THE RANGE OF TWO DIMENSIONAL SIMPLE RANDOM WALK

Jian (Kevin) Jiao

A THESIS

in

Mathematics

Presented to the Faculties of the University of Pennsylvania in Partial

Fulfillment of the Requirements for the Degree of Master of Arts

2014

Robin Pemantle
Supervisor of Thesis

David Harbater
Graduate Group Chairman

Abstract

The goal of this paper is to prove that the range of a two dimensional simple random walk at time n has roughly the size of the form $cn/\log n$ where c is a constant. We begin by decomposing the range using several new random variables. The whole proof also requires generating functions and Tauberian theorem on series convergence. We show that the range not only has a expected size of $\pi n/\log n$, but also converges to it in probability.

1 Introduction

Random walk is one of the most studied topics in probability theory. The intuition of the simplest random walk process is the following: you start from somewhere on a straight line. Every time before you move, you flip a coin to decide which way to go. If the head turns up, you move one step to the right. Otherwise, you move one step to the left (or vice versa). The simple one dimensional random walk is the random variable that determines your location after n tosses.

In this paper, we are interested in the range of two dimensional simple random walk on integer lattices. We start by setting the basic definitions and notations. The standard basis of vectors in $\mathbb{Z}^2$ is denoted by $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$. Our discrete time, simple random walk starts from the origin $(0, 0) \in \mathbb{Z}^2$. The random walk variable S_n can be considered the sum of a sequence of independent, identically distributed random variables

$$S_n = X_1 + X_2 + \cdots + X_n \tag{1.1}$$

where $\mathbb{P}\{X_i = e_k\} = \mathbb{P}\{X_i = -e_k\} = 1/4$, $k = 1, 2, i = 1, 2, \ldots, n$. Hence, for each step there are four choices on which direction to go. All four directions have equal chances of being chosen.

Next, we need to define the transition probability. In general, the n-step distribution is denoted by $p_n(x, y)$ with

$$p_n(x, y) = \mathbb{P}\{S_n = y | S_0 = x\} \tag{1.2}$$

Since we have $S_0 = 0$ by assumption, we use $p_n(x)$ for $p_n(x, 0)$. Note that p_n describes the distribution of the location variable after n steps.

Our goal is to give an estimation of the size of the two dimensional random walk range. Throughout the paper, the term *step* and *time* are interchangeable, and n-th step and time n have the same meaning. The range of an n-step random walk, R_n, is a random variable that characterizes the number of distinct points visited at time n:

$$R_n = card\{S_0, S_1, \ldots, S_n\}, \forall n \in \mathbb{N} \tag{1.1}$$

In the following sections, we will examine various properties of R_n and $p_n(x)$. In section two, the range is decomposed and analyzed. In section three and four, we compute the expected value and the variance of R_n, respectively. The generating function method follows some of the arguments in [LL10].

2 Decomposition of the Range

First, we create a new random variable γ_n, $n \in \mathbb{N}^+$

$$\gamma_n = \begin{cases} 1 & \text{if } n\text{-th step hits a new spot} \\ \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

Note that $R_n = \sum_{i=1}^{n} \gamma_i$ simply because γ exactly characterizes the chance of visiting a new location in each step, thus summing up to the total number of distinct locations visited after n steps. Now we have our first proposition on the property of γ.

Proposition 1. *The expected value of γ_n is equal to the probability of never going back to the origin in n steps.*

Proof. we can decompose γ_n in the following way:

$$\begin{aligned}
\mathbb{E}[\gamma_n] &= \mathbb{P}\{n\text{-th step hits a new spot}\} \\
&= \mathbb{P}\{S_n \neq S_{n-1}, S_n \neq S_{n-2}, \ldots, S_n \neq S_0 = 0\} \\
&= \mathbb{P}\{S_n - S_{n-1} \neq 0, S_n - S_{n-2} \neq 0, \ldots, S_n \neq 0\} \\
&= \mathbb{P}\{X_n \neq 0, X_n + X_{n-1} \neq 0, \ldots, X_n + X_{n-1} + \cdots + X_1 \neq 0\} \\
&= \mathbb{P}\{X_1 \neq 0, X_1 + X_2 \neq 0, \ldots, X_1 + X_2 + \cdots + X_n \neq 0\} \\
&= \mathbb{P}\{S_i \neq 0, i = 1, 2, \ldots, n\}
\end{aligned}$$

□

we reverse the path between the fourth and the fifth line because all X_i are *i.i.d* random variables. Since $p_n(0) = \mathbb{P}\{S_n = 0\} = 0$ for all odd n, we extend our index to $2n$ to study some properties of $p_{2n}(0)$ and γ_{2n}.

Proposition 2. *Let $u_n(0)$ denote the probability of visiting the origin for the first time (not including $n = 0$). We have the following recurrence relation:*

$$\sum_{i=1}^{n} u_{2i}(0)p_{2n-2i}(0) = p_{2n}(0) \tag{2.2}$$

Proof. $p_{2n}(0)$ simply represents the probability of arriving at the origin at time $2n$. Using the basic counting technique and conditioning on the first visit to the origin, we have n different cases. Because the first visiting time is a stopping time, by Strong Markov Property, the process after $2i$-th step is identical to the random walk starting from $(0, 0)$ with $2n - 2i$ future steps. Therefore, $u_{2i}(0)p_{2n-2i}(0)$ is the probability of visiting the origin for the first time at $2i$-th step and finally coming back to the origin after another $2n - 2i$ steps. Adding all different cases, we have

$$p_{2n}(0) = u_2(0)p_{2n-2}(0) + u_4(0)p_{2n-4}(0) + \cdots + u_{2n}(0)p_0(0)$$

which means $\sum_{i=1}^{n} u_{2i}(0)p_{2n-2i}(0) = p_{2n}(0)$. □

So how can we estimate $p_{2n}(0)$? In literature, it is popular to use local central limit theorem for high dimensional random walks. Here, we can use a combinatorial approach for the two dimensional case. A $2n$-step random walk has in total 4^{2n} choices. In order to get back to the origin at time $2n$, the walker must have k steps $(0 \le k \le n)$ moving towards the direction of $\mathbf{e_1}$ and the other k steps towards the exact opposite direction $(-\mathbf{e_1})$. Similarly, there are $n-k$ steps towards $\mathbf{e_2}$ and another $n - k$ steps towards $-\mathbf{e_2}$. As a result, we are able to write $p_{2n}(0)$ as

$$p_{2n}(0) = \mathbb{P}\{S_{2n} = 0\} = 4^{-2n} \sum_{k=0}^{n} \frac{(2n)!}{k!k!(n-k)!(n-k)!} = 4^{-2n} \binom{2n}{n}\binom{2n}{n} \tag{2.3}$$

Using the Stirling's formula, we have

$$\begin{aligned} p_{2n}(0) &= 4^{-2n}\binom{2n}{n}\binom{2n}{n} \\ &= \frac{(2n)!^2}{4^{2n}n!^4} \\ &\sim \frac{2\pi 4^{2n+(1/2)}n^{4n+1}e^{-4n}}{4^{2n}4\pi^2 n^{4n+2}e^{-4n}} \\ &= \frac{1}{n\pi} \end{aligned}$$

This formula holds for $n \in \mathbb{N}^+$. When $n = 0$, $p_0(0) = \mathbb{P}\{S_0 = 0\} = 1$.

3 Estimating the Expectation

For the two dimensional random walk, the *Green's generating function* is defined to be

$$G(x) = \sum_{n=0}^{\infty} x^n p_n(0) \tag{3.1}$$

Since $p_n(0) \in (0,1)$ when $n > 1, n$ even, the power series converges absolutely when $|x| < 1$.

Similarly, we can define the generating function of first visits to be

$$F(x) = \sum_{n=1}^{\infty} x^n u_n(0) \tag{3.2}$$

The sum also converges absolutely when $|x| < 1$. We can then prove the following proposition:

Proposition 3.

$$G(x) = \frac{1}{1 - F(x)} \tag{3.3}$$

Proof. By **Proposition 2**, we can multiply both sides by x^n and rearrange to get

$$\begin{aligned} x^n p_{2n}(0) &= x^n \sum_{i=1}^{n} u_{2i}(0) p_{2n-2i}(0) \\ &= \sum_{i=1}^{n} \left(x^i u_{2i}(0)\right) \left(x^{n-i} p_{2n-2i}(0)\right) \end{aligned}$$

Including odd terms does not affect this equation. As $n \to \infty$, we have

$$\sum_{n=1}^{\infty} x^n p_n(0) = \left[\sum_{n=1}^{\infty} x^n u_n(0)\right] \left[\sum_{m=0}^{\infty} x^m p_m(0)\right] \tag{3.4}$$

which leads to

$$G(x) - p_0(0) = F(x) G(x)$$

$$G(x)\,(1-F(x)) = 1$$

Since $0 < F(x) < 1$, we have

$$G(x) = \frac{1}{1-F(x)}$$

$\square$

Now we need to build the link between γ_n and $u_n(0)$. Let $\mathbb{E}[\gamma_n] = b_n$, we have the following proposition:

Proposition 4. *For $x \in (0,1)$,*

$$(1-x)\sum_{n=0}^{\infty} x^n b_n = 1 - \sum_{n=1}^{\infty} x^n u_n(0) \tag{3.5}$$

Proof. By **Proposition 1**,

$$b_n = \mathbb{E}[\gamma_n] = \mathbb{P}\{S_i \neq 0, i = 1, 2, \ldots, n\}$$

Consider the probablistic meaning behind these two sums. Think of $1-x$ as the geometric *killing rate* of the random walk. That is, for each step the walker has $1-x$ probability of being killed. Since $u_n(0) = \mathbb{P}\{S_1 \neq 0, \ldots, S_{n-1} \neq 0, S_n = 0\}$, $\sum_{n=1}^{\infty} x^n u_n(0)$ represents the probability of returning the origin before being killed. Thus, the right hand side of (3.5) is the probability of being killed without ever getting back to the origin. This is exactly the probablistic meaning of the left hand side. $\square$

Now that we have figured out the relationship between generating functions of b_n, u_n, and p_n, we can derive the explicit form of the generating function of b_n.

Proposition 5. *The generating function of b_n has the form*

$$\sum_{n=0}^{\infty} x^n b_n \sim \frac{\pi}{1-x}\left[\log\left(\frac{1}{1-x}\right)\right]^{-1} \tag{3.6}$$

when $x \to 1-$.

Proof. By the definition of $G(x)$ and the estimation of $p_n(0)$ derived in section 2, we have

$$G(x) = \sum_{n=0}^{\infty} x^n p_n(0) \sim 1 + \sum_{n=1}^{\infty} \frac{x^{2n}}{n\pi} \tag{3.7}$$

the power series is convergent for $|x| < 1$. So as $x \to 1-$, we can differentiate term-by-term to get

$$\frac{\mathrm{d}}{\mathrm{d}x}G(x) \sim \frac{2}{\pi}\sum_{n=1}^{\infty} x^{2n-1} = \frac{2x}{\pi(1-x^2)}$$

$$G(x) \sim \frac{1}{\pi}\left[\log\left(\frac{1}{1-x}\right)\right] \tag{3.8}$$

Together with **Proposition 3** and **4**, we have

$$\sum_{n=0}^{\infty} x^n b_n = \frac{1}{1-x}\left(1 - F(x)\right) = \frac{1}{1-x}G(x)^{-1} \sim \frac{\pi}{1-x}\left[\log\left(\frac{1}{1-x}\right)\right]^{-1}$$

□

Now it is possible to give an estimation of the expected range. This next proposition is proved using some of the Tauberian thoughts on series convergence.

Proposition 6.

$$\sum_{i=0}^{n} b_i \sim \frac{n\pi}{\log n}, \qquad n \to \infty \tag{3.9}$$

Proof. This result is a simplified version of Proposition A.5.3 of [LL10]. In particular, we use the special case when $\alpha = -1$. □

Therefore, the expected range of two dimensional simple random walk at time n is approximately $n\pi/\log n$. This is a result we desired.

4 Estimating the Variance

Now we want to estimate the variance for the range of two dimensional simple random walk. Again, we need to define a new random variable to calculate the second moment of R_n.

$$\gamma_{m,n} = \begin{cases} 1 & \text{if both } m\text{-th and } n\text{-th step hit a new spot} \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

Similarly, let $b_{m,n}$ denote $\mathbb{E}[\gamma_{m,n}]$. We can use the following decomposition, given the condition that $1 \leq m \leq n$:

$$\begin{aligned} b_{m,n} &= \mathbb{P}\{S_m \neq S_{m-1}, \ldots, S_m \neq S_0; S_n \neq S_{n-1}, \ldots, S_n \neq S_0\} \\ &\leq \mathbb{P}\{S_m \neq S_{m-1}, \ldots, S_m \neq S_0; S_n \neq S_{n-1}, \ldots, S_n \neq S_m\} \\ &= \mathbb{P}\{S_m \neq S_{m-1}, \ldots, S_m \neq S_0\} * \mathbb{P}\{S_{n-m} \neq S_{n-m-1}, \ldots, S_{n-m} \neq S_0\} \\ &= b_m * b_{n-m} \end{aligned}$$

Together with the fact that $\mathbb{E}[R_n^2] = \sum_{i,j=0}^{n} b_{i,j}$, we have

$$\begin{aligned} Var[R_n] &= \mathbb{E}[R_n^2] - \mathbb{E}[R_n]^2 \\ &= \sum_{i,j=0}^{n} b_{i,j} - \sum_{i=0}^{n} b_i \sum_{j=0}^{n} b_j \\ &= \sum_{i=0}^{n} \sum_{j=0}^{n} (b_{i,j} - b_i b_j) \\ &\leq 2 \sum_{0 \leq i \leq j \leq n} (b_{i,j} - b_i b_j) \\ &\leq 2 \sum_{0 \leq i \leq j \leq n} (b_i b_{j-i} - b_i b_j) \end{aligned}$$

The next step involves using the property of the sequence $\{b_n\}$. Remember that $b_n = \mathbb{P}\{S_i \neq 0, i = 1, 2, \dots, n\}$. It is a decreasing sequence, which leads to the fact that

$$\sum_{j=i}^{n}(b_{j-i} - b_j) = (b_0 + b_1 + \cdots + b_{n-i}) - (b_i + b_{i+1} + \cdots + b_n)$$

is maximized when we let i be around $n/2$ (round off $n/2$ downward). Therefore, we have

$$\begin{aligned}
Var[R_n] &\leq 2 \sum_{0\leq i\leq j\leq n} (b_i b_{j-i} - b_i b_j) \\
&= 2\sum_{i=0}^{n} b_i \sum_{j=i}^{n} (b_{j-i} - b_j) \\
&\leq 2\sum_{i=0}^{n} b_i \left(\sum_{i=0}^{n/2} 2b_i - \sum_{i=0}^{n} b_i \right) \\
&= 2\mathbb{E}[R_n] \left(2\mathbb{E}[R_{\frac{n}{2}}] - \mathbb{E}[R_n]\right)
\end{aligned}$$

Plug in the value of $\mathbb{E}[R_n]$ derived in **Proposition 6**, we can obtain the upper bound of the variance

$$Var[R_n] \leq \frac{2n\pi}{\log n} \left(\frac{n\pi}{\log \frac{n}{2}} - \frac{n\pi}{\log n} \right) = \frac{n^2\pi^2 \log 4}{\log \frac{n}{2} \log n \log n} \tag{4.2}$$

Now we can apply Chebyshev's Inequality to show that R_n obeys the weak law of large numbers.

Proposition 7.

$$\bar{R_n} \xrightarrow{p} \frac{n\pi}{\log n} \tag{4.3}$$

when $n \to \infty$.

Proof. First, we define $\sigma_n = Var[R_n]^{\frac{1}{2}}$ and $\phi_n = \sqrt{\frac{\log \frac{n}{2}}{\log 4}}$. By (4.2), the standard deviation of R_n is

$$\sigma_n \leq \sqrt{\frac{n^2\pi^2 \log 4}{\log \frac{n}{2} \log n \log n}} \sim \mathbb{E}[R_n]\sqrt{\frac{\log 4}{\log \frac{n}{2}}} = \mathbb{E}[R_n]\phi_n^{-1} \tag{4.4}$$

Therefore, by Chebyshev's Inequality,

$$\mathbb{P}\left(\left|R_n - \frac{n\pi}{\log n}\right| \geq \frac{n\pi}{\log n}\right) = \mathbb{P}\left(\left|R_n - \frac{n\pi}{\log n}\right| \geq \phi_n\sigma_n\right) \leq \frac{1}{\phi_n^2}$$

This result gives a good estimation of the distribution of realized random walk range. Meanwhile, as $n \to \infty$, we have

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}\left(\left|R_n - \frac{n\pi}{\log n}\right| \geq \phi_n\sigma_n\right) &\leq \lim_{n\to\infty} \frac{1}{\phi_n^2} \\ &= \lim_{n\to\infty} \frac{\log 4}{\log \frac{n}{2}} \\ &= 0 \end{aligned}$$

$\square$

This result gives a good estimation of the distribution of realized random walk range. Because

$$\mathbb{P}\left(\left|R_n - \frac{n\pi}{\log n}\right| \geq \frac{n\pi}{k\log n}\right) = \mathbb{P}\left(\left|R_n - \frac{n\pi}{\log n}\right| \geq \frac{\phi_n\sigma_n}{k}\right) \leq \frac{k^2}{\phi_n^2}$$

as $n \to \infty$, we are able to bound the realized range almost surely for any $k > 0$. For example, when $k = 2$, we can construct the following table:

n	min % within half expected range	max % beyond half expected range
10	0%	100%
20	0%	100%
30	0%	100%
40	0%	100%
50	13.1%	86.9%
60	35.1%	64.9%
70	52.9%	47.1%
80	67.6%	32.4%
90	80.2%	19.8%
100	91.1%	8.9%

References

[Cer07] J. Cerny. Moments and Distribution of the Local Time of a Two Dimensional Random Walk. *Stochastic Processes and their Applications*, Vol. 117, p. 262-270, 2007

[DE51] A. Dvoretzky and P. Erdos. Some Problems on Random Walk in Space. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, p. 353-367, 1951

[FO90] P. Flajolet and A. Odlyzko. Singularity Analysis of Generating Functions. *SIAM Journal of Discrete Mathematics*, Vol. 3, p. 216-240, 1990

[JP72] N. Jain and W. Pruitt. The Range of Random Walk. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 3, p. 31-50, 1972

[LL10] G. Lawler and V. Limic. *Random Walk: A Modern Introduction*, Cambridge University Press, 2010

[LR91] J. Le Gall and J. Rosen. The Range of Stable Random Walks. *The Annals of Probability*, Vol. 19, p. 650-705, 1991

[Spi01] F. Spitzer. *Principles of Random Walk*, Second Edition, Springer, 2001