega(2^n/n)$.

Propositions 3.6 and 3.10 together complete the proof of Proposition 3.1 and hence also Theorem 2.2.

4 Proof of $B_n \sim 2^{n-2}/n$

In this section prove Theorem 1.2 which states that $B_n \sim 2^{n-2}/n$. The proof is independent from the work done in previous sections. It involves first obtaining an exact formula for B_n and then evaluating the limit of $nB_n/2^n$.

4.1 A formula for B_n

In this subsection, we derive a formula for B_n. We also introduce some notation for subsequent analysis.

In this section we will use "lattice walk" to mean a walk with steps $(1,1)$ or $(1,-1)$. Let $N_{n,x}$ denote the the number of lattice walks from $(0,0)$ to (n,x), with no other restrictions. It is easy to see that

$$N_{n,x}=\binom{n}{\frac{1}{2}(n+x)} \quad \text{whenever } n+x \text{ is even.} \tag{5}$$

If $n+x$ is odd, then $N_{n,x}=0$.

We quote the following result from Feller [1, III.10, Prob. 3, p. 96].

Proposition 4.1. *Let a and b be positive integers, and $-b<c<a$. The number of lattice walks from $(0,0)$ to (n,c) which meet neither the line $y=-b$ nor $y=a$ is given by the series*

$$\sum_{k=-\infty}^{\infty}\left(N_{n,2k(a+b)+c}-N_{n,2k(a+b)+2a-c}\right).$$

Note that the summation only contains finitely many non-zero terms.

Proposition 4.1 is connected to the Gambler's ruin problem. Although we do not prove the proposition here, we mention that it can be proven using repeated applications of the reflection principle, restated below for readers' interest. Also see Figure 7.

Proposition 4.2 (Reflection principle)**.** *Let A and B be integral points above the x-axis, and let A' be the reflection of A across the x-axis. Then the number of lattice walks from A to B that touch or cross the x-axis at some point equals the number of lattice walks from A' to B.*

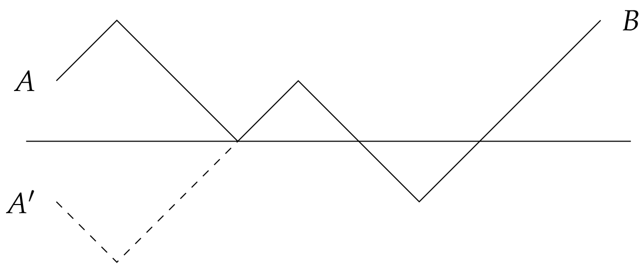

Figure 7: An illustration of the reflection principle.

After a translation by $(-1,-1)$ and the removal of the first and last steps, a bidirectional ballot walk from $(0,0)$ to $(n, c+2)$ is equivalent to a lattice walk from $(0,0)$ to $(n-2, c)$ that avoids the lines $x=-1$ and $x=c+1$, where c is a non-negative integer with the same parity as n. Let $B_n(c)$ denote the number of bidirectional ballot walks from $(0,0)$ to $(n, c+2)$. Using Proposition 4.1, we have

$$B_n(c) = \sum_{k=-\infty}^{\infty} \left(N_{n-2,(2k+1)(c+2)-2} - N_{n-2,(2k+1)(c+2)} \right). \tag{6}$$

Let us group together terms indexed by k and $-k-1$. Using $N_{n,x} = N_{n,-x}$, we obtain

$$B_n(c) = \sum_{k=0}^{\infty} T(n,c,k), \tag{7}$$

where

$$\begin{aligned} T(n,c,k) &= N_{n-2,(2k+1)(c+2)-2} - 2N_{n-2,(2k+1)(c+2)} + N_{n-2,(2k+1)(c+2)+2} \\ &= \frac{(2k+1)^2(c+2)^2 - n}{n(n-1)} \binom{n}{\frac{1}{2}\left(n + (2k+1)(c+2)\right)}. \end{aligned}$$

Note that $T(n,c,k)$ is only defined for $n \equiv c \pmod 2$. The second equality follows from a straightforward computation. Thus we obtain the following formula for B_n,

$$B_n = \sum_{\substack{c \geq 0 \\ c \equiv n (2)}} B_n(c) = \sum_{\substack{c \geq 0 \\ c \equiv n (2)}} \sum_{k \geq 0} T(n,c,k). \tag{8}$$

Recall that our goal is to evaluate $\lim_{n\to\infty} nB_n/2^n$. We do so by first restricting to the subset of bidirectional ballot walks whose final height lies in a certain interval. Let

$$B_n(\alpha\sqrt{n}, \beta\sqrt{n}) = \sum_{\substack{\alpha\sqrt{n} < c < \beta\sqrt{n} \\ c \equiv n (2)}} B_n(c). \tag{9}$$

Our attack on B_n starts with $B_n(\alpha\sqrt{n}, \beta\sqrt{n})$.

4.2 Outline

Here we outline the main steps of our proof of Theorem 1.2.

First, we would like to determine

$$\lim_{n\to\infty} \frac{n}{2^n} B_n(\alpha\sqrt{n}, \beta\sqrt{n}) = \lim_{n\to\infty} \sum_{\substack{c\equiv n(2) \\ \alpha\sqrt{n}<c<\beta\sqrt{n}}} \sum_{k\geq 0} \frac{n}{2^n} T(n,c,k) \tag{10}$$

for $0 < \alpha < \beta$ (Lemma 4.5). For any fixed k, we can evaluate the limit

$$\lim_{n\to\infty} \sum_{\substack{c\equiv n(2) \\ \alpha\sqrt{n}<c<\beta\sqrt{n}}} \frac{n}{2^n} T(n,c,k)$$

(Lemma 4.3). Next, we switch the sum over k with the limit on n to see that (10) is equal to

$$\sum_{k\geq 0} \left(\lim_{n\to\infty} \sum_{\substack{c\equiv n(2) \\ \alpha\sqrt{n}<c<\beta\sqrt{n}}} \frac{n}{2^n} T(n,c,k) \right).$$

To justify the switching of limits, we bound the tail of the sum over k,

$$\sum_{\substack{c\equiv n(2) \\ \alpha\sqrt{n}<c<\beta\sqrt{n}}} \sum_{k\geq \gamma} \frac{n}{2^n} T(n,c,k),$$

where γ is a sufficiently large (depending on α) positive integer (Lemma 4.4). This would complete the evaluation of the limit in (10). Next, we would like to use the limit in (10) to evaluate the limit of $nB_n/2^n$ as $n\to\infty$. This involves the following switching of limits

$$\lim_{n\to\infty} \frac{n}{2^n} B_n = \lim_{n\to\infty} \lim_{\substack{\alpha\to 0 \\ \beta\to\infty}} \frac{n}{2^n} B_n(\alpha\sqrt{n}, \beta\sqrt{n}) = \lim_{\substack{\alpha\to 0 \\ \beta\to\infty}} \lim_{n\to\infty} \frac{n}{2^n} B_n(\alpha\sqrt{n}, \beta\sqrt{n}). \tag{11}$$

We know how to evaluate the final limit (Lemma 4.8). However, once again, this limit switch needs to be justified. We do so by bounding the quantities $B_n(0, \alpha\sqrt{n}]$ and $B_n[\beta\sqrt{n}, \infty)$, which are defined to be the number of n-step bidirectional ballot walks whose ending point has height in the interval $(0, \alpha\sqrt{n}]$ and $[\beta\sqrt{n}, \infty)$ respectively (Lemma 4.6). The argument used to bound $B_n(0, \alpha\sqrt{n}]$ and $B_n[\beta\sqrt{n}, \infty)$ is in the same spirit as the arguments used to prove the weaker result $B_n = \Theta(2^n/n)$; we observe that a bidirectional ballot sequence is necessarily the concatenation of two ballot sequences (the second one reversed) and then bound each half separately.

Once we can justify (11) and perform all the limit evaluations, the proof of Theorem 1.2 is complete.

4.3 Analysis

Now we prove Theorem 1.2 through a series of lemmas as outlined in the previous section.

Lemma 4.3. *For fixed* $0 < \alpha < \beta$ *and* k,

$$\lim_{n\to\infty} \sum_{\substack{\alpha\sqrt{n}<c<\beta\sqrt{n} \\ c\equiv n(2)}} \frac{n}{2^n} T(n,c,k) = \frac{1}{\sqrt{2\pi}} \left(\alpha e^{-\frac{1}{2}(2k+1)^2\alpha^2} - \beta e^{-\frac{1}{2}(2k+1)^2\beta^2} \right).$$

Proof. It is known (see [1, p. 180]) that if $x = O(\sqrt{n})$, then

$$\binom{n}{\frac{1}{2}(n+x)} = \binom{n}{\lfloor n/2 \rfloor} e^{-x^2/2n} \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right). \tag{12}$$

So, for $\alpha\sqrt{n} < c < \beta\sqrt{n}$, we have

$$\begin{aligned} T(n,c,k) &= \frac{(2k+1)^2(c+2)^2 - n}{n(n-1)} \binom{n}{\frac{1}{2}\left(n + (2k+1)(c+2)\right)} \\ &= \frac{(2k+1)^2(c+2)^2 - n}{n(n-1)} \binom{n}{\lfloor n/2 \rfloor} \exp\left(-\frac{(2k+1)^2(c+2)^2}{2n}\right) \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right), \end{aligned}$$

where the constants in the big-O depend on k and β. It follows that

$$\begin{aligned} &\sum_{\substack{\alpha\sqrt{n}<c<\beta\sqrt{n} \\ c\equiv n(2)}} \frac{n}{2^n} T(n,c,k) \\ &= \frac{n}{2^n} \binom{n}{\lfloor n/2 \rfloor} \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right) \sum_{\substack{c\equiv n(2) \\ \alpha\sqrt{n}<c<\beta\sqrt{n}}} \frac{(2k+1)^2(c+2)^2 - n}{n(n-1)} \exp\left(-\frac{(2k+1)^2(c+2)^2}{2n}\right). \end{aligned}$$

From Proposition 3.4 we have

$$\lim_{n\to\infty} \frac{\sqrt{n}}{2^n} \binom{n}{\lfloor n/2 \rfloor} = \sqrt{\frac{2}{\pi}}.$$

So, it suffices to prove that

$$\lim_{n\to\infty} \sum_{\substack{\alpha\sqrt{n}<c<\beta\sqrt{n} \\ c\equiv n(2)}} \frac{(2k+1)^2(c+2)^2 - n}{n\sqrt{n}} \exp\left(-\frac{(2k+1)^2(c+2)^2}{2n}\right) = \frac{1}{2} \left(\alpha e^{-\frac{1}{2}(2k+1)^2\alpha^2} - \beta e^{-\frac{1}{2}(2k+1)^2\beta^2} \right),$$

which follows directly from the Riemann sum approximation of the integral

$$\int_\alpha^\beta ((2k+1)^2x^2 - 1) e^{-\frac{1}{2}(2k+1)^2x^2}\, dx = \alpha e^{-\frac{1}{2}(2k+1)^2\alpha^2} - \beta e^{-\frac{1}{2}(2k+1)^2\beta^2}. \qquad \square$$

Lemma 4.4. *For fixed* $0 < \alpha < \beta$, *and any positive integers* $\gamma > \frac{\sqrt{3}}{2\alpha} + \frac{1}{2}$ *and* $n \geq 2$, *we have* $T(n,c,k) \geq$

0 *for all* $\alpha\sqrt{n} < c < \beta\sqrt{n}$ *and* $k \geq \gamma$*, and also*

$$\limsup_{n\to\infty} \sum_{\substack{c\equiv n(2)\\ \alpha\sqrt{n}<c<\beta\sqrt{n}}} \sum_{k\geq\gamma} \frac{n}{2^n} T(n,c,k) \leq \frac{(\beta-\alpha)(2\gamma-1)}{\sqrt{8\pi}} e^{-\frac{1}{2}(2\gamma-1)^2\alpha^2}.$$

Proof. For $k \geq \gamma$ and $c > \alpha\sqrt{n}$, we have

$$(2k+1)^2(c+2)^2 > (2\gamma+1)^2(\alpha\sqrt{n}+2)^2 > \left(\frac{\sqrt{3}}{\alpha}\right)^2 \alpha^2 n > n,$$

so that

$$T(n,c,k) = \frac{(2k+1)^2(c+2)^2 - n}{n(n-1)} \binom{n}{\frac{1}{2}\left(n+(2k+1)(c+2)\right)} \geq 0.$$

Consider the following function defined for $n+a$ even,

$$f(n,a) = \frac{a^2-n}{n(n-1)} \binom{n}{\frac{1}{2}(n+a)}.$$

When n is fixed, $f(n,a)$ is monotonically non-increasing with respect to a as long as $(a+1)^2 > 3n$. Indeed,

$$\begin{aligned}
f(n,a) - f(n,a+2) &= \frac{a^2-n}{n(n-1)}\binom{n}{\frac{1}{2}(n+a)} - \frac{(a+2)^2-n}{n(n-1)}\binom{n}{\frac{1}{2}(n+a+2)} \\
&= \frac{(a^2-n)(n+a+2)}{2n(n-1)(n+1)}\binom{n+1}{\frac{1}{2}(n+a+2)} - \frac{((a+2)^2-n)(n-a)}{2n(n-1)(n+1)}\binom{n+1}{\frac{1}{2}(n+a+2)} \\
&= \frac{(a+1)((a+1)^2-3n-1)}{n(n-1)(n+1)}\binom{n+1}{\frac{1}{2}(n+a+2)},
\end{aligned}$$

which is nonnegative whenever $(a+1)^2 > 3n$.

Note that

$$T(n,c,k) = f(n,(2k+1)(c+2)).$$

Since $f(n,a)$ is non-increasing with respect to a, we have

$$(c+2)T(n,c,k) \leq \sum_{\substack{a\equiv n(2)\\ (2k-1)(c+2)<a\leq(2k+1)(c+2)}} f(n,a) \tag{13}$$

as long as $(2k-1)^2(c+2)^2 > 3n$. Also,

$$f(n,a) = \frac{a-1}{n-1}\binom{n-1}{\frac{1}{2}(n+a-2)} - \frac{a+1}{n-1}\binom{n-1}{\frac{1}{2}(n+a)}. \tag{14}$$

The restrictions $\gamma > \frac{\sqrt{3}}{2\alpha} + \frac{1}{2}$ and $c > \alpha\sqrt{n}$ imply that

$$(2\gamma - 1)^2(c+2)^2 > \left(\frac{\sqrt{3}}{\alpha}\right)^2 (\alpha\sqrt{n}+2)^2 > 3n.$$

So (13) holds for any $k \geq \gamma$. Using (13) and then evaluating a telescoping sum through (14), we obtain

$$\begin{aligned}
\sum_{k\geq\gamma} T(n,c,k) &\leq \frac{1}{c+2} \sum_{\substack{a\equiv n(2) \\ a>(2\gamma-1)(c+2)}} f(n,a) \\
&= \frac{1}{c+2} \cdot \frac{(2\gamma-1)(c+2)+1}{n-1} \binom{n-1}{\frac{1}{2}(n+(2\gamma-1)(c+2))} \\
&= \left(\frac{2\gamma-1}{n-1} + \frac{1}{(c+2)(n-1)}\right) \binom{n-1}{\frac{1}{2}(n+(2\gamma-1)(c+2))}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\sum_{\substack{c\equiv n(2) \\ \alpha\sqrt{n}<c<\beta\sqrt{n}}} \sum_{k\geq\gamma} T(n,c,k) &\leq \sum_{\substack{c\equiv n(2) \\ \alpha\sqrt{n}<c<\beta\sqrt{n}}} \left(\frac{2\gamma-1}{n-1} + \frac{1}{(c+2)(n-1)}\right) \binom{n-1}{\frac{1}{2}(n+(2\gamma-1)(c+2))} \\
&\leq \frac{1}{2}(\beta\sqrt{n} - \alpha\sqrt{n} + 2) \left(\frac{2\gamma-1}{n-1} + \frac{1}{(\lceil\alpha\sqrt{n}\rceil+2)(n-1)}\right) \binom{n-1}{\frac{1}{2}(n+(2\gamma-1)(\lceil\alpha\sqrt{n}\rceil+2))}
\end{aligned}$$

where the last step follows from the observation that the largest term occurs when c is the smallest. Finally, multiplying by $n/2^n$, letting $n\to\infty$, and using Proposition 3.4 we obtain the second claim in the lemma. □

Lemma 4.5. *For any* $0 < \alpha < \beta$,

$$\lim_{n\to\infty} \frac{n}{2^n} B_n(\alpha\sqrt{n}, \beta\sqrt{n}) = \frac{1}{\sqrt{2\pi}} \sum_{k\geq 0} \left(\alpha e^{-\frac{1}{2}(2k+1)^2\alpha^2} - \beta e^{-\frac{1}{2}(2k+1)^2\beta^2}\right). \tag{15}$$

Proof. From the definition, we have

$$B_n(\alpha\sqrt{n}, \beta\sqrt{n}) = \sum_{\substack{c\equiv n(2) \\ \alpha\sqrt{n}<c<\beta\sqrt{n}}} \sum_{k\geq 0} T(n,c,k).$$

For any positive integer $\gamma > \frac{\sqrt{3}}{2\alpha} + \frac{1}{2}$, by Lemma 4.4, the terms with $k \geq \gamma$ are all nonnegative. Also, using Lemma 4.3, we have

$$\lim_{n\to\infty} \sum_{\substack{c\equiv n(2) \\ \alpha\sqrt{n}<c<\beta\sqrt{n}}} \sum_{k=0}^{\gamma-1} \frac{n}{2^n} T(n,c,k) = \sum_{k=0}^{\gamma-1} \frac{1}{\sqrt{2\pi}} \left(\alpha e^{-\frac{1}{2}(2k+1)^2\alpha^2} - \beta e^{-\frac{1}{2}(2k+1)^2\beta^2}\right). \tag{16}$$

It follows that

$$\liminf_{n\to\infty} \frac{n}{2^n} B_n(\alpha\sqrt{n}, \beta\sqrt{n}) \geq \sum_{k=0}^{\gamma-1} \frac{1}{\sqrt{2\pi}} \left(\alpha e^{-\frac{1}{2}(2k+1)^2\alpha^2} - \beta e^{-\frac{1}{2}(2k+1)^2\beta^2} \right)$$

for every sufficiently large γ, and hence

$$\liminf_{n\to\infty} \frac{n}{2^n} B_n(\alpha\sqrt{n}, \beta\sqrt{n}) \geq \sum_{k\geq 0} \frac{1}{\sqrt{2\pi}} \left(\alpha e^{-\frac{1}{2}(2k+1)^2\alpha^2} - \beta e^{-\frac{1}{2}(2k+1)^2\beta^2} \right). \tag{17}$$

Now we give an upper bound. We have

$$\frac{n}{2^n} B_n(\alpha\sqrt{n}, \beta\sqrt{n}) = \sum_{\substack{c\equiv n(2) \\ \alpha\sqrt{n}<c<\beta\sqrt{n}}} \sum_{k=0}^{\gamma-1} \frac{n}{2^n} T(n,c,k) + \sum_{\substack{c\equiv n(2) \\ \alpha\sqrt{n}<c<\beta\sqrt{n}}} \sum_{k\geq\gamma} \frac{n}{2^n} T(n,c,k). \tag{18}$$

Using Lemma 4.4, for sufficiently large γ, we have

$$\limsup_{n\to\infty} \sum_{\substack{c\equiv n(2) \\ \alpha\sqrt{n}<c<\beta\sqrt{n}}} \sum_{k\geq\gamma} \frac{n}{2^n} T(n,c,k) \leq \frac{(\beta-\alpha)(2\gamma-1)}{\sqrt{8\pi}} e^{-\frac{1}{2}(2\gamma-1)^2\alpha^2}. \tag{19}$$

Combining (16) and (19) and using (18) we get

$$\begin{aligned} &\limsup_{n\to\infty} \frac{n}{2^n} B_n(\alpha\sqrt{n}, \beta\sqrt{n}) \\ &\qquad \leq \frac{(\beta-\alpha)(2\gamma-1)}{\sqrt{8\pi}} e^{-\frac{1}{2}(2\gamma-1)^2\alpha^2} + \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\gamma-1} \left(\alpha e^{-\frac{1}{2}(2k+1)^2\alpha^2} - \beta e^{-\frac{1}{2}(2k+1)^2\beta^2} \right). \end{aligned}$$

Since γ can be arbitrarily large, letting $\gamma \to \infty$ gives

$$\limsup_{n\to\infty} \frac{n}{2^n} B_n(\alpha\sqrt{n}, \beta\sqrt{n}) \leq \frac{1}{\sqrt{2\pi}} \sum_{k\geq 0} \left(\alpha e^{-\frac{1}{2}(2k+1)^2\alpha^2} - \beta e^{-\frac{1}{2}(2k+1)^2\beta^2} \right). \tag{20}$$

Combining (17) and (20), we obtain

$$\lim_{n\to\infty} \frac{n}{2^n} B_n(\alpha\sqrt{n}, \beta\sqrt{n}) = \frac{1}{\sqrt{2\pi}} \sum_{k\geq 0} \left(\alpha e^{-\frac{1}{2}(2k+1)^2\alpha^2} - \beta e^{-\frac{1}{2}(2k+1)^2\beta^2} \right). \qquad \square$$

Let $B_n(0, \alpha\sqrt{n}]$ and $B_n[\beta\sqrt{n}, \infty)$ denote the number of bidirectional ballot walks that end at a height in the intervals $(0, \alpha\sqrt{n}]$ and $[\beta\sqrt{n}, \infty)$, respectively. Next we would like to bound these two quantities.

Lemma 4.6. *For $0 < \alpha < \beta$, we have*

$$\limsup_{n\to\infty} \frac{n}{2^n} B_n(0, \alpha\sqrt{n}] \leq \frac{1}{\pi} \left(1 - e^{-\alpha^2}\right)^2. \tag{21}$$

and

$$\limsup_{n\to\infty} \frac{n}{2^n} B_n[\beta\sqrt{n}, \infty) \leq \frac{2}{\pi} e^{-\frac{1}{4}\beta^2}. \tag{22}$$

Proof. Let $n_0 = \lfloor n/2 \rfloor$ and $n_1 = \lceil n/2 \rceil$. To prove (21) we observe that a bidirectional ballot walk of length n that ends at height at most $\alpha\sqrt{n}$ is necessarily the concatenation of a ballot walk of length n_0 that ends at height at most $\alpha\sqrt{n}$ with the reverse of a ballot walk of length n_1 that ends at height at most $\alpha\sqrt{n}$. Using Corollary 3.3, we see that the number of ballot walks with n_0 steps that end at height at most $\alpha\sqrt{n}$ is

$$\binom{n_0 - 1}{\lceil \frac{1}{2} n_0 \rceil - 1} - \binom{n_0 - 1}{\lfloor \frac{1}{2}(n_0 + \alpha\sqrt{n}) \rfloor} \tag{23}$$

and similarly with n_1. Then (21) follows from applying Proposition 3.4 to (23).

Similarly, a bidirectional ballot walk of length n that ends at height at least $\beta\sqrt{n}$ is necessarily the concatenation of a ballot walk of length $\lfloor n/2 \rfloor$ with the reverse of a ballot walk of length $\lceil n/2 \rceil$, where at least one of the two walks end at height at least $\frac{1}{2}\beta\sqrt{n}$. Using Corollary 3.3 and Proposition 3.4 in a similar way gives us (22). We omit the details. □

The following lemma shows that the sum in (15) is absolutely convergent (when we split up the terms in each parenthesis) and analyzes it through Riemann sum approximations.

Lemma 4.7. *For any α, the sum*

$$\sum_{k\geq 0} \alpha e^{-\frac{1}{2}(2k+1)^2\alpha^2}$$

is finite. Also we have the following limits:

$$\lim_{\alpha\to 0} \frac{1}{\sqrt{2\pi}} \sum_{k\geq 0} \alpha e^{-\frac{1}{2}(2k+1)^2\alpha^2} = \frac{1}{4}, \quad \textit{and} \quad \lim_{\beta\to\infty} \frac{1}{\sqrt{2\pi}} \sum_{k\geq 0} \beta e^{-\frac{1}{2}(2k+1)^2\beta^2} = 0.$$

Proof. The function $x \mapsto e^{-\frac{1}{2}x^2}$ is monotonically decreasing for $x \geq 0$, so the lower Riemann sum (with width 2α) gives us

$$\sum_{k\geq 0} \alpha e^{-\frac{1}{2}(2k+1)^2\alpha^2} \leq \alpha e^{-\frac{1}{2}\alpha^2} + \sum_{k\geq 1} \alpha e^{-\frac{1}{2}(2k+1)^2\alpha^2} \leq \alpha e^{-\frac{1}{2}\alpha^2} + \frac{1}{2}\int_{\alpha}^{\infty} e^{-\frac{1}{2}x^2}\, dx < \infty.$$

This shows that the sum is finite. Next, by the Riemann sum approximation, we have

$$\lim_{\alpha\to 0} \frac{1}{\sqrt{2\pi}} \sum_{k\geq 0} \alpha e^{-\frac{1}{2}(2k+1)^2\alpha^2} = \frac{1}{2\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{1}{2}x^2}\, dx = \frac{1}{4}.$$

For the other limit, using the lower Riemann sum, we have

$$\begin{aligned}\sum_{k\geq 0}\beta e^{-\frac{1}{2}(2k+1)^2\beta^2} &\leq \beta e^{-\frac{1}{2}\beta^2}+\sum_{k\geq 1}\beta e^{-\frac{1}{2}(2k+1)^2\beta^2}\\ &\leq \beta e^{-\frac{1}{2}\beta^2}+\frac{1}{2}\int_{\beta}^{\infty} e^{-\frac{1}{2}x^2}\,dx\\ &\to 0,\quad \text{as } \beta\to\infty.\end{aligned}$$

This completes the proof of the limits. □

Lemma 4.8. $\lim\limits_{\substack{\alpha\to 0\\ \beta\to\infty}}\lim\limits_{n\to\infty}\dfrac{n}{2^n}B_n(\alpha\sqrt{n},\beta\sqrt{n})=\dfrac{1}{4}.$

Proof. This follows from Lemmas 4.5 and 4.7. □

Now we finally have come to the proof that $B_n\sim 2^{n-2}/n$.

Proof of Theorem 1.2. For $0<\alpha<\beta$, we have

$$\frac{n}{2^n}B_n(\alpha\sqrt{n},\beta\sqrt{n})\leq\frac{n}{2^n}B_n=\frac{n}{2^n}B_n(\alpha\sqrt{n},\beta\sqrt{n})+\frac{n}{2^n}B_n(0,\alpha\sqrt{n}]+\frac{n}{2^n}B_n[\beta\sqrt{n},\infty).$$

Letting $n\to\infty$ and using Lemmas 4.5 and 4.6, we obtain

$$\lim_{n\to\infty}\frac{n}{2^n}B_n(\alpha\sqrt{n},\beta\sqrt{n})\leq\lim_{n\to\infty}\frac{n}{2^n}B_n\leq\lim_{n\to\infty}\frac{n}{2^n}B_n(\alpha\sqrt{n},\beta\sqrt{n})+\frac{1}{\pi}\left(1-e^{-\alpha^2}\right)^2+\frac{2}{\pi}e^{-\frac{1}{4}\beta^2}.$$

Letting $\alpha\to 0,\beta\to\infty$ and using Lemma 4.8, we get

$$\lim_{n\to\infty}\frac{n}{2^n}B_n=\frac{1}{4},$$

as desired. □

5 Higher dimensional generalization

Now we turn our attention to walks on the d-dimensional lattice $\mathbb{Z}^d$ with unit steps $\pm e_i$, where e_i is some unit coordinate vector. The following definition gives a higher dimensional generalization of bidirectional ballot walks.

Definition 5.1. A *d-confined* walk is a walk on $\mathbb{Z}^d$ with unit steps such that, among all the points visited by the walk, the starting point is the coordinate-wise minimum and the ending point is the coordinate-wise maximum. In other words, if the walk goes from $(x_1,\ldots,x_d)$ to $(y_1,\ldots,y_d)$, then the walk is confined if $x_i\leq y_i$ for each i and the walk stays inside the box $[x_1,y_1]\times\cdots\times[x_d,y_d]$. The number of d-confined walks with n steps starting at the origin is denoted $C_{d,n}$.

For example, a 1-confined walk of n steps is equivalent to a bidirectional ballot walk of $n+2$ steps with the first and last steps removed, so $C_{1,n} = B_{n+2}$. See Figure 8 for an example of a 2-confined walk and Table 3 for some values of $C_{d,n}$.

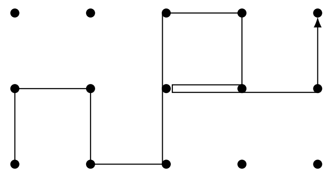

Figure 8: A 2-confined walk of 12 steps from $(0,0)$ to $(4,2)$.

Table 3: Some values of $C_{d,n}$.

$C_{d,n}$	n										
d	0	1	2	3	4	5	6	7	8	9	10
1	1	1	1	2	3	5	9	15	28	49	91
2	1	2	4	10	28	80	248	786	2550	8486	28522
3	1	3	9	30	111	435	1797	7773	34698	159549	750843
4	1	4	16	68	312	1520	7776	41556	230284	1316260	7730164
5	1	5	25	130	715	4145	25145	158835	1039800	7023425	48777355
6	1	6	36	222	1428	9600	67224	488550	3672258	28454394	226606446
7	1	7	49	350	2583	19775	156933	1288161	10909990	95103421	851248027
8	1	8	64	520	4336	37280	330752	3025128	28477848	275426984	2731751848
9	1	9	81	738	6867	65565	643401	6488631	67188564	713451825	7758117819
10	1	10	100	1010	10380	109040	1173240	12936570	146126350	1689544750	19975761010

The following result is the generalization of Theorem 1.2 for d-confined walks.

Theorem 5.2. *For a fixed positive integer d,*

$$\lim_{n\to\infty} \frac{n^d C_{d,n}}{2^n d^{n+d}} = 1$$

In other words, for a fixed d, the probability that an n-step random walk on $\mathbb{Z}^d$ is d-confined is

$$\frac{C_{d,n}}{(2d)^n} \sim \left(\frac{d}{n}\right)^d, \text{ as } n \to \infty.$$

The remainder of this section contain the proof of Theorem 5.2. First, let us give a formula for $C_{d,n}$ in terms of the number of bidirectional ballot walks B_n.

Proposition 5.3. *For any positive integers d and n, we have*

$$C_{d,n} = \sum_{\substack{a_1,\ldots,a_d \\ \sum a_i = n}} \binom{n}{a_1,\ldots,a_d} B_{a_1+2}\cdots B_{a_d+2},$$

where $\binom{n}{a_1,\ldots,a_d}$ is a multinomial coefficient. Equivalently, for every positive integer d, we have the generating *function identity*

$$\sum_{n\geq 0} C_{d,n}\frac{x^n}{n!} = \left(\sum_{n\geq 0} B_{n+2}\frac{x^n}{n!}\right)^d.$$

Proof. Consider the set of d-confined walks with a_i steps that change the i-th coordinate for each i. The steps that change the i-th coordinate form a walk on a line where the starting point is minimum and the ending point is maximum, and hence there are B_{a_i+2} choices. There are $\binom{n}{a_1,\dots,a_d}$ ways to combine the coordinates together to form one walk on $\mathbb{Z}^d$. The formula then follows. $\square$

Since

$$\binom{n}{a_1,\dots,a_d} = \binom{n+d}{a_1+1,\dots,a_d+1}\frac{(a_1+1)\cdots(a_d+1)}{(n+1)(n+2)\cdots(n+d)},$$

we have

$$\begin{aligned}\frac{n^d C_{d,n}}{2^n d^{n+d}} &= \frac{n^d}{d^{n+d}}\sum_{\substack{a_1,\dots,a_d\\ \sum a_i=n}}\binom{n}{a_1,\dots,a_d}\prod_{i=1}^d\frac{B_{a_i+2}}{2^{a_i}}\\ &= \frac{n^d}{(n+1)(n+2)\cdots(n+d)d^{n+d}}\sum_{\substack{a_1,\dots,a_d\\ \sum a_i=n}}\binom{n+d}{a_1+1,\dots,a_d+1}\prod_{i=1}^d\frac{(a_i+1)B_{a_i+2}}{2^{a_i}}. \qquad (24)\end{aligned}$$

From Theorem 1.2 we know that $(a_i+1)B_{a_i+2}2^{-a_i}$ is close to 1. The next lemma shows that this approximation gives us the correct answer. Afterwards, we justify this approximation.

Lemma 5.4. *For a positive integer d,*

$$\lim_{n\to\infty}\frac{1}{d^{n+d}}\sum_{\substack{a_1,\dots,a_d\geq 0\\ \sum a_i=n}}\binom{n+d}{a_1+1,\dots,a_d+1} = 1.$$

Proof. The sum

$$S = \sum_{\substack{a_1,\dots,a_d\geq 0\\ \sum a_i=n}}\binom{n+d}{a_1+1,\dots,a_d+1}$$

equals to the number surjective maps $f\colon \{1,2,\dots,n+d\}\to\{1,2,\dots,d\}$. If we drop the surjective condition, then there are d^{n+d} maps, so $S\leq d^{n+d}$. The number of non-surjective maps is at most $d(d-1)^{n+d}$, since there are exactly $(d-1)^{n+d}$ maps that miss any fixed element of $\{1,2,\dots,d\}$. Therefore

$$d^{n+d} - d(d-1)^{n+d} \leq S \leq d^{n+d}.$$

The lemma follows after letting $n\to\infty$. $\square$

Lemma 5.5. *For positive integers d and γ,*

$$\lim_{n\to\infty}\frac{1}{d^{n+d}}\sum_{\substack{a_1,\dots,a_d\geq 0\\ \sum a_i=n\\ a_1<\gamma}}\binom{n+d}{a_1+1,\dots,a_d+1} = 0.$$

Proof. The sum

$$S = \sum_{\substack{a_1,\dots,a_d \geq 0 \\ \sum a_i = n \\ a_1 < \gamma}} \binom{n+d}{a_1+1,\dots,a_d+1}$$

equals to the number of surjective maps $f\colon \{1,2,\dots,n+d\} \to \{1,2,\dots,d\}$ such that $\left|f^{-1}(1)\right| \leq \gamma$. The number of (not necessarily surjective) maps with $\left|f^{-1}(1)\right| = k$ is $\binom{n+d}{k}(d-1)^{n+d-k}$. Therefore, assuming that $n+d \geq 2\gamma$, we have

$$S \leq \sum_{k=1}^{\gamma} \binom{n+d}{k}(d-1)^{n+d-k} \leq \gamma \binom{n+d}{\gamma}(d-1)^{n+d-1} \leq \gamma(n+d)^{\gamma}(d-1)^{n+d-1}.$$

Then,

$$\frac{S}{d^{n+d}} \leq \frac{\gamma(n+d)^{\gamma}(d-1)^{n+d-1}}{d^{n+d}}.$$

The lemma follows after letting $n \to \infty$. □

Lemma 5.6. *Let d be a positive integer. Suppose that $\epsilon\colon \mathbb{Z}_{\geq 0} \to \mathbb{R}$ is some function such that $\epsilon(k) \to 0$ as $k \to 0$. Then*

$$\lim_{n\to\infty} \frac{1}{d^{n+d}} \sum_{\substack{a_1,\dots,a_d \geq 0 \\ \sum a_i = n}} \binom{n+d}{a_1+1,\dots,a_d+1} \epsilon(a_1) = 0.$$

Proof. Let $E = \max_k |\epsilon(k)|$, which is finite since $\epsilon(k) \to 0$ as $k \to \infty$. For any positive integer γ, we have

$$\begin{aligned}
&\frac{1}{d^{n+d}} \sum_{\substack{a_1,\dots,a_d \geq 0 \\ \sum a_i = n}} \binom{n+d}{a_1+1,\dots,a_d+1} |\epsilon(a_1)| \\
&\leq \frac{E}{d^{n+d}} \sum_{\substack{a_1,\dots,a_d \geq 0 \\ \sum a_i = n \\ a_i < \gamma}} \binom{n+d}{a_1+1,\dots,a_d+1} + \frac{\max_{k\geq\gamma} |\epsilon(k)|}{d^{n+d}} \sum_{\substack{a_1,\dots,a_d \geq 0 \\ \sum a_i = n}} \binom{n+d}{a_1+1,\dots,a_d+1} \\
&\leq \frac{E}{d^{n+d}} \sum_{\substack{a_1,\dots,a_d \geq 0 \\ \sum a_i = n \\ a_i < \gamma}} \binom{n+d}{a_1+1,\dots,a_d+1} + \max_{k\geq\gamma} |\epsilon(k)|.
\end{aligned}$$

Now we use Lemma 5.5. Keeping γ fixed and letting $n \to \infty$, we get

$$\limsup_{n\to\infty} \frac{1}{d^{n+d}} \sum_{\substack{a_1,\dots,a_d \geq 0 \\ \sum a_i = n}} \binom{n+d}{a_1+1,\dots,a_d+1} |\epsilon(a_1)| \leq \max_{k\geq\gamma} |\epsilon(k)|.$$

Since $\epsilon(k) \to 0$ as $k \to \infty$, the RHS can be made arbitrarily small by choosing sufficiently large γ. Thus, the LHS must be zero. □

Proof of Theorem 5.2. For every nonnegative integer k, let

$$\epsilon(k) = \frac{(k+1)B_{k+2}}{2^k} - 1.$$

From Theorem 1.2 we know that $\epsilon(k) \to 0$ as $k \to \infty$. In (24) we have

$$\lim_{n\to\infty} \frac{n^d}{(n+1)(n+2)\cdots(n+d)} = 1,$$

so it suffices to show that the following quantity approaches 1 as $n \to \infty$,

$$\begin{aligned} &\frac{1}{d^{n+d}} \sum_{\substack{a_1,\ldots,a_d \\ \sum a_i = n}} \binom{n+d}{a_1+1,\ldots,a_d+1} \prod_{i=1}^{d} \frac{(a_i+1)B_{a_i+2}}{2^{a_i}} \\ &\qquad = \frac{1}{d^{n+d}} \sum_{\substack{a_1,\ldots,a_d \\ \sum a_i = n}} \binom{n+d}{a_1+1,\ldots,a_d+1} (1+\epsilon(a_1))\cdots(1+\epsilon(a_d)). \end{aligned} \tag{25}$$

Lemma 5.4 reduces this task to showing that

$$\frac{1}{d^{n+d}} \sum_{\substack{a_1,\ldots,a_d \\ \sum a_i = n}} \binom{n+d}{a_1+1,\ldots,a_d+1} \left((1+\epsilon(a_1))\cdots(1+\epsilon(a_d)) - 1\right) \to 0, \quad \text{as } n \to \infty. \tag{26}$$

Let $E = \max_k |\epsilon(k)|$. We have

$$|(1+\epsilon(a_1))\cdots(1+\epsilon(a_d)) - 1| \leq 2^d \max(E,1)^{d-1} \left(|\epsilon(a_1)| + \cdots + |\epsilon(a_d|\right)$$

since every term in the expansion of the LHS is the product of some $\epsilon(a_i)$'s. So (26) reduces to showing that

$$\frac{2^d \max(E,1)^{d-1}}{d^{n+d}} \sum_{\substack{a_1,\ldots,a_d \\ \sum a_i = n}} \binom{n+d}{a_1+1,\ldots,a_d+1} \left(|\epsilon(a_1)| + \cdots + |\epsilon(a_d|\right) \to 0, \quad \text{as } n \to \infty,$$

which follows immediately from Lemma 5.6. This completes the proof of Theorem 5.2. $\square$

6 Conjectures and further questions

Using the data in Table 2, we make the following conjecture about the asymptotic expansion of $B_n/2^n$.

Conjecture 6.1. $\dfrac{B_n}{2^n} = \dfrac{1}{4n} + \dfrac{1}{6n^2} + O\left(\dfrac{1}{n^3}\right).$

If we can show that $\frac{B_n}{2^n} = \frac{1}{4n} + O\left(\frac{1}{n^2}\right)$, then the proof of Theorem 5.2 can be modified to prove the following conjecture.

Conjecture 6.2. *For a fixed positive integer d,*

$$\frac{C_{d,n}}{(2d)^n} = \left(\frac{d}{n}\right)^d \left(1 + O\left(\frac{1}{n}\right)\right).$$

Our method for proving $B_n \sim 2^{n-2}/n$ does not give the second term in the asymptotic expansion, since the approximation of binomial coefficients in (12) is too weak for this purpose. It would be interesting to determine the other terms in the expansion.

Question 6.3. What is the asymptotic expansion of $B_n/2^n$ and $C_{d,n}/(2d)^n$?

Theorem 1.2, which gives the formula $B_n \sim 2^{n-2}/n$, seems too simple to be just a coincidence. (In contrast, the number of classical ballot sequences of length n is asymptotically $2^n/\sqrt{2n\pi}$; see Lemma 3.5.) It would be nice to have a simpler proof of this result that gives some intuition for formula. Such intuition could also give us the insight to tackle the two conjectures.

Bidirectional ballot sequences look superficially similar to Dyck paths and Catalan numbers. However, the former lack the nice enumerative properties enjoyed by the latter. There does not seem to be any simple recursive structure in bidirectional ballot sequences, and we were unable to find any useful recurrence relations or generating functions for B_n. This is what makes the enumeration of bidirectional ballot sequences particularly difficult. However, the surprisingly nice asymptotics hint that perhaps there is more to the picture that we have just seen.

Question 6.4. What are nice enumerative properties of bidirectional ballot sequences and d-confined walks?

Finally, in the context of random walks, what happens when we change the probability model?

Question 6.5. In a random walk on a line where each step move to the right with probability p and to the left with probability $1-p$, what is the probability that we end up with a bidirectional ballot walk? A similar question can be asked for d-confined walks.

Acknowledgments

This research was carried out at the University of Minnesota Duluth under the supervision of Joseph Gallian with the financial support of the National Science Foundation and the Department of Defense (grant number DMS 0754106), the National Security Agency (grant number H98230-06-1-0013), and the MIT Department of Mathematics. The author would like to thank Joseph Gallian for his encouragement and support. The author would also like to thank Nathan Kaplan and Ricky Liu for reading the paper and making valuable suggestions.

References

[1] W. Feller, *An introduction to probability theory and its applications*, 3 ed., vol. 1, Wiley, 1968.

[2] P. Hegarty, *Some explicit constructions of sets with more sums than differences*, Acta Arith. **130** (2007), no. 1, 61–77.

[3] P. Hegarty and S. J. Miller, *When almost all sets are difference dominated*, Random Structures Algorithms (2009), no. 1, 118–136.

[4] G. Martin and K. O'Bryant, *Many sets have more sums than differences*, Additive combinatorics, CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, 2007, pp. 287–305.

[5] S. J. Miller, B. Orosz, and D. Scheinerman, *Explicit constructions of infinite families of MSTD sets*, 2008, Submitted to J. Number Theory, preprint at `http://arxiv.org/abs/0809.4621`.

[6] M. B. Nathanson, *Problems in additive number theory. I*, Additive combinatorics, CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, 2007, pp. 263–270.

[7] ———, *Sets with more sums than differences*, Integers **7** (2007), A5, 24 pp. (electronic).

[8] M. Renault, *Four proofs of the ballot theorem*, Math. Mag. **80** (2007), no. 5, 345–352.

[9] N. J. A. Sloane, *The On-Line Encyclopedia of Integer Sequences*, published electronically at `www.research.att.com/~njas/sequences/`, 2009.

[10] Y. Zhao, *Counting MSTD sets in finite abelian groups*, in preparation.

[11] ———, *Subsets characterized by the number of missing sums and differences*, in preparation.