# INTERSECTIONS AND COLLISIONS OF SIMPLE RANDOM WALKS IN $\mathbb{Z}^d$

NITYA GADHIWALA

Abstract. In this paper, we prove the known result of the expectation of the number of intersections and collisions of two independent simple symmetric random walks on $\mathbb{Z}^d$. We also present a demonstration of the calculated result using simulations and plot the empirical expectation obtained for both the intersections and collisions. One of the main tools used to compute the result is the local central limit theorem which is used to estimate the probability of finding the random walk at a point $x$ after $n$ steps.

## Contents



## 1. Introduction

In "Two incidents"[1], George Pòlya talks about the incident that led him to study the intersections of random walks, and his famous result on recurrence and transience.

> *I had then the habit of doing my mathematical work in an agreeable and healthy way in strolling through the woods. I carried paper and pencil and occasionally a few books. Sometimes I sat down at a table and scribbled a few formulas. Then I continued my leisurely walk in thinking about my problem until another table invited me to sit down and scribble a little more or look up something in a book.*

---

*Date*: October 10, 2020.

*At the hotel there lived also some students with whom I usually took my meals and had friendly relations. On a certain day one of them expected the visit of his fiancée, what I knew, but I did not foresee that he and his fiancée would also set out for a stroll in the woods, and then suddenly I met them there. And then I met them the same morning repeatedly, I don't remember how many times, but certainly much too often and I felt embarrassed: It looked as if I was snooping around which was, I assure you, not the case.*

*I met them by accident–but how likely was it that it happened by an accident at not on purpose? There is a question of probability–but if the question is conceived too narrowly, too realistically, au pied de la lettre, with the actual data about the network of winding footpaths behind the hotel, it becomes unmanageably complicated and, moreover, uninteresting.*

One reason to study the intersections of random walks is to understand better the random walks where no intersections occur. These walks are called self avoiding walks, and have significance in many subjects other than mathematics. In biology, self avoiding walks are a model used to study the folding of protein molecules. Self avoiding walks is also a tool used to study the structure and geometry of long chain polymer molecules. They have also found applications elsewhere in the sciences, such as the physics of magnetic materials and the study of phase transitions. Many of the problems involving intersections of random walks arise in studying statistical physics and other critical phenomena.

Problems dealing with the non-intersection of paths of random walks have been studied in detail by Professor Gregory F. Lawler, in his book "Intersections of Random Walks"[2].

### 1.1. **Definitions.**

The random walks considered in this paper are all on the $d$ dimensional integer lattice $\mathbb{Z}^d$.

**Definition 1.1.** Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables defined on $\mathbb{Z}^d$. They take values in

$$\text{Range}(X_i) := \{e_1, e_2, \ldots, e_d, -e_1, -e_2, \ldots, -e_d\}$$

where $e_i$s are the standard basis for $\mathbb{R}^d$. The probability distribution of $X_i$ for $i \in \{1, 2, \ldots, n\}$ is given by

$$\mathbb{P}\{X_i = x\} = \frac{1}{2d} \qquad \text{for } x \in \{e_1, e_2, \ldots, e_d, -e_1, -e_2, \ldots, -e_d\}.$$

The simple symmetric random walk, starting at a point $x$, is a stochastic process $S_n$ indexed by non-negative integers, with $S_0 = x$ and

$$S_n = S_0 + X_1 + X_2 + \cdots + X_n.$$

The probability distribution of $S_n$ is denoted by $\mathbb{P}^x\{S_n = y\}$ for every $y \in \mathbb{Z}^d$.

All the random walks in this paper have been assumed to start at 0 (i.e., $S_0 = 0$). Hence, $S_n = X_1 + X_2 + \cdots + X_n$.

Calculating $\mathbb{P}\{S_n = y\} := \mathbb{P}^0\{S_n = y\}$ for some $y \in \mathbb{Z}^d$ is non-trivial. One combinatorial approach to calculate the required probability is by counting the number of possible walks in $n$ steps and then multiplying it with the probability of getting a particular walk[1]. This method works exceptionally well for random walks on $\mathbb{Z}$, where $\mathbb{P}\{S_n = y\}$ can be calculated precisely to get

$$\mathbb{P}\{S_n = y\} = \begin{cases} \binom{n}{\frac{n+y}{2}} \left(\frac{1}{2}\right)^n, & n+y \text{ even and } |y| \leq n, \\ 0, & \text{Otherwise.} \end{cases}$$

This method of counting the number of $n$ step walks becomes increasingly harder to calculate for random walks on $\mathbb{Z}^d$, for $d$ greater than 1.

Another method to estimate $\mathbb{P}\{S_n = y\}$, is using the characteristic function, which will be described in detail in Section 3.1.

**Definition 1.2.** A random walk is called time reversible if the probability of $\mathbb{P}\{S_0 = x_0, S_1 = x_1, \ldots, S_k = x_k\}$ and $\mathbb{P}\{S_0 = x_k, S_1 = x_{k-1}, \ldots, S_k = x_0\}$ are equal.

It can be noticed that the simple symmetric random walk on $\mathbb{Z}^d$ is time reversible as

$$\mathbb{P}\{S_0 = x_0, S_1 = x_1, \ldots, S_k = x_k\} = \mathbb{P}\{X_1 = x_1 - x_0, \ldots, X_k = x_k - x_{k-1}\} = \frac{1}{(2d)^k}$$

and

$$\mathbb{P}\{S_0 = x_k, S_1 = x_{k-1}, \ldots, S_k = x_0\} = \mathbb{P}\{X_1 = x_{k-1} - x_k, \ldots, X_k = x_1 - x_0\} = \frac{1}{(2d)^k}.$$

Since the probability of a walk traversing the path $\{x_0, x_1, x_2, \ldots, x_k\}$ is the same as that of traversing it's reverse, $\{x_k, x_{k-1}, \ldots, x_1, x_0\}$, the walks are time reversible.

**Definition 1.3.** Let $S^1$ and $S^2$ be two independent simple random walks on a given lattice. The number of intersections of the two random walks in the first $n$ steps is denoted by a random variable $R_n$ given by

$$R_n = \sum_{i=1}^{n} \sum_{j=1}^{n} 1\{S_i^1 = S_j^2\}.$$

Given any two random walks, $S^1$ and $S^2$, an intersection is said to have occured when for any two times (they may be same or different), both the random walks have been at the same point in the state space.

**Definition 1.4.** Let $S^1$ and $S^2$ be two independent simple random walks on a given lattice. The number of collisions of the two random walks in the first $n$ steps is denoted by a random variable $C_n$ given by

$$C_n = \sum_{i=1}^{n} 1\{S_i^1 = S_i^2\}.$$

[1]In the case of a simple symmetric random walk on $\mathbb{Z}^d$, this will be $\frac{1}{(2d)^n}$ for each walk of $n$ steps.

For a collision to occur, both the random walks require to be at the same point in the state space at the same time. To count the number of intersections, the path travelled by the random walks need to be drawn out and for each pair of indices $(i, j)$ where $1 \leq i, j \leq n$, it must be checked if an intersection has occured. For the collisions of random walks, the number of times the two random walkers meet is counted. [2]

It can be observed that every collision is an intersection but the converse isn't true. Therefore, for any given pair of random walks $R_n \geq C_n$, for all $n$.

### 1.2. **Results.**

In this paper, we calculate the expected number of collisions and intersections of simple symmetric random walks in $\mathbb{Z}^d$.

**Theorem 1.5.** *Let $S^1$ and $S^2$ be two simple random walks on $\mathbb{Z}^d$ starting at the origin. Let $R_n$ be the number of intersections between the first $n$ steps of $S^1$ and the first $n$ steps of $S^2$. As $n \to \infty$,*

$$\mathbb{E}[R_n] = \begin{cases} c_1 n^{\frac{3}{2}} + O(n^{\frac{1}{2}}), & d = 1, \\ c_2 n + O(\ln(n)), & d = 2, \\ c_3 n^{\frac{1}{2}} + O(1), & d = 3, \\ c_4 \ln(n) + O(1), & d = 4, \\ c_d + O(n^{\frac{4-d}{2}}), & d \geq 5. \end{cases} \tag{1.6}$$

**Theorem 1.7.** *Let $S^1$ and $S^2$ be two simple random walks on $\mathbb{Z}^d$ starting at the origin. Let $C_n$ be the number of collisions occuring in the first $n$ steps of $S^1$ and $S^2$. As $n \to \infty$*

$$\mathbb{E}[C_n] = \begin{cases} c_1 \sqrt{n} + O(1), & d = 1, \\ c_2 \log(n) + O(1), & d = 2, \\ c_d + O(\frac{1}{n^{\frac{d}{2}}}), & d \geq 3. \end{cases} \tag{1.8}$$

The above theorems have been proved using the local central limit theorem.

The rest of the paper is organised as follows: In Section 2, we will take a numerical approach where we simulate 100 pairs of random walks and count their intersections and collisions. We also verify the result using these simulations. In Section 3, we discuss some preliminaries, which will be used to prove the above result on the expected number of random walks. These include the characteristic function (see Section 3.1) and the local central limit theorem (see Section 3.2). In Section 4, we will calculate the expected number of intersections and collisions of two simple symmetric random walks on $\mathbb{Z}^d$.

## 2. Simulations

To understand the order of the expected number of intersections, we can simulate random walks and count their intersections and collisions. This approach provides

[2]George Pòlya meeting his student would be a collision not an intersection.

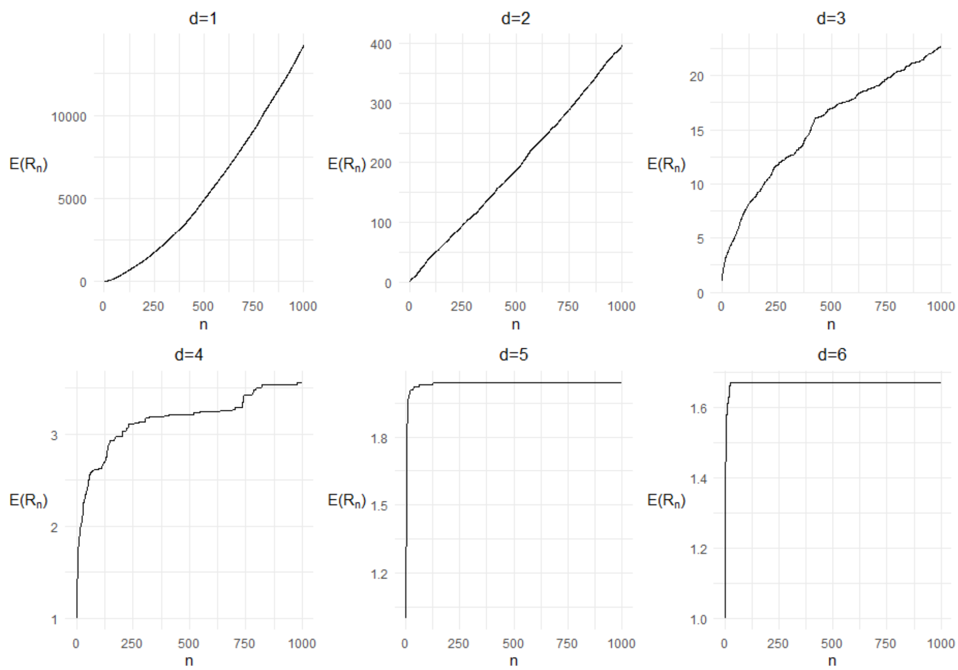



FIGURE 1. For the simulation as explained in Section 2.1, the following result is seen for the expected number of intersections for $n$ steps of the random walks on $\mathbb{Z}^d$ for $d = 1, 2, \ldots, 6$.

an intuition on what the result should look like. In this section, we simulate 100 pairs of simple symmetric random walks in $\mathbb{Z}^d$, for $d = 1, 2, \ldots, 6$ and find the empirical average of the number of intersections and collisions.

## 2.1. Intersections of Random Walks.

For random walks on $\mathbb{Z}^d$, the following procedure is performed.

- Choose some dimension $d$.
- Simulate a pair of simple random walks till 1000 steps.
- Count the number of intersections after $n$ steps for $n = 1, 2, \ldots, 1000$.
- Repeat this experiment 100 times.
- Take the means of the 100 counts of intersections for each $n$, this should give an approximate value of $E[R_n]$ for $n = 1, 2, \ldots, 1000$.

The values of $E[R_n]$ calculated in the above method have been plotted in Figure 1.

Several observations can be made from the graphs. First, it is visible that the behaviour across different dimensions is very different.

For dimension 1, we observe that the number of intersections increases very quickly. Another observation from the simulations is that out of the 100 pairs of random walks observed, the minimum number of intersections was 301 while the maximum number of intersections was 40371.[3]

It can be seen that for dimensions 1 to 4, as $n$ increases, the number of intersections keeps on increasing with $n$, whereas for dimensions 5 and 6, in less than 250 steps, the mean number of intersections becomes a constant. This means that for the 100 pairs of random walks observed here for dimensions 5 and 6 each, after around 250 steps, no new intersections were observed. This indicates that the critical dimension here is 4, since for larger dimensions, the number of intersections becomes constant very quickly.

### 2.2. **Collisions of Random Walks.**

Now, we describe a similar procedure for the collisions of the random walks. For $d = 1, 2, 3, 4$, we count the average number of collisions for 100 pairs random walks on $\mathbb{Z}^d$ taken up to 1000 steps.

The steps followed are the same as those described for the intersections in Section 2.1. Figure 2 contains the empirical means of the collisions (i.e., $\mathbb{E}[C_n]$) that occured during the first $n$ steps for $1 \le n \le 1000$.

In Figure 2, for dimensions 1 and 2, the expected number of collisions (i.e., $\mathbb{E}[C_n]$) continues to increase with $n$. While for dimensions 3 and 4, after the first 100 steps, no further collisions are observed. This indicates that the critical dimension is 2.

On comparing Figure 1 Figure 2, it can be observed that the expected number of collisions and intersections differ very much for all the values of $d$ observed. Observing only the random walk on $\mathbb{Z}$, it is seen that the empirical mean number of intersections exceeds 10000 while the mean number of collisions reaches only 30.

Furthermore, it can be observed that the critical dimensions for the expected number of collisions and intersections are quite different. For the intersections of random walks, the critical dimension for the expectation is observed to be 4, whereas for the collisions, it is observed to be 2.

### 2.3. **Verification.**

It can be observed that the expected number of simulations as seen in Figures 1 and 2 resembles the results as seen in the proofs of Theorem 4.1 and Theorem 4.17. The critical dimension for the intersections and collisions can be seen clearly in the graphs as 4 and 2 respectively, which matches the result from Theorem 4.1 and Theorem 4.17. It is clearly seen that for simple symmetric random walks on $\mathbb{Z}^d$, for $d \geq 5$, the number of intersections eventually resembles a constant. And similarly for collisions of random walks and $d \geq 3$.

To further confirm the order of the first term in the simulations seen in Figure 1 for random walks on the integer lattice for dimensions 1 to 4, the following has been done. Figure 3, contains the simulations for the random walks simulated in Section 2. For $d = 1$, the plot is that of $\frac{\mathbb{E}[R_n]}{n^{\frac{3}{2}}}$ against $n$, for $d = 2$ we plot $\frac{\mathbb{E}[R_n]}{n}$

[3]Since while counting the number of intersections of random walks, we consider every pair of indices, 40371 is a valid number of intersections for a pair of random walk to have traversed 1000 steps.

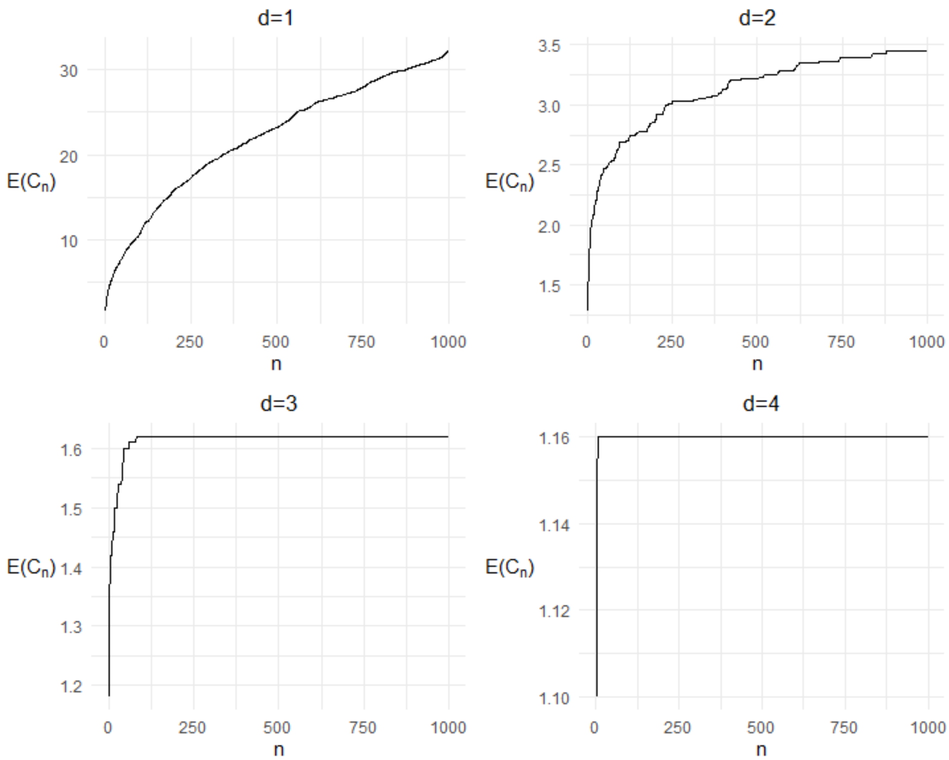



FIGURE 2. For the simulation as explained in Section 2.2, the following result is seen for the expected number of collisions for $n$ steps of the random walks on $\mathbb{Z}^d$ for $d = 1, 2, 3$ and 4.

against $n$, for $d = 3$ we plot $\frac{\mathbb{E}[R_n]}{n^{\frac{1}{2}}}$ against $n$ and for $d = 4$ the plot is of $\frac{\mathbb{E}[R_n]}{\log(n)}$ against $n$.

As $n \to \infty$, it can be seen that the error term divided by the order of the first term goes to 0. Hence, Figure 3 should resemble a non-zero constant function as $n \to \infty$.

A similar analysis for the collisions of random walks, requires us to plot for $d = 1$, $\frac{\mathbb{E}[C_n]}{\sqrt{n}}$ against $n$ and for $d = 2$, we plot $\frac{\mathbb{E}[C_n]}{\log n}$. These plots should resemble a non-zero constant functions as $n \to \infty$.

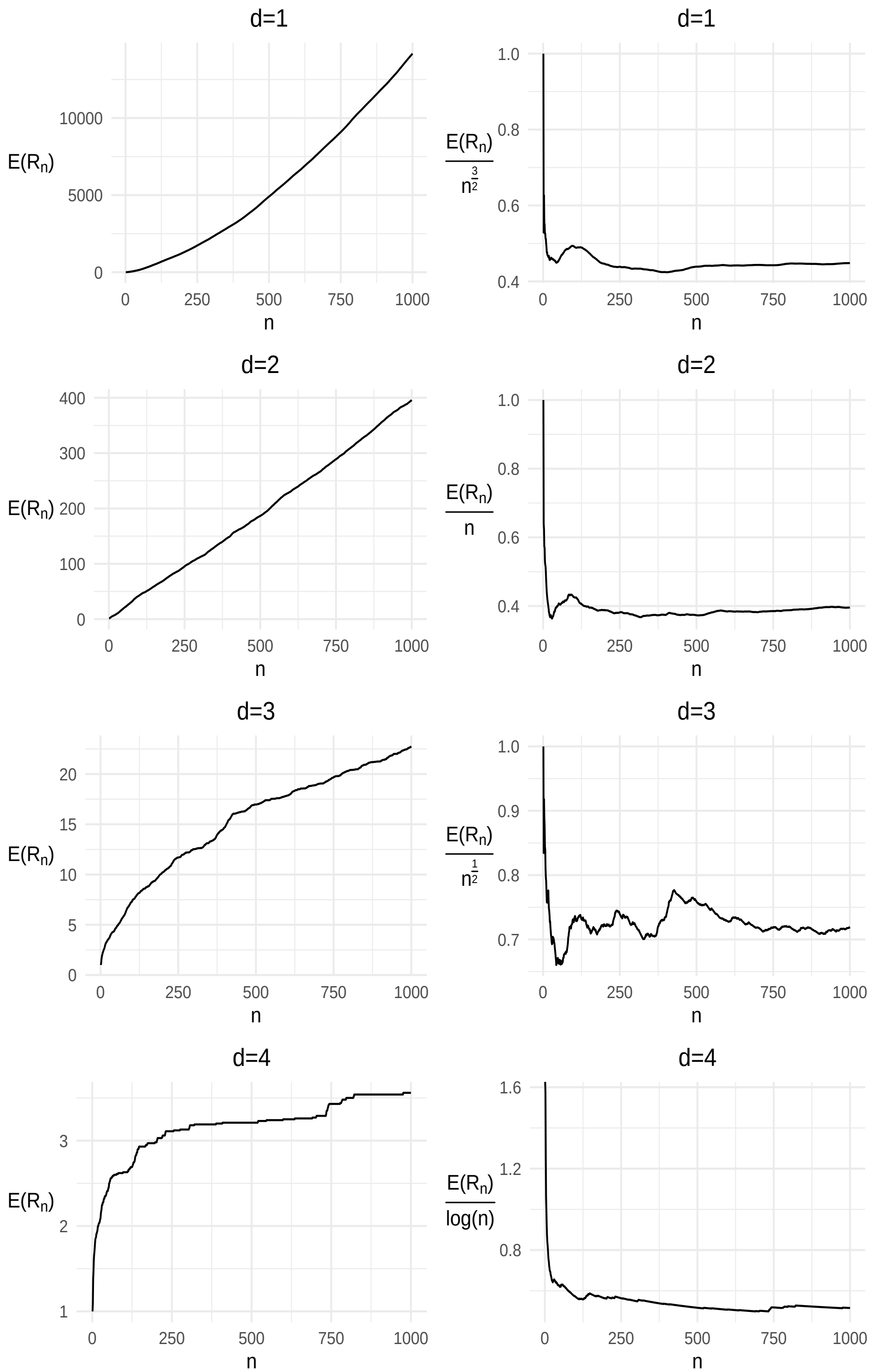



FIGURE 3. For the simulation seen in Figure 1, the following result is the expected number of intersections divided by the order of the leading term of the result as seen in Theorem 4.1 for $d = 1, 2, 3$ and 4.

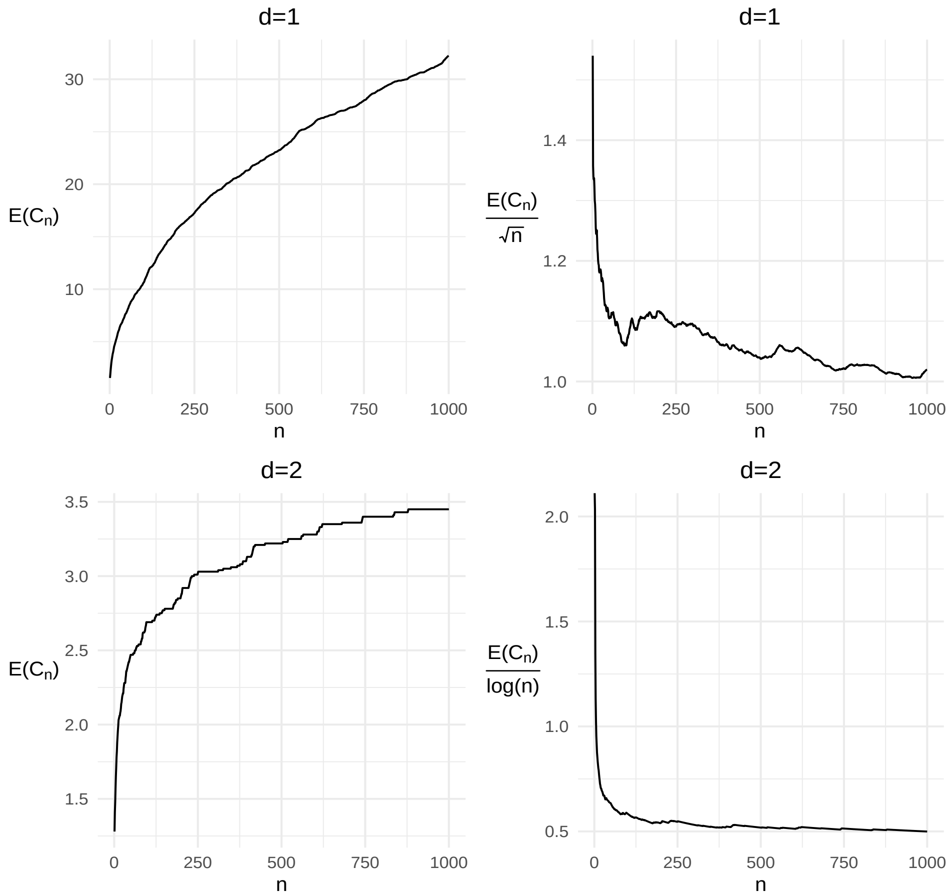



FIGURE 4. For the simulation seen in Figure 1, the following result is the expected number of intersections divided by the order of the leading term of the result as seen in Theorem 4.1 for $d = 1, 2, 3$ and 4.

It can be seen that in Figure 4, that both the plots resemble a non-zero constant function.

## 3. Preliminaries

For a simple symmetric random walk on $\mathbb{Z}^d$, since the combinatorial method is technically tedious to calculate $\mathbb{P}\{S_n = x\}$ for dimensions greater than 1, we outline an alternative method. Computing the characteristic function of each of the $X_i$'s is easier than the probability that the walk is at a particular location after $n$ steps. Hence, a method to calculate $\mathbb{P}\{S_n = x\}$ using the characteristic function is described in this section. Using the characteristic function, we prove the local central limit theorem. While the local central limit theorem has been proven here for simple symmetric random walks on $\mathbb{Z}^d$, it can be proven for any random walk whose increments have mean zero and finite variance.

### 3.1. **Characteristic Function.**

**Definition 3.1.** The characteristic function of a random variable $Y = (Y_1, Y_2, \dots, Y_d)$ taking values in $\mathbb{R}^d$ is a function $\phi : \mathbb{R}^d \to \mathbb{C}$ given by

$$\phi_Y(\theta) = \mathbb{E}[e^{i\theta \cdot Y}]$$

for $\theta \in \mathbb{R}^d$.

Let us calculate the characteristic function for $X_1$ following the distribution described in Definition 1.1. Let $\theta = (\theta_1, \dots, \theta_d) \in \mathbb{R}^d$.

$$\begin{aligned}\phi_{X_1}(\theta) = \mathbb{E}[e^{i \cdot X_1 \theta}] &= \sum_{y \in \text{Range}(X_1)} \mathbb{P}\{X_1 = y\} e^{iy\cdot\theta} \\ &= \frac{1}{2d} \sum_{y \in \text{Range}(X_1)} e^{iy\cdot\theta} \\ &= \frac{1}{2d} \sum_{j=1}^{d} \left(e^{i\theta_j} + e^{-i\theta_j}\right) \\ &= \frac{1}{2d} \sum_{j=1}^{d} 2\cos(\theta_j) = \frac{1}{d} \sum_{j=1}^{d} \cos(\theta_j). \qquad (3.2)\end{aligned}$$

Equation (3.2) describes the characteristic function for $X_1$. Since each of the $X_i$'s follow the same distribution as $X_1$, their characteristic function can be calculated in the same way to give the same result. Let this function be denoted as $\phi_X$. For some $\theta = (\theta_1, \dots, \theta_d) \in \mathbb{R}^d$,

$$\phi_X(\theta) := \phi_{X_1}(\theta) = \phi_{X_2}(\theta) = \cdots = \phi_{X_n}(\theta) = \frac{1}{d} \sum_{j=1}^{d} \cos(\theta_j). \qquad (3.3)$$

Since, $S_n$ is the sum of the $X_i$'s, its charactersistic function can be found in the following way

$$\begin{aligned}\phi_{S_n}(\theta) &= \mathbb{E}[e^{i(X_1 + \cdots + X_n)\theta}] \\ &= \mathbb{E}[e^{iX_1\theta} e^{iX_2\theta} \cdots e^{iX_n\theta}] \\ &= \mathbb{E}[e^{iX_1\theta}]\mathbb{E}[e^{iX_2\theta}] \cdots \mathbb{E}[e^{iX_2\theta}] \\ &= [\phi_X(\theta)]^n = \phi_X^n(\theta) \\ &= \frac{1}{d^n} \left( \sum_{j=1}^{d} \cos(\theta_j) \right)^n. \qquad (3.4)\end{aligned}$$

Above, the third equality follows from the fact that the $X_i$s are independent and the fifth equality follows from Equation (3.2).

We observe that for random walks on $\mathbb{Z}^d$, for dimensions greater than 1, it is easier to compute the characteristic function of $S_n$ when compared to the probability distribution of $S_n$. Furthermore, inverting the characteristic function as described below can be used to calculate $\mathbb{P}\{S_n = x\}$.

**Theorem 3.5.** *If $Y = (Y_1, \dots, Y_d)$ is a $\mathbb{Z}^d$-valued random variable whose characteristic function is given by $\phi_Y(\theta)$, then the following equation holds*

$$\mathbb{P}\{Y = x\} = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \phi_Y(\theta) e^{-i\theta\cdot x} d\theta. \tag{3.6}$$

*Proof.* From the definition of the characteristic function we have

$$\phi_Y(\theta) = \sum_{y \in \mathbb{Z}^d} \mathbb{P}\{Y = y\} e^{i\cdot y\theta}.$$

For $x \in \mathbb{Z}^d$,

$$\phi_Y(\theta) e^{-i\theta\cdot x} = \sum_{y \in \mathbb{Z}^d} \mathbb{P}\{Y = y\} e^{i\theta\cdot y} e^{-i\theta\cdot x}.$$

On integrating both sides the equation over $[-\pi, \pi]^d$, we get

$$\int_{[-\pi,\pi]^d} \phi_Y(\theta) e^{-i\theta\cdot x} d\theta = \int_{[-\pi,\pi]^d} \left( \sum_{y \in \mathbb{Z}^d} \mathbb{P}\{Y = y\} e^{i\theta\cdot(y-x)} \right) d\theta. \tag{3.7}$$

Verifying the condition for Fubini's theorem (See [4]) on the right side of Equation (3.7) above we get

$$\begin{aligned}\int_{[-\pi,\pi]^d} \sum_{y \in \mathbb{Z}^d} \left| \mathbb{P}\{Y = y\} e^{i\theta\cdot(y-x)} \right| d\theta &= \int_{[-\pi,\pi]^d} \left( \sum_{y \in \mathbb{Z}^d} \mathbb{P}\{Y = y\} \right) d\theta \\ &= \int_{[-\pi,\pi]^d} d\theta = (2\pi)^d < \infty.\end{aligned}$$

Hence we can apply Fubini's Theorem to Equation (3.7) to get

$$\int_{[-\pi,\pi]^d} \phi_Y(\theta) e^{-i\theta\cdot x} d\theta = \sum_{y \in \mathbb{Z}^d} \mathbb{P}\{Y = y\} \int_{[-\pi,\pi]^d} e^{i\theta\cdot(y-x)} d\theta. \tag{3.8}$$

If $y \neq x + 2\pi k$ for any $k \in \mathbb{Z}^d$, then

$$\int_{[-\pi,\pi]^d} e^{i\theta\cdot(y-x)} d\theta = 0.$$

If $y - x = 2\pi k$ for some $k \in \mathbb{Z}^d$, the integral becomes

$$\int_{[-\pi,\pi]^d} e^{i(2\pi k)\theta} d\theta = \int_{[-\pi,\pi]^d} 1 d\theta = (2\pi)^d.$$

Substituting the above values in (3.8) and rearranging, we have:

$$\mathbb{P}\{Y = y\} = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \phi_Y(\theta) e^{-i\theta\cdot y} d\theta. \tag{3.9}$$

□

**Corollary 3.10.** *In the above equation, replacing $Y$ with $S_n$, we get*

$$\mathbb{P}\{S_n = x\} = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} [\phi_X(\theta)]^n e^{i\theta x} d\theta. \tag{3.11}$$

Further, on replacing the value of $\phi_X(\theta)$ from Equation (3.4), we get the following equation

$$\mathbb{P}\{S_n = x\} = \frac{1}{(2\pi)^d} \frac{1}{d^n} \int_{[-\pi,\pi]^d} \left( \sum_{j=1}^{d} \cos(\theta_j) \right)^n e^{i\theta x} d\theta. \tag{3.12}$$

3.2. **The Local Central Limit Theorem.**

We aim to estimate $\mathbb{P}\{S_n = x\}$ for large $n$, where $S_n$ is a random variable denoting the position of a simple symmetric random walk in $\mathbb{Z}^d$ after $n$ steps.

If we consider $d = 1$, we have $X_i$s to be real valued variables and $S_n$ to be their sum. Since $X_1, X_2, \ldots, X_n$ are independent and identically distributed, the central limit theorem can be applied here.

The central limit theorem states that,

$$\lim_{n\to\infty} \mathbb{P}\{a \leq \frac{S_n}{\sqrt{n}} \leq b\} = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} dy. \tag{3.13}$$

As $S_n$ is a discrete random variable and the normal distribution is a continuous distribution, to approximate the value of $\mathbb{P}\{S_n = x\}$, one can consider $a = \frac{x}{\sqrt{n}}$ and $b = \frac{x+1}{\sqrt{n}}$. The right side of equation (3.13) becomes

$$\begin{aligned} \int_{\frac{x}{\sqrt{n}}}^{\frac{x+1}{\sqrt{n}}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} dy &= \int_x^{x+1} \frac{1}{\sqrt{2\pi\sigma^2 n}} e^{-\frac{u^2}{2\sigma^2 n}} du \\ &\approx \frac{1}{\sqrt{2\pi\sigma^2 n}} e^{-\frac{x^2}{2\sigma^2 n}}. \end{aligned} \tag{3.14}$$

Equation (3.14) uses the Riemann sum to approximate the integral.

The points of $\mathbb{Z}^d$ can be partitioned into odd and even points, where odd points are the ones that can be reached in an odd number of steps and even points are those that can be reached in an even number of steps.

**Definition 3.15.** Consider a simple random walk on $\mathbb{Z}^d$. Let $x = (x_1, x_2, \ldots, x_d) \in \mathbb{Z}^d$ and $n \in \mathbb{N} \cup \{0\}$. We say that $n$ and $x$ have the same parity if $n + x_1 + x_2 + \cdots + x_d$ is even.

For any $n$, every alternate point will have the same parity as $n$. Hence, we can approximate Equation (3.14) to the following

$$\mathbb{P}\{S_n = x\} = p_n(x) \approx \frac{2}{\sqrt{2\pi\sigma^2 n}} e^{-\frac{x^2}{2\sigma^2 n}}. \tag{3.16}$$

The local central limit theorem makes Equation (3.16) precise by justifying the approximation.

**Theorem 3.17. Local Central Limit Theorem**

*Let $S_n$ denote a simple symmetric random walk in $\mathbb{Z}^d$. For a positive integer $n$ and a point $x \in \mathbb{Z}^d$, define*

$$p_n(x) := \mathbb{P}\{S_n = x\}$$

*and*

$$\overline{p}_n(x) := 2\left(\frac{d}{2\pi n}\right)^{d/2} e^{-\frac{d|x|^2}{2n}}.$$

*Then, for $n$ and $x$ having the same parity,*

$$p_n(x) = \overline{p}_n(x) + O\left(\frac{1}{n^{\frac{d}{2}+1}}\right). \tag{3.18}$$

3.3. **Proof of the Local Central Limit Theorem.**

Local central limit theorem is proven by splitting the domain of the integral in Equation (3.11). The different integrals are approximated differently.

From Equation (3.11), we have

$$\mathbb{P}\{S_n = x\} = \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} [\phi_X(\theta)]^n\, e^{i\theta\cdot x} d\theta.$$

It can be noticed that in the integral above, on replacing the term $\theta$ with $\theta + (\pi, \pi, \ldots, \pi)$, the absolute value of the integrand doesn't change. Hence, the integral over $A = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \times [-\pi, \pi]^{(d-1)}$, will be half of the original integral.

Hence, the above equation can be rewritten to give

$$\mathbb{P}\{S_n = x\} = \frac{2}{(2\pi)^d}\int_A [\phi_X(\theta)]^n\, e^{i\theta\cdot x} d\theta. \tag{3.19}$$

Earlier, in equation (3.2) we saw that using the Taylor Expansion

$$\begin{aligned}
\phi_X(\theta) &= \frac{1}{d}\sum_{j=1}^{d}\cos(\theta_j)\\
&= \frac{1}{d}\sum_{j=1}^{d}\left(1 - \frac{\theta_j^2}{2!} + \frac{\theta_j^4}{4!} - \cdots\right)\\
&= 1 - \sum_{j=1}^{d}\frac{\theta_j^2}{2!d} + \sum_{j=1}^{d}\frac{\theta_j^4}{4!d} - \cdots\\
&= 1 - \frac{|\theta|^2}{2d} + O(|\theta|^4).
\end{aligned} \tag{3.20}$$

Equation (3.20) follows from the fact that

$$\phi_X(\theta) - \left(1 - \frac{|\theta|^2}{2d}\right) \le \sum_{j=1}^{d}\frac{\theta_j^4}{4!d} \le \left(\sum_{j=1}^{d}\frac{1}{4!d}\right)|\theta|^4.$$

Equation (3.20) can be rewritten as

$$\phi_X(\theta) = 1 - \frac{1}{4d}|\theta|^2 - \frac{1}{4d}|\theta|^2 + O(|\theta|^4).$$

For $\theta$ close to 0,

$$\frac{1}{4d}|\theta|^4 = O(|\theta|^4).$$

So, there exists $r > 0$ such that for $|\theta| < r$, $\phi_X(\theta) \le 1 - \frac{1}{4d}|\theta|^2$.

There exists $0 < \rho = \rho(r) < 1$ such that $\phi_X(\theta) \leq \rho$ for $\theta > r$.

Furthermore, let us consider Equation (3.19) and split the integral as a sum of the following two integrals as $A = (A \cap \{|\theta| \leq r\}) \sqcup (A \cap \{|\theta| > r\})$.

Let $\mathbb{P}\{S_n = x\} = I(n,x) + J(n,x)$ where,

$$I(n,x) = \frac{2}{(2\pi)^d} \int_{A\cap\{|\theta|\leq r\}} [\phi_X(\theta)]^n e^{i\theta\cdot x} d\theta \tag{3.21}$$

$$J(n,x) = \frac{2}{(2\pi)^d} \int_{A\cap\{|\theta|> r\}} [\phi_X(\theta)]^n e^{i\theta\cdot x} d\theta. \tag{3.22}$$

First, we estimate $J(n,x)$. Earlier we saw that for all $|\theta| > r$, $\phi_X(\theta) \leq \rho < 1$. Substituting this in Equation (3.22) and simplifying we get the following:

$$\begin{aligned}
|J(n,x)| &= \frac{2}{(2\pi)^d} \left| \int_{A\cap\{|\theta|> r\}} [\phi_X(\theta)]^n e^{i\theta\cdot x} d\theta \right| \\
&\leq \frac{2}{(2\pi)^d} \int_{A\cap\{|\theta|> r\}} |\, [\phi_X(\theta)]^n e^{i\theta\cdot x}| d\theta \\
&= \frac{2}{(2\pi)^d} \int_{A\cap\{|\theta|> r\}} |\phi_X(\theta)|^n d\theta \\
&\leq \frac{2}{(2\pi)^d} \int_{A\cap\{|\theta|> r\}} |\rho|^n d\theta \\
&\leq \frac{2\rho^n}{(2\pi)^d} \int_{A\cap\{|\theta|> r\}} d\theta \\
&\leq \frac{2\rho^n}{(2\pi)^d} C \leq \frac{2\rho^n}{(2\pi)^d} (2\pi)^{d-1}\pi = O(\rho^n). \qquad (3.23)
\end{aligned}$$

In the above calculation, in Equation (3.23), $C$ represents the area over which the integral is being done (i.e., $A \cap \{|\theta| > r\}$). This area can be bounded by the area of $A$.

Equation (3.23) tells us that as $J(n,x) = O(\rho^n)$.

It remains to compute the integral $I(n,x)$. It is simplified by substituting $\theta$ with $\frac{\alpha}{\sqrt{n}}$ where $\alpha \in \mathbb{Z}^d$. Equation (3.21) can be rewritten as

$$\begin{aligned}
I(n,x) &= \frac{2}{(2\pi)^d} \int_{|\theta|\leq r} [\phi_X(\theta)]^n e^{i\theta\cdot x} d\theta \\
&= \frac{2}{(2\pi)^d} \int_{|\alpha|\leq r\sqrt{n}} e^{-ix\cdot\frac{\alpha}{\sqrt{n}}} \phi_X^n \left(\frac{\alpha}{\sqrt{n}}\right) \frac{d\alpha}{\sqrt{n^d}} \\
&= \frac{2}{(2\pi\sqrt{n})^d} \int_{|\alpha|\leq r\sqrt{n}} e^{-ix\cdot\frac{\alpha}{\sqrt{n}}} \phi_X^n \left(\frac{\alpha}{\sqrt{n}}\right) d\alpha
\end{aligned}$$

$$\frac{(2\pi\sqrt{n})^d}{2} I(n,x) = \int_{|\alpha|\leq r\sqrt{n}} e^{-ix\cdot\frac{\alpha}{\sqrt{n}}} \phi_X^n \left(\frac{\alpha}{\sqrt{n}}\right) d\alpha. \tag{3.24}$$

Let $I_1$, $I_2$, $I_3$ and $I_4$ be defined as follows

$$I_1 = \int_{|\alpha|\le n^{\frac14}} \exp\left\{-\frac{ix\cdot\alpha}{\sqrt{n}}\right\}\left[\phi^n\left(\frac{\alpha}{\sqrt{n}}\right) - \exp\left\{-\frac{\alpha^2}{2d}\right\}\right] d\alpha \tag{3.25}$$

$$I_2 = \int_{\mathbb{R}^d} \exp\left\{-\frac{ix\cdot\alpha}{\sqrt{n}}\right\}\exp\left\{-\frac{\alpha^2}{2d}\right\} d\alpha \tag{3.26}$$

$$I_3 = -\int_{|\alpha|\ge n^{\frac14}} \exp\left\{-\frac{ix\cdot\alpha}{\sqrt{n}}\right\}\exp\left\{-\frac{\alpha^2}{2d}\right\} d\alpha \tag{3.27}$$

$$I_4 = \int_{n^{\frac14}\le|\alpha|\le rn^{\frac12}} e^{-ix\cdot\frac{\alpha}{\sqrt{n}}}\phi_X^n\left(\frac{\alpha}{\sqrt{n}}\right) d\alpha. \tag{3.28}$$

Let $n > \left(\frac{1}{r}\right)^2$. Consider the integral $I(n,x)$ from Equation (3.24)

$$\begin{aligned}
\frac{(2\pi\sqrt{n})^d}{2} I(n,x) &= \int_{|\alpha|\le n^{\frac14}} e^{-ix\cdot\frac{\alpha}{\sqrt{n}}}\phi_X^n\left(\frac{\alpha}{\sqrt{n}}\right) d\alpha + \int_{n^{\frac14}\le|\alpha|\le rn^{\frac12}} e^{-ix\cdot\frac{\alpha}{\sqrt{n}}}\phi_X^n\left(\frac{\alpha}{\sqrt{n}}\right) d\alpha \\
&= \int_{|\alpha|\le n^{\frac14}} e^{-ix\cdot\frac{\alpha}{\sqrt{n}}}\phi_X^n\left(\frac{\alpha}{\sqrt{n}}\right) d\alpha + I_4 \\
&= \int_{|\alpha|\le n^{\frac14}} \exp\left\{-\frac{ix\cdot\alpha}{\sqrt{n}}\right\}\left[\phi^n\left(\frac{\alpha}{\sqrt{n}}\right) - \exp\left\{-\frac{\alpha^2}{2d}\right\}\right] d\alpha \\
&\qquad + \int_{|\alpha|\le n^{\frac14}} \exp\left\{-\frac{ix\cdot\alpha}{\sqrt{n}}\right\}\exp\left\{-\frac{\alpha^2}{2d}\right\} d\alpha + I_4 \\
&= I_1 + \int_{|\alpha|\le n^{\frac14}} \exp\left\{-\frac{ix\cdot\alpha}{\sqrt{n}}\right\}\exp\left\{-\frac{\alpha^2}{2d}\right\} d\alpha + I_4 \\
&= I_1 + \int_{\mathbb{R}^d} \exp\left\{-\frac{ix\cdot\alpha}{\sqrt{n}}\right\}\exp\{-\frac{\alpha^2}{2d}\}d\alpha - \int_{|\alpha|\ge n^{\frac14}} \exp\left\{-\frac{ix\cdot\alpha}{\sqrt{n}}\right\}\exp\left\{-\frac{\alpha^2}{2d}\right\} d\alpha + I_4 \\
&= I_1 + I_2 + I_3 + I_4.
\end{aligned} \tag{3.29}$$

The four integrals $I_1$, $I_2$, $I_3$ and $I_4$ are approximated separately.

**Approximating $I_1$:**

We begin by approximating the integral $I_1 = I_1(n,x)$. From Equation (3.25), we have

$$\begin{aligned}
I_1 &= \int_{|\alpha|\le n^{\frac14}} \exp\{-\frac{ix\cdot\alpha}{\sqrt{n}}\}\left[\phi^n(\frac{\alpha}{\sqrt{n}}) - \exp\{-\frac{\alpha^2}{2d}\}\right] d\alpha \\
|I_1| &\le \int_{|\alpha|\le n^{\frac14}} \left|\phi^n(\frac{\alpha}{\sqrt{n}}) - \exp\{-\frac{\alpha^2}{2d}\}\right| d\alpha.
\end{aligned} \tag{3.30}$$

In the above equation, to estimate $\phi\left(\frac{\alpha}{\sqrt{n}}\right)$, we replace the value from Equation (3.20).

$$\begin{aligned}
\phi\left(\frac{\alpha}{\sqrt{n}}\right) &= 1 - \frac{|\alpha|^2}{2dn} + |\alpha|^4 O(n^{-2}) \\
\phi^n\left(\frac{\alpha}{\sqrt{n}}\right) &= \left[\phi\left(\frac{\alpha}{\sqrt{n}}\right)\right]^n = \left(\left(1 - \frac{|\alpha|^2}{2dn}\right) + |\alpha|^4 O(n^{-2})\right)^n.
\end{aligned} \tag{3.31}$$

Using the binomial expansion to simplify the term on the right, we get: $\sum_{r=0}^{n}\binom{n}{r}\left(1-\frac{|\alpha|^2}{2dn}\right)^{n-r}|\alpha|^{4r}O(n^{-2r})$. Which on simplifying becomes $\sum_{r=0}^{n}\left(1-\frac{|\alpha|^2}{2dn}\right)^{n-r}|\alpha|^{4r}O(n^{-r})$. If we consider $r \gg 1$, the $O(n^{-r})$ goes to 0 as $n \to \infty$. Since $\left(1-\frac{|\alpha|^2}{2dn}\right)^{n-r}|\alpha|^{4r}$ can be bounded by a constant for $|\alpha| \leq n^{\frac{1}{4}}$, the the term for $r \gg 1$ is negligible. Hence, we only need to consider the values of $r$ comparable to 1.

If $n \gg r$, the term inside the summation can be written as, $\left(1-\frac{|\alpha|^2}{2dn}\right)^{n}|\alpha|^{4r}O(n^{-r})$. Replacing this in Equation (3.31)

$$\begin{aligned}\phi^n\left(\frac{\alpha}{\sqrt{n}}\right) &= \left(1-\frac{|\alpha|^2}{2dn}\right)^n + \left(1-\frac{|\alpha|^2}{2dn}\right)^n |\alpha|^4 O(n^{-1}) \\ &= \left(1-\frac{|\alpha|^2}{2dn}\right)^n \left(1+|\alpha|^4 O(n^{-1})\right) \\ &= \exp\left\{-\frac{|\alpha|^2}{2d}\right\}\left(1+|\alpha|^4 O(n^{-1})\right). \end{aligned} \tag{3.32}$$

In Equation (3.32), we assume $n \to \infty$. Replacing the above equation in Equation (3.30), we get

$$\begin{aligned}|I_1| &\leq \int_{|\alpha|\leq n^{\frac{1}{4}}} \exp\left\{-\frac{|\alpha|^2}{2d}\right\}|\alpha|^4 O(n^{-1})d\alpha \\ &= O(n^{-1})\int_{|\alpha|\leq n^{\frac{1}{4}}} \exp\left\{-\frac{|\alpha|^2}{2d}\right\}|\alpha|^4 d\alpha \\ &= O(n^{-1})c\int_0^{n^{\frac{1}{4}}} \exp\left\{-\frac{r^2}{2d}\right\} r^4 r^{d-1} dr \\ &= O(n^{-1})c\int_0^{n^{\frac{1}{2}}} \exp\left\{-\frac{t}{2d}\right\} t^{(\frac{d}{2}+2)} dt \\ &\leq O(n^{-1})c\int_0^{\infty} \exp\left\{-\frac{t}{2d}\right\} t^{(\frac{d}{2}+2)} dt \\ &\leq O(n^{-1})c'\int_0^{\infty} \left(\frac{1}{2d}\right)^{(\frac{d}{2}+2)} \frac{\exp\left\{-\frac{t}{2d}\right\} t^{(\frac{d}{2}+2)}}{\Gamma(\frac{d}{2}+2)} dt = O(n^{-1}). \end{aligned} \tag{3.33}$$

The term inside the integral in equation (3.33), is the density function of a random variable following the Gamma $\left(\frac{d}{2}+2, \frac{1}{2d}\right)$. Hence, the value of the integral is 1.

From equation (3.33), we have $I_1(n,x) = O(n^{-1})$.

**Approximating $I_2$:**

From Equation (3.26), we have:

$$I_2 = \int_{\mathbb{R}^d} \exp\left\{-\frac{ix\cdot\alpha}{\sqrt{n}}\right\}\exp\left\{-\frac{|\alpha|^2}{2d}\right\} d\alpha.$$

Multiplying and dividing the integral on the right with $(2\pi d)^{\frac{d}{2}}$, we get

$$I_2 = (2\pi d)^{\frac{d}{2}} \int_{\mathbb{R}^d} e^{-i\alpha\frac{x}{\sqrt{n}}}\left[\frac{1}{(2\pi d)^{\frac{d}{2}}} e^{-\frac{|\alpha|^2}{2d}}\right] d\alpha \tag{3.34}$$

In the above equation, we can observe that the term inside the box brackets is the probability density function of a multivariate normal distribution $Z \sim \text{Normal}(0, dI_{d\times d})$. Hence the integral denotes the characteristic function of $Z$ at $-\frac{x}{\sqrt{n}}$. Equation (3.34) can be rewritten to get

$$\begin{aligned} I_2 &= (2\pi d)^{\frac{d}{2}} \phi_Z\left(-\frac{x}{\sqrt{n}}\right) \\ &= (2\pi d)^{\frac{d}{2}} \exp\left\{\left(-\frac{\alpha}{\sqrt{n}}\right)\cdot\left(i0 - \frac{1}{2}dI\left(-\frac{x}{\sqrt{n}}\right)\right)\right\} \\ &= (2\pi d)^{\frac{d}{2}} \exp\left\{-\frac{d|\alpha|^2}{2n}\right\}. \end{aligned} \tag{3.35}$$

Therefore, we have $I_2 = (2\pi d)^{\frac{d}{2}}\, e^{-\frac{|x|^2 d}{2n}}$.

**Approximating $I_3$:**
From Equation (3.27), we have:

$$I_3 = -\int_{|\alpha|\ge n^{\frac{1}{4}}} \exp\left\{-\frac{ix\cdot\alpha}{\sqrt{n}}\right\}\exp\left\{-\frac{|\alpha|^2}{2d}\right\} d\alpha.$$

Taking the absolute value, we get:

$$\begin{aligned} |I_3| &\le \int_{|\alpha|\ge n^{\frac{1}{4}}} \exp\left\{-\frac{|\alpha|^2}{2d}\right\} d\alpha \\ &\le c\int_{n^{\frac{1}{4}}}^{\infty} \exp\left\{-\frac{r^2}{2d}\right\} r^{d-1} dr \\ &\le c'\int_{n^{\frac{1}{4}}}^{\infty} \exp\left\{-\frac{r^2}{4d}\right\}\left(\exp\left\{-\frac{r^2}{4d}\right\} r^{d-2} dr\right) r dr. \end{aligned} \tag{3.36}$$

Consider the term inside the parenthesis. Let $g(r) = \exp\left\{-\frac{r^2}{4d}\right\} r^{d-2}$. It can be observed that $g(0) = 0$ and $\lim_{r\to\infty} g(r) = 0$. The latter follows because $\exp\left\{\frac{x^2}{4d}\right\} \le \frac{1}{d!}\left(\frac{r^2}{4d}\right)^d$. And replacing this in the limit one gets $\lim_{r\to\infty} g(r) \ge \lim_{r\to\infty} c\frac{r^{d-2}}{r^2 d} = 0$. Further, we see that the function $g$ is continuous, hence it must have an upper bound. Replacing this in the equation (3.36)

$$\begin{aligned} |I_3| &\le c'\int_{n^{\frac{1}{4}}}^{\infty} \exp\{-\frac{r^2}{4d}\} r dr \\ &= c''\int_{n^{\frac{1}{2}}}^{\infty} \exp\{-\frac{t}{4d}\} dt \\ &= c'''\left[-e^{-\frac{x}{4d}}\right]_{n^{\frac{1}{2}}}^{\infty} \\ &= ce^{-\frac{n^{\frac{1}{2}}}{4d}} \\ &= O(n^{-1}) \end{aligned} \tag{3.37}$$

This follows because $\lim_{y\to\infty} \frac{e^{-y^{\frac{1}{2}}}}{(\frac{1}{y})} = 0$.

Therefore, $I_3 = O(n^{-1})$.

**Approximating $I_4$:**
From Equation (3.28), we have

$$I_4 = \int_{n^{\frac{1}{4}} \leq |\alpha| \leq rn^{\frac{1}{2}}} e^{-iX\frac{\alpha}{\sqrt{n}}} \phi_X^n\left(\frac{\alpha}{\sqrt{n}}\right) d\alpha$$

$$|I_4| \leq \int_{n^{\frac{1}{4}} \leq |\alpha| \leq rn^{\frac{1}{2}}} \left|\phi^n\left(\frac{\alpha}{\sqrt{n}}\right)\right| d\alpha. \tag{3.38}$$

Recall that for $|\theta| \leq r$, $\phi(\theta) \leq 1 - \frac{|\theta|^2}{4d}$. For $|\alpha| \leq r\sqrt{n}$, $|\frac{\alpha}{\sqrt{n}}| \leq r$. So,

$$|\phi(\frac{\alpha}{\sqrt{n}})| \leq 1 - \frac{|\alpha|^2}{4dn} \leq e^{-\frac{|\alpha|^2}{4dn}}.$$

Hence,

$$\begin{aligned}
|I_4| &\leq \int_{n^{\frac{1}{4}} \leq |\alpha| \leq rn^{\frac{1}{2}}} \left|\phi\left(\frac{\alpha}{\sqrt{n}}\right)\right|^n d\alpha \\
&\leq \int_{n^{\frac{1}{4}} \leq |\alpha| \leq rn^{\frac{1}{2}}} e^{-\frac{|\alpha|^2}{4d}} d\alpha \\
&\leq \int_{|\alpha| \geq n^{\frac{1}{4}}} e^{-\frac{|\alpha|^2}{4d}} d\alpha \\
&= c\int_{n^{\frac{1}{4}}}^{\infty} e^{-\frac{r^2}{4d}} r^{d-1} dr = O(n^{-1}). \qquad (3.39)
\end{aligned}$$

Equation (3.39) follows from the same argument as that for $I_3$. Hence, $I_4 = O(n^{-1})$.

Now we estimate $I(n,x)$ from the above estimates of $I_1$, $I_2$, $I_3$ and $I_4$. From equation (3.29), we have:

$$\frac{(2\pi\sqrt{n})^d}{2} I(n,x) = (2\pi d)^{\frac{d}{2}} e^{-\frac{|x|^2 d}{2n}} + O(n^{-1})$$

$$I(n,x) = 2\left(\frac{d}{2\pi n}\right)^{\frac{d}{2}} e^{-\frac{|x|^2 d}{2n}} + O(n^{-\frac{d}{2}-1}). \tag{3.40}$$

Since $J(n,x) = O(\rho^n)$, where $\rho < 1$, we get

$$p_n(x) = \mathbb{P}\{S_n = x\} = 2\left(\frac{d}{2\pi n}\right)^{\frac{d}{2}} e^{-\frac{|x|^2 d}{2n}} + O(n^{-\frac{d}{2}-1}). \tag{3.41}$$

Since $\overline{p}_n(x) = 2\left(\frac{d}{2\pi n}\right)^{\frac{d}{2}} e^{-\frac{|x|^2 d}{2n}}$, we have

$$E(n,x) = |p_n(x) - \overline{p}_n(x)| = O(n^{-\frac{d}{2}-1}) \tag{3.42}$$

concluding the proof of the Local Central Limit Theorem.

## 4. Expected Number of Collisions and Intersections

**Theorem 4.1.** *Let $S^1$ and $S^2$ be two simple random walks on $\mathbb{Z}^d$ starting at the origin. Let $R_n$ be the number of intersections between the first $n$ steps of $S^1$ and the first $n$ steps of $S^2$. As $n \to \infty$*

$$\mathbb{E}[R_n] = \begin{cases} c_1 n^{\frac{3}{2}} + O(n^{\frac{1}{2}}), & d = 1, \\ c_2 n + O(\ln(n)), & d = 2, \\ c_3 n^{\frac{1}{2}} + O(1), & d = 3, \\ c_4 \ln(n) + O(1), & d = 4, \\ c_d + O(n^{\frac{4-d}{2}}), & d \geq 5. \end{cases} \tag{4.2}$$

*Proof.* Recall from Definition 1.3,

$$R_n = \sum_{i=0}^{n} \sum_{j=0}^{n} 1\{S_i^1 = S_j^2\}.$$

Due to the linearity of expectation,

$$\begin{aligned} \mathbb{E}[R_n] &= \sum_{i=0}^{n} \sum_{j=0}^{n} \mathbb{E}[1\{S_i^1 = S_j^2\}] \\ &= \sum_{i=0}^{n} \sum_{j=0}^{n} \mathbb{P}\{S_i^1 = S_j^2\}. \end{aligned} \tag{4.3}$$

Both $S^1$ and $S^2$ are simple random walks in $\mathbb{Z}^d$. The term on the right is the probability that $S^1$, after $i$ steps, and $S^2$, after $j$ steps, are the the same point, say $x \in \mathbb{Z}^d$.

Fix $i$ and $j$. Let $A$ be the collection of pairs of paths starting from the origin and reaching the same terminal points in $i$ and $j$ steps respectively. Let $B$ be the collection of paths of length $i+j$ such that the terminal points in these paths is at the origin.

$$A = \big\{([0, \omega_1, \ldots, \omega_i = y], [0, \eta_1, \ldots, \eta_j = y]) : \omega_1, \ldots, \omega_i, \eta_1, \ldots, \eta_j \in \mathbb{Z}^d, \\ |\omega_{k-1} - \omega_k| = 1 \text{ for } 1 \leq k \leq i \text{ and } |\eta_{k-1} - \eta_k| = 1 \text{for} 1 \leq k \leq j\big\}$$

$$B = \big\{[0, \omega_1, \ldots, \omega_{i+j-1}, 0] : \omega_1, \ldots, \omega_{i+j-1} \in \mathbb{Z}^d, |\omega_{k-1} - \omega_k| = 1 \text{ for } 1 \leq k \leq i+j\big\}$$

**Claim:** There exists a bijection from $A$ to $B$.

*Proof.* Consider the map $\Psi : A \to B$ such that

$$\Psi([0, \omega_1, \ldots, \omega_i = y], [0, \eta_1, \ldots, \eta_j = y]) = [0, \omega_1, \ldots, \omega_i = y = \eta_j, \ldots, \eta_1, 0]$$

This is the required bijection. □

Recall from definition 1.2, that the simple symmetric random walk on $\mathbb{Z}^d$ is time reversible. Since $S^2$ is a simple symmetric random walk, it is also time reversible. The probability that the random walk $S^2$ occurs is the same as that of it's reverse occuring. Hence, the probability of the two walks $S^1$, after $i$ steps, and $S^2$, after $j$ steps, being at the same point is the same as the probability of a single random walk traversing the path of $S^1$ for the first $i$ steps and the reverse of $S^2$ for the following $j$ steps and returning to 0 in $i+j$ steps.

Therefore, Equation (4.3) becomes,

$$\mathbb{E}[R_n] = \sum_{i=0}^{n}\sum_{j=0}^{n}\mathbb{P}\{S_{i+j}=0\}. \tag{4.4}$$

On replacing $i+j$ with $k$ in equation (4.4), it can be noticed that each value of $k$ corresponds to $k+1$ pairs of $(i,j)$. Equation (4.4) becomes

$$\mathbb{E}[R_n] = \sum_{k=0}^{2n}(k+1)\mathbb{P}\{S_k=0\} = \sum_{k=0}^{2n}(k+1)p_k(0).$$

Furthermore, since $S_0 = 0$ we get $p_0(0) = 1$ and

$$\mathbb{E}[R_n] = 1 + \sum_{k=1}^{2n}(k+1)p_k(0). \tag{4.5}$$

For large $k$, we use the local central limit theorem to estimate the value for $p_k(0)$.

Recall from the local central limit theorem (Theorem 3.17) that $p_k(0)$ can be approximated with $\overline{p}_k(0)$ where

$$\overline{p}_k(0) := 2\left(\frac{d}{2\pi k}\right)^{d/2} e^{-\frac{d|0|^2}{2n}} = 2\left(\frac{d}{2\pi k}\right)^{d/2}.$$

The error term $E(k,0) = |p_k(0) - \overline{p}_k(0)| = O\left(\frac{1}{k^{\frac{d}{2}+1}}\right)$.

Equation (4.5) can be approximated with

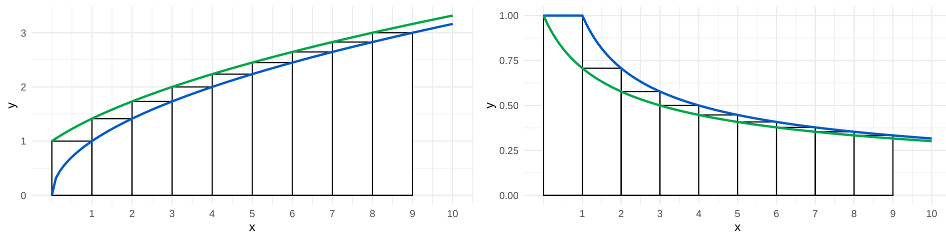
$$\begin{aligned}\mathbb{E}[R_n] &= 1 + \sum_{k=1}^{2n}(k+1)(\bar{p_k}(0) + E(k,0))\\ &= 1 + \sum_{k=1}^{2n}(k+1)2\left(\frac{d}{2\pi k}\right)^{\frac{d}{2}} + \sum_{k=1}^{2n}(k+1)O\left(\frac{1}{k^{\frac{d}{2}+1}}\right)\\ &= 1 + 2\left(\frac{d}{2\pi}\right)^{\frac{d}{2}}\left[\sum_{1}^{2n}\frac{1}{k^{\frac{d}{2}-1}} + \sum_{1}^{2n}\frac{1}{k^{\frac{d}{2}}}\right] + \sum_{k=1}^{2n}(k+1)O\left(\frac{1}{k^{\frac{d}{2}+1}}\right). \end{aligned} \tag{4.6}$$

Let

$$\begin{aligned}T_n &= \sum_{k=1}^{2n}(k+1)O\left(\frac{1}{k^{\frac{d}{2}+1}}\right)\\ &\leq \sum_{k=1}^{2n}(k+1)\frac{c}{k^{\frac{d}{2}+1}} \qquad \text{for some c}\\ &\leq c\left[\sum_{k=1}^{2n}\frac{1}{k^{\frac{d}{2}}} + \sum_{k=1}^{2n}\frac{1}{k^{\frac{d}{2}+1}}\right]. \end{aligned} \tag{4.7}$$

Since, for all $k > 1$,

$$\frac{c}{k^{\frac{d}{2}}} \geq \frac{c}{k^{\frac{d}{2}+1}},$$




(A) A picture proof for Equation 4.10: The curve in blue is that of the equation of $\sqrt{x}$ and the one in green is of $\sqrt{x+1}$. The bars represent the value of $\Sigma^1$ for $d = 1$.

(B) A picture proof for Equation 4.11: The curve in blue is that of the equation of $\frac{1}{\sqrt{x}}$ and the one in green is of $\frac{1}{\sqrt{x+1}}$. The bars represent the value of $\Sigma^1$ for $d = 3$.

FIGURE 5. It can be seen that the area under the green curve is less than that under the bars which is less than that under the blue curve.

we can say that as $n \to \infty$,

$$T_n = O\left(\sum_{k=1}^{2n} \frac{1}{k^{\frac{d}{2}}}\right). \tag{4.8}$$

It remains to estimate the values of the two summations $\Sigma^1 = \sum_{k=1}^{2n} \frac{1}{k^{\frac{d}{2}-1}}$ and $\Sigma^2 = \sum_{k=1}^{2n} \frac{1}{k^{\frac{d}{2}}}$.

We have

$$\mathbb{E}[R_n] = 1 + 2\left(\frac{d}{2\pi}\right)^{\frac{d}{2}} \left[\Sigma^1 + \Sigma^2\right] + O(\Sigma^2) = 1 + 2\left(\frac{d}{2\pi}\right)^{\frac{d}{2}} [\Sigma^1] + O(\Sigma^2). \tag{4.9}$$

First we consider $\Sigma^1 = \sum_{k=1}^{2n} \frac{1}{k^{\frac{d}{2}-1}}$ taking different cases for different values of $d$.

**For d=1:** $\Sigma^1 = \sum_{k=1}^{2n} \sqrt{k}$.

A picture proof for the following statement can be seen in Figure 5a.

$$\int_0^{2n} \sqrt{x}dx \leq \sum_{k=1}^{2n} \sqrt{n} \leq \int_0^{2n} \sqrt{x+1}dx$$

$$\left[\frac{2x^{\frac{3}{2}}}{3}\right]_0^{2n} \leq \sum_{k=1}^{2n} \sqrt{n} \leq \left[\frac{2(x+1)^{\frac{3}{2}}}{3}\right]_0^{2n}$$

(4.10)

$$\frac{2(2n)^{\frac{3}{2}}}{3} \leq \sum_{k=1}^{2n} \sqrt{n} \leq \frac{2(2n)^{\frac{3}{2}}}{3} + \frac{3}{2} \times \frac{2}{3}(2n)^{\frac{1}{2}} + \frac{3}{2} \times \frac{1}{2} \times \frac{2}{3}(2n)^{-\frac{1}{2}} + \cdots = \frac{2(2n)^{\frac{3}{2}}}{3} + O(n^{\frac{1}{2}}).$$

Therefore, for $d = 1$, $\Sigma^1 = \frac{2(2n)^{\frac{3}{2}}}{3} + O(n^{\frac{1}{2}})$.

**For d=2:** $\Sigma^1 = \sum_{k=1}^{2n} 1 = 2n$.

**For d=3:** $\Sigma^1 = \sum_{k=1}^{2n} \frac{1}{k^{\frac{1}{2}}}$.

A picture proof for the following statement can be seen in Figure 5b.

$$\int_0^{2n} \frac{1}{\sqrt{x+1}}dx \leq \sum_{k=1}^{2n} \frac{1}{k^{\frac{1}{2}}} \leq 1 + \int_1^{2n} \frac{1}{\sqrt{x}}dx$$

$$\left[2\sqrt{x+1}\right]_0^{2n} \leq \sum_{k=1}^{2n} \frac{1}{k^{\frac{1}{2}}} \leq 1 + \left[2\sqrt{x}\right]_1^{2n}$$

$$2\sqrt{2n+1} - 2 \leq \sum_{k=1}^{2n} \frac{1}{k^{\frac{1}{2}}} \leq 1 + 2\sqrt{2n} - 2$$

$$2\sqrt{2n} - 2 \leq \sum_{k=1}^{2n} \frac{1}{k^{\frac{1}{2}}} \leq 1 + 2\sqrt{2n} - 2. \tag{4.11}$$

Therefore, for $d = 3$, $\Sigma^1 = 2\sqrt{2n} + O(1)$.

**For d=4:** $\Sigma^1 = \sum_{k=1}^{2n} \frac{1}{k}$.

The picture proof for the following statement looks very much like Figure 5b.

$$\int_0^{2n} \frac{1}{x+1}dx \leq \sum_{k=1}^{2n} \frac{1}{k} \leq 1 + \int_1^{2n} \frac{1}{x}dx$$

$$\int_0^{2n-1} \frac{1}{x+1}dx \leq \sum_{k=1}^{2n} \frac{1}{k} \leq 1 + [\log(x)]_1^{2n}$$

$$\log(2n) \leq \sum_{k=1}^{2n} \frac{1}{k} \leq 1 + \log(2n). \tag{4.12}$$

Therefore, for $d = 4$, $\Sigma^1 = \log(2n) + O(1)$.

**For d$\geq$ 5:** $\Sigma^1 = \sum_{k=1}^{2n} \frac{1}{k^{\frac{d}{2}-1}}$.

First, we show that $\sum_{k=1}^{\infty} \frac{1}{k^{\frac{d}{2}-1}} = c_d < \infty$.

$$\begin{aligned}
\sum_{k=1}^{m} \frac{1}{k^{\frac{d}{2}-1}} &\leq 1 + \int_0^m \frac{1}{x^{\frac{d}{2}-1}}dx \\
&\leq 1 + \left[\frac{2}{4-d}x^{(2-\frac{d}{2})}\right]_1^m \\
&\leq 1 + \frac{2}{d-4} - \frac{2m^{(2-\frac{d}{2})}}{4-d} \leq 1 + \frac{2}{d-4}. \qquad (4.13)
\end{aligned}$$

Since the series $\sum_{k=1}^{\infty} \frac{1}{k^{\frac{d}{2}-1}}$ is bounded (as seen above) and increasing, it converges to some $c_d < \infty$.

We observe that

$$\begin{aligned}\sum_{k=1}^{2n}\frac{1}{k^{\frac{d}{2}-1}} &= \sum_{k=1}^{\infty}\frac{1}{k^{\frac{d}{2}-1}} - \sum_{k=2n+1}^{\infty}\frac{1}{k^{\frac{d}{2}-1}}\\ &\geq c_d - \int_{2n}^{\infty}\frac{1}{x^{\frac{d}{2}-1}}dx\\ &\geq c_d - \left[\frac{2}{(4-d)x^{\frac{d}{2}-2}}\right]_{2n}^{\infty} = c_d + \frac{2}{(d-4)(2n)^{\frac{d}{2}-2}}.\end{aligned}$$

Furthermore, we have

$$\begin{aligned}\sum_{k=1}^{\infty}\frac{1}{k^{\frac{d}{2}-1}} \geq \sum_{k=1}^{2n}\frac{1}{k^{\frac{d}{2}-1}} &\geq c_d + \frac{2}{(d-4)(2n)^{\frac{d}{2}-2}}\\ c_d \geq \sum_{k=1}^{2n}\frac{1}{k^{\frac{d}{2}-1}} &\geq c_d + \frac{2}{(d-4)(2n)^{\frac{d}{2}-2}}.\end{aligned}$$

Therefore, for $d=5$, $\Sigma^1 = c_d + O(\frac{1}{n^{\frac{d}{2}-2}})$.

The estimate for $\Sigma^2$ for the different dimensions can be done similarly. Hence, we have

$$\Sigma^1 = \sum_{k=1}^{2n}\frac{1}{k^{\frac{d}{2}-1}} = \begin{cases} \frac{2(2n)^{\frac{3}{2}}}{3} + O(\sqrt{n}), & d=1,\\ 2n, & d=2,\\ 2\sqrt{2n}+O(1), & d=3,\\ \log(2n)+O(1), & d=4,\\ c_d + O(\frac{1}{n^{\frac{d}{2}-2}}), & d\geq 5.\end{cases} \tag{4.14}$$

$$\Sigma^2 = \sum_{k=1}^{2n}\frac{1}{k^{\frac{d}{2}}} = \begin{cases} 2\sqrt{2n}+O(1), & d=1,\\ log(2n)+O(1), & d=2,\\ c_d + O(\frac{1}{n^{\frac{d}{2}-1}}), & d\geq 3.\end{cases} \tag{4.15}$$

Replacing the values from Equations (4.14) and (4.15) in Equation (4.9), the following result is obtained, proving Theorem 4.1.

$$\mathbb{E}[R_n] = \begin{cases} \frac{8}{3\sqrt{\pi}}n^{\frac{3}{2}} + O(\sqrt{n}), & d=1,\\ \frac{4}{\pi}n + O(\log(n)), & d=2,\\ \frac{6}{\pi}\sqrt{\frac{3}{pi}}\sqrt{n}+O(1), & d=3,\\ \frac{8}{\pi^2}\log(n)+O(1), & d=4,\\ c_d + O(\frac{1}{n^{\frac{d}{2}-2}}), & d\geq 5.\end{cases} \tag{4.16}$$

□

**Theorem 4.17.** *Let $S^1$ and $S^2$ be two simple random walks on $\mathbb{Z}^d$ starting at the origin. Let $C_n$ be the number of collisions occuring in the first $n$ steps of $S^1$ and $S^2$. As $n\to\infty$*

$$\mathbb{E}[C_n] = \begin{cases} c_1\sqrt{n}+O(1), & d=1,\\ c_2\log(n)+O(1), & d=2,\\ c_3 + O(\frac{1}{n^{\frac{d}{2}}}), & d\geq 3.\end{cases} \tag{4.18}$$

*Proof.* The proof for the expected number of collisions is much like the one of their intersections. From the linearity of expectation we have

$$\begin{aligned}\mathbb{E}[C_n] &= \sum_{i=1}^{n} \mathbb{E}[1\{S_i^1 = S_i^2\}] \\ &= \sum_{i=1}^{n} \mathbb{P}\{S_i^1 = S_i^2\}. \end{aligned} \tag{4.19}$$

Since the simple symmetric random walk on $\mathbb{Z}$ is time reversible, using the same argument as in the case of intersections can be done. It is observed that the probability of two random walks colliding on the $i^{\text{th}}$ step is the same as that of a single random walk returning to the origin in $2i$ steps. Hence Equation (4.19) becomes

$$\mathbb{E}[C_n] = \sum_{i=0}^{n} \mathbb{P}\{S_{2i} = 0\}, \tag{4.20}$$

where $S$ denotes a simple symmetric random walk on $\mathbb{Z}^d$ starting at the origin.

To evaluate the right side of Equation (4.20), the local central limit theorem (Theorem 3.17) is used. The analysis is the same as the the one in the previous section in the case of intersections.

The final result obtained is

$$\mathbb{E}[C_n] = \begin{cases} \frac{4}{\sqrt{\pi}}\sqrt{n} + O(1), & d = 1, \\ \frac{2}{\pi}\log(n) + O(1), & d = 2, \\ c_3 + O(\frac{1}{n^{\frac{d}{2}}}), & d \geq 3. \end{cases} \tag{4.21}$$

□

## Acknowledgments

I would like to thank my mentor, Jin Woo Sung, for his instruction, guidance, and endless patience over the course of the summer. I greatly appreciate all the time and effort he put in answering all the questione I had. I would also like to thank Professor Gregory Lawler for his lecture series on "Probability Random Fields, and Geometry in Statistical Physics" for introducing many ideas used in this paper. Furthermore, I would also like to thank Professor Siva Athreya and Professor Yogeshwaran D. for feedback on an earlier draft. Last, but not least, I would like to thank Professor Peter May for organising this REU online. I am very grateful to have gotten the opportunity to participate.

## References

[1] P´olya, G., *George Pòlya: Collected Papers* **volume IV**, 582–585. The MIT Press, Cambridge, Massachusetts

[2] Lawler, Gregory F. *Intersections of random walks.* Springer Science & Business Media, (2013).

[3] Lawler, Gregory & Limic, Vlada. *Random Walk: A Modern Introduction.* 10.1017/CBO9780511750854. (2010).

[4] https://encyclopediaofmath.org/wiki/Fubini_theorem