

# Cover Times of Random Walks

## MATH 0710 Thesis

Nick Bermingham

May 11, 2020

### Abstract

If a random walk is performed on a connected graph, a cover time is the number of steps required for each vertex to be visited at least once. To begin, we discuss in simple language a few examples drawn from graph theory, until we find that our developed strategy fails with larger, unstructured graphs. We then look at the cover time distribution of a complete graph with self-edges, and prove that the number of unvisited vertices approaches a Poisson distribution in a special case. Finally, we introduce Pólya's Recurrence Theorem, which states that a simple random walk on a  $d$ -dimensional lattice is recurrent for  $d = 1, 2$ , and transient for  $d > 2$ . We discuss the 1-dimensional, 2-dimensional, and 3-dimensional components of this theorem using The Cliff Hanger, The Drunk Man, and The Drunk Bird problems, respectively.

## Contents

<b>1</b>	<b>Cover Times</b>	<b>3</b>
1.1	Preliminaries . . . . .	3
1.2	A Special Graph . . . . .	3
1.3	Polygons . . . . .	6
1.4	Complete Graphs . . . . .	9
1.5	Cover Time Distribution of a Complete Graph . . . . .	10
<b>2</b>	<b>Pólya's Recurrence Theorem</b>	<b>14</b>
2.1	One-Dimensional Random Walk: The Cliff-Hanger . . . . .	15
2.2	Stirling's Approximation Proof . . . . .	16
2.3	Two-Dimensional Random Walk: The Drunk Man . . . . .	19
2.4	Three-Dimensional Random Walk: The Drunk Bird . . . . .	21
<b>A</b>	<b>Random Walk on a Chessboard: Source Code</b>	<b>25</b>

# 1 Cover Times

## 1.1 Preliminaries

We define a graph  $G = (V, E)$  with vertex set  $V$  and edge set  $E$ . We say  $|V| = n$ , and  $|E| = e$ , where  $n$  and  $e$  can be infinite. A graph is considered to be *connected* if

$$\forall v_1, v_2 \in V, \exists \{e_1, e_2, \dots\} \subseteq E \text{ that connects } v_1 \rightarrow v_2. \quad (1)$$

Consider the following random walk on graph  $G$ . A particle is at a certain *starting-vertex* at time  $t = 0$ , which we can denote vertex  $i$ . From vertex  $i$ , the particle can move to any of the vertexes with which this vertex is directly connected at  $t = 1$ , each with probability  $1/d_i$ , where  $d_i$  is the number of edges incident to  $i$ . This process continues for  $t = 2, 3, \dots$ . When each choice in a random walk is made with equal probability, a graph is considered to be *unweighted*. A weighted graph is presented in the discussion of Pólya's Recurrence Theorem, in the Gambler's Ruin problem.

The *cover time* - the expected time until all vertices have been visited - can be expressed as  $m_{in}$ . We can also call this the *absorption time*. When determining a simple cover time, we assume that the particle stops when the each vertex has been reached. But what if we wanted to let the walk continue until the particle returns to its starting vertex? We call this the *extended cover time*, and we can express this as  $u_{in}$ . Clearly,  $u_{in} > m_{in}$  for all graphs.

It can also be of interest to consider the expected time for given number  $k < n$  vertices to be visited. We denote this expected value  $m_{ik}$ , and call it a *partial cover time*. If we wish to again compute the time for the particle to return to the starting vertex, we can denote this  $u_{ik}$  - the *extended partial cover time*.

All of the different kinds of cover times can be independent of the starting-vertex, and can always be found using an appropriate system of linear equations. This process is explained in the following example, and follow from [1].

## 1.2 A Special Graph

To begin, let us examine the graph in figure 1, where particle rests at vertex 1 at  $t = 0$ . From this starting vertex, the particle can move to vertex 2, 3, or 4 - each with probability  $1/3$ . In order to determine the partial cover time  $m_{13}$ , the expected time until  $k = 3$  vertices have been visited, we consider each of the possible first moves for the particle.

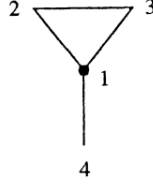


Figure 1: A Special Graph [1]

1. The particle moves first to vertex 2 with probability  $1/3$ . We can denote the expected time required to visit a third and final vertex  $a$ . This case is represented in the leftmost illustration in figure 2. The bolded vertex represents the vertex where the particle is present, and the circled vertex represents a vertex that has already been visited.
2. The particle moves first to vertex 3 with probability  $1/3$ . The expected time to visit a third vertex is again  $a$ , by symmetry.
3. The particle moves first to vertex 4 with probability  $1/3$ . We can set the expected remaining time as  $b$ .

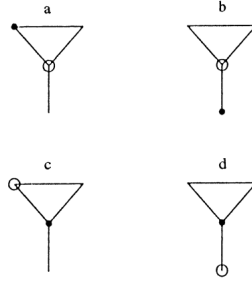


Figure 2: Random walk on the special graph [1]

Combining the cases enumerated above, we obtain the following equation, where 1 represents the step already taken to reach one of the vertices

$$m_{13} = 1 + \frac{2}{3} \cdot a + \frac{1}{3} \cdot b \quad (2)$$

If we allow the particle to continue its walk from its current vertex until three vertices have been visited, we can calculate remaining times that correspond to the graphs in figure 2. Thus, we have the four following additional equations

$$\begin{aligned}
a &= 1 + \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot c & b &= 1 + d \\
c &= 1 + \frac{2}{3} \cdot 0 + \frac{1}{3} \cdot a & d &= 1 + \frac{2}{3} \cdot a + \frac{1}{3} \cdot b
\end{aligned} \tag{3}$$

Each letter in these equations represents the particle being positioned according to the corresponding illustration, and a 0 means the particle has reached 3 vertices and requires no additional traversing across the graph. Each coefficient represents the likelihood that the particle will choose each vertices (accounting for symmetry).

If we plug these four equations into the original partial cover time equation, we find that  $m_{13} = 16/5$ . Using this general strategy, we can derive similar equations to compute the cover time, partial cover times, extended cover time, and extended partial cover times on this special graph. Used on larger, more complicated graphs, this strategy quickly becomes tedious. To demonstrate this, we briefly study the following example: let the particle rest on the eight corners of a cube - compute the mean number of steps required to pass across all eight edges. Although ostensibly simple, this computation involves a system of equations of 387 unknowns (compared to five unknowns in our special graph).

If we place a king somewhere on an empty chessboard, how many random moves does it take for him to reach all 64 squares? Clearly a system of equations would be futile, but we can make a strong estimate of the expected value by calculating the mean of a large number of trials. Using a simulation I created of a 250 thousand trials, the mean cover time of a king is 615 moves. If we put a knight through the same experiment, we get a mean of 556 moves.

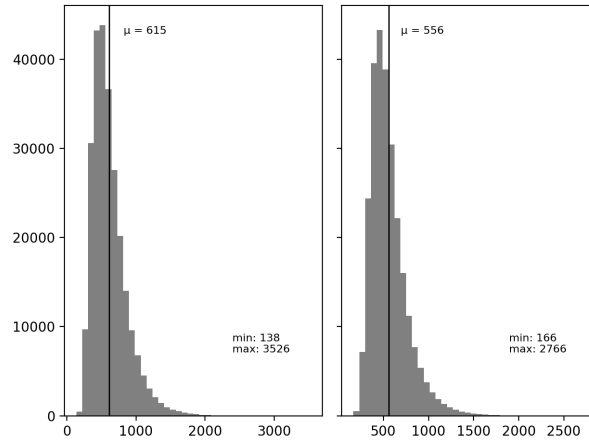


Figure 3: Random walk of a king and a knight on an empty chessboard, where the initial position is chosen at random.

The main method of this code can be found in the appendix. It may come as a surprise to readers that a knight has a smaller mean value. Although we won't use simulation to determine any more cover times in this paper, I thought that this chess board example might serve as a simple yet relatable example of a cover time, and that the result of my simulation might be found interesting. [1]

### 1.3 Polygons

Next, we consider a polygon with  $n$  vertices, numbered  $0, 1, \dots, n-1$ . A particle starts at vertex 0 and moves to one of the two adjacent corners, 1 or  $n-1$  in this case. This binary choice occurs at time  $t = 2, 3, \dots$  until all  $n$  vertices have been visited.

In order to compute cover time and extended cover time, we can consider the polygonal graph equivalent to a walk on the x-axis, starting at zero. A hexagon, for example, is equivalent to a number line with range  $[-5, 5]$ . The starting vertex is now the origin, and each corner  $1, \dots, n-1$  is now represented twice - corner 1 is now 1 and -5 on the number line, corner 2 is now 2 and -4, and so on. The walk on the number line continues until the *range* of the walk becomes 6. This is equivalent to a particle reaching all corners on a polygon of  $n = 6$ . The general case for this transformation will be used to prove the expected time until absorption.

Consider a classical symmetric walk that starts at the point 0, with absorbing barriers  $a$  and  $b$ . The average time until one of these points is reached is equal to  $ab$ . Although this result is not proved here, we will use it to prove the expectation of  $T_i$  - the amount of time it takes for the number of points visited to increase to a given number  $i$ . Consider a particle's path up to  $t = 8$  is

$$0 \rightarrow 1 \rightarrow 0 \rightarrow 1 \rightarrow 2 \rightarrow 1 \rightarrow 0 \rightarrow -1. \quad (4)$$

Here,  $T_1 = 0$ ,  $T_2 = 1$ ,  $T_3 = 5$ , and  $T_4 = 8$ . The range of this walk at  $t = 8$  is 4.

Suppose that, at time  $T_i$ , the range is  $[-i+k+1, k]$ , and the particle is positioned at the rightmost point  $k$ . At time  $T_{i+1}$ , the particle is just outside this range, at either  $-i+k$  or  $k+1$ . The distance to these points are  $i$  and  $1$ , respectively, making the expected time to reach either of these points  $1 \cdot i$ , by the classical symmetric walk quoted earlier. Thus we have proved

$$E(T_{i+1}) - E(T_i) = i \quad (5)$$

By this recursive definition, we get

$$E(T_i) = 1 + 2 + \dots + (i-1) = \frac{1}{2}i(i-1) \quad (6)$$

Remembering our goal to find the cover time of a polygonal graph, we get the result

$$m_n = E(T_n) = \frac{n(n-1)}{2} \quad (7)$$

This result holds for all starting positions of the particle. Only the number of vertices  $n$  is relevant.

Finding the extended cover time  $u_n$  is more complicated. We want to find an expression for the absorption time, and add this to the cover time we just determined. Thus we get the sum

$$u_n = m_n + c_n, \quad (8)$$

where

$$c_n = \sum_{j=1}^{n-1} c_{nj} p_j \quad (9)$$

Here,  $p_j$  is the probability that, at  $m_n$ , the particle is at point  $j$ , and  $c_{nj}$  is the expected absorption time. For all vertices  $j$ , regardless of their distance from the starting vertex,  $p_j = 1/(n-1)$ . It may come as a surprise that the particle is not less likely to end up at a vertex close to the starting point than one further away, but it is indeed true.

We consider different alternatives for the range at time  $T_i$ . We can denote  $R_i$  the range of the particle at time  $T_i$ . There are always  $i$  alternative ranges for  $R_i$ , which are as follows. Clearly, at time  $T_i$ , the particle is at either end point of the range.

$$A_0 = [-i+1, 0]; A_1 = [-i+2, 1]; \dots; A_{i-1} = [0, i-1] \quad (10)$$

We now wish to prove that these alternatives occur with equal probability  $1/i$ . To prove this, we consider the resting point of the particle at  $T_i$ . For each alternative, the particle could be resting at either end point. We can denote these possibilities as pairs of events  $(A_0^+, A_0^-), \dots, (A_{i-1}^+, A_{i-1}^-)$ , where  $A_0^+$  indicates that the particle is at the right end point of the range  $A_0$  at  $T_i$ , and  $A_0^-$  indicates that it is on the left end point.

Consider the general case of these pairings, a pair  $(A_k^+, A_k^-)$ , where the range is defined as

$$A_k = [-i+k+1, k] \quad (11)$$

Clearly, in order for event  $A_k^+$  to occur, the particle must visit point  $-i+k+1$  before point  $k$ . This occurs, thanks to a proof not included in this paper, with probability

$$\frac{k}{k + (i - k - 1)} = \frac{k}{i - 1} \quad (12)$$

Additionally, the particle must visit point  $k$  before point  $-i + k$  in order for event  $A_k^+$  to occur. This occurs with probability

$$\frac{1}{k + (i - k)} = \frac{1}{i} \quad (13)$$

Thus, by multiplying these probabilities together, we get

$$P(A_k^+) = \frac{k}{i - 1} \cdot \frac{1}{i}. \quad (14)$$

By a similar proof on the left side of the range, we get

$$P(A_k^-) = \frac{i - k - 1}{i - 1} \cdot \frac{1}{i}. \quad (15)$$

If we add these two equations together, we get the overall probability of  $A_k$ , namely

$$P(A_k) = \frac{1}{i}, \quad (16)$$

Since there are  $n - 1$  possible values for  $i$  in a polygonal graph with  $n$  edges, we have proven that  $p_j = 1/(n - 1)$ .

Now we wish to determine the expected amount of time for a particle to return from point  $j$  to the starting vertex, which we denoted  $c_{nj}$  earlier. This is again equivalent to a random walk on the x-axis from point  $j$  and stopping at a point  $j$  steps to the right or  $n - j$  steps to the left. The expected time until absorption in this case is  $c_{nj} = j(n - j)$ . Thus we have

$$c_n = \sum_{j=1}^{n-1} j(n - j) \cdot \frac{1}{n - 1} = \frac{n(n + 1)}{6} \quad (17)$$

If we add (7) and (17) into (8), we get

$$u_n = \frac{n(n - 1)}{2} + \frac{n(n + 1)}{6} = \frac{n(2n - 1)}{3}, \quad (18)$$

which is the correct extended cover time for an  $n$ -sided polygonal graph. [1]



## 1.4 Complete Graphs

A type of graph that presents a simple cover time computation is the *complete graph*, in which each vertex is connected to every other vertex in the graph by a single edge. That is, a particle on the graph can reach any vertex from any other vertex in just one move. We denote the cover time of a complete graph  $m_n$ , as the starting vertex does not matter due to symmetry.

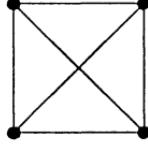


Figure 4: A complete graph with four vertices.

To derive the cover time of a complete graph of size  $n$ , we use a simple trick. We can write

$$m_n = E[Y_1] + E[Y_2] + \dots + E[Y_n], \quad (19)$$

where  $E[Y_j]$  is the expected amount of time required for the number of visited vertices to go from  $j - 1$  to  $j$ . The amount of moves required to reach one more vertex has the geometric probability function

$$P(Y_j = k) = (1 - p_j)^{k-1} p_j, \quad k = 1, 2, \dots \quad (20)$$

For  $k = 1, 2, \dots$ , with  $(1 - p_j)^{k-1}$  representing the probability that the particle will remain in the set of visited vertices for  $k - 1$  moves, and  $p_j$  the probability of the particle traveling to a new vertex on move  $k$ . At time  $k - 1$ , the particle has visited  $(j - 1)$  vertices, out of  $n$ , thus  $p_j = [n - (j - 1)]/n$ . Using the mean of a geometric distribution, we get

$$E[Y_j] = \frac{1}{p_j} = \frac{n}{n - j + 1}, \quad (21)$$

and so

$$m_n = (n - 1) \left( 1 + \frac{1}{2} + \dots + \frac{1}{n - 1} \right). \quad (22)$$

Using this equation, we can solve for the cover time of the complete graph in figure 4, and we get  $m_4 = 33/6$ . [1]

## 1.5 Cover Time Distribution of a Complete Graph

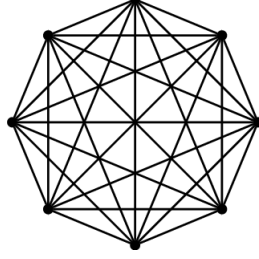


Figure 5: A Complete Graph with 8 vertices

What if we are interested in the *distribution* of this cover time variable, which we define as  $\tau_{cov}$ . Consider a complete graph of size  $n$  with *self-edges* - a particle on a given step can stay at it's current vertex. An example of such a graph is a man randomly dropping  $r$  balls into  $n$  boxes - each box is a vertex and each ball is an edge connected the previous box to the next one. Clearly, each of the  $n^r$  possible arrangements occurs with equal probability. If we consider  $N_n$  the number of empty boxes, and  $A_i$  the event that box  $i$  is empty, we clearly have

$$P(A_i) = (1 - 1/n)^r \quad \text{and} \quad E[N_n] = n(1 - 1/n)^r. \quad (23)$$

We can show that if  $rn \rightarrow c$ ,  $E[N_n] \rightarrow e^{-c}$  using logarithms as follows

$$E[N_n]/n = (1 - 1/n)^r = [(1 - 1/n)^n]^{r/n} \rightarrow (e^{-1})^c = e^{-c} \quad (24)$$

In order to compute the variance of  $N_n$ , we need to find both  $E[N_n]^2$  and  $E[N_n^2]$ . We define  $\mathbb{1}_i$  to be the indicator random variable of  $A_i$ , so  $\mathbb{1}_i = 1$  if box  $i$  is empty and 0 otherwise. Starting with  $E[N_n]^2$ , we have

$$E[N_n]^2 = E\left[\left(\sum_{i=1}^n \mathbb{1}_i\right)\left(\sum_{j=1}^n \mathbb{1}_j\right)\right] = \sum_{1 \leq i, j \leq n} E[\mathbb{1}_i]E[\mathbb{1}_j] = \sum_{1 \leq i, j \leq n} P(A_i)P(A_j) \quad (25)$$

Next we find  $E[N_n^2]$

$$E[N_n^2] = E\left[\left(\sum_{i=1}^n \mathbb{1}_i\right)^2\right] = E\left[\left(\sum_{i=1}^n \mathbb{1}_i\right)\left(\sum_{j=1}^n \mathbb{1}_j\right)\right] = \sum_{1 \leq i, j \leq n} E[\mathbb{1}_i \mathbb{1}_j] = \sum_{1 \leq i, j \leq n} P(A_i \cap A_j). \quad (26)$$

Using these to compute the variance of  $N_n$ , we get

$$\begin{aligned}
\text{var}(N_n) &= E[N_n^2] - E[N_n]^2 = \sum_{1 \leq i, j \leq n} P(A_i \cap A_j) - P(A_i)P(A_j) \\
&= n(n-1)\{P(A_i \cap A_j) - P(A_i)P(A_j)\} + n\{P(A_i \cap A_j) - P(A_i)P(A_j)\} \\
&= n(n-1)\{(1-2/n)^r - (1-1/n)^{2r}\} + n\{(1-2/n)^r - (1-1/n)^{2r}\}.
\end{aligned}$$

In the first term,  $k \neq m$ , and in the second term,  $k = m$ . We know, by logarithms, that  $(1-2/n)^r \rightarrow e^{-2c}$  and  $(1-1/n)^r \rightarrow e^{-c}$ . We can use this to determine, as  $n \rightarrow \infty$ ,

$$\begin{aligned}
\text{Var}(N_n/n) &= \text{Var}(N_n)/n^2 \\
&= (1-2/n)^r(1-1/n)^{2r} + o(n) \\
&\rightarrow e^{-2c} - (e^{-c})^2 = 0.
\end{aligned}$$

The second equality holds because, as  $n \rightarrow \infty$ , the second part of  $\text{Var}[N_n]$  becomes obsolete (we have denoted it  $o(n)$ ). To conclude that  $N_n/n \rightarrow e^{-c}$ , we use Chebyshev's Inequality

$$P(|N_n/n - E[N_n/n]| \geq \epsilon) \leq \text{Var}(N_n/n)/\epsilon^2. \quad (27)$$

We use this result to get

$$N_n/n - E[N_n/n] \xrightarrow{p} 0 \quad (28)$$

Here,  $\xrightarrow{p}$  denotes convergence *in probability*. We say that  $Y_n$  converges to  $Y$  in probability if  $\forall \epsilon > 0$ ,  $P(|Y_n - Y| > \epsilon) \rightarrow 0$  as  $n \rightarrow \infty$ .

Plugging in our solved value of  $E[N_n]$ , we get

$$N_n/n - (1-1/n)^r \rightarrow 0 \quad \text{so} \quad N_n/n \rightarrow e^{-c} \quad (29)$$

Thus, the fraction of empty boxes approaches  $e^{-c}$ . Based on the Poisson approximation to the binomial, we know that if  $n \rightarrow \infty$  and  $r/n \rightarrow c$ , that the number of balls in any given box approaches a Poisson distribution with mean  $c$ . Considering these results, we will now show, generalizing again to edges and vertices,

**Theorem 1.1.** *If  $ne^{r/n} \rightarrow \lambda \in [0, \infty)$ , the number of unvisited vertices approaches a Poisson distribution with mean  $\lambda$ .*

*Proof.* We start with the observation that  $k$  unvisited vertices occurs with probability

$$P(\text{vertices } i_1, i_2, \dots, i_k \text{ are unvisited}) = (1 - k/n)^r. \quad (30)$$

Now we denote the probability that exactly  $m$  out of  $n$  vertices are empty after taken

$r$  steps  $p_m(r, n)$ . Clearly  $P(\text{no unvisited vertices}) = 1 - P(\text{at least one unvisited vertex})$ , so by inclusion-exclusion

$$p_0(r, n) = \binom{n}{0} 1^r - \binom{n}{1} (1 - 1/n)^r + \binom{n}{2} (1 - 2/n)^r - \dots \quad (31)$$

$$= \sum_{k=0}^n (-1)^k \binom{n}{k} (1 - k/n)^r. \quad (32)$$

Using the possibilities for different locations of unvisited vertices, we get a more sophisticated version of (30)

$$p_m(r, n) = \binom{n}{m} \left(1 - \frac{m}{n}\right)^r p_0(r, n - m) \quad (33)$$

To prove that the limit of  $p_m(r, n)$  is a Poisson distribution with mean  $\lambda$ , we first need to prove that if  $ne^{-r/n} \rightarrow \lambda$ , then

$$\binom{n}{m} \left(1 - \frac{m}{n}\right)^r \rightarrow \frac{\lambda^m}{m!} \quad (34)$$

We can prove this using the squeeze theorem. We start with the upper bound. Since  $(1 - x) \leq e^{-x}$  and  $ne^{-r/n} \rightarrow \lambda$ , we have the inequality

$$\binom{n}{m} \left(1 - \frac{m}{n}\right)^r \leq \left(\frac{n^m}{m!}\right) e^{-mr/n} \rightarrow \lambda^m/m! \quad (35)$$

Now for the lower bound, which proves to be more difficult. We notice  $\binom{n}{m} \geq (n - m)^m/m!$ , and we use this to get

$$\begin{aligned} \binom{n}{m} \left(1 - \frac{m}{n}\right)^r &\geq \left(1 - \frac{m}{n}\right)^r \frac{(n - m)^m}{m!} \left(\frac{n^m}{n^m}\right) \\ &\geq \left(1 - \frac{m}{n}\right)^{m+r} \left(\frac{n^m}{m!}\right) \end{aligned}$$

We can use a couple tricks to simplify this. We know that, for a fixed  $m$ ,  $(1 - m/n)^m \rightarrow 1$  as  $n \rightarrow \infty$ , and that  $1/m!$  is a constant. To bound the rest, we use a Taylor Series, noting that if  $0 \leq t \leq 1/2$  then

$$\begin{aligned} \log(1 - t) &= -t - t^2/2 - t^3/3 - \dots \\ &\geq -t - \frac{t^2}{2} \left(1 + 2^{-1} + 2^{-2} + \dots\right) = -t - t^2 \end{aligned}$$

using this, we have

$$\begin{aligned}\log \left( n^m \left( 1 - \frac{m}{n} \right)^r \right) &= m \log n + r \log \left( 1 - \frac{m}{n} \right) \\ &\geq m \log n - r \left( \frac{m}{n} \right) - r \left( \frac{m}{n} \right)^2\end{aligned}$$

In order to solve for  $r$ , we use our assumption  $ne^{-r/n} \rightarrow \lambda$ ,

$$\begin{aligned}e^{\log n - r/n} &\rightarrow e^{\log \lambda} \\ \log n - r/n &= \log \lambda + o(n) \\ r &= n \log n - \log \lambda + n \cdot o(n)\end{aligned}$$

from which we get  $r(m/n)^2 \rightarrow 0$ . Now we can get

$$\begin{aligned}\frac{rm}{n} &= m \log n - m \log \lambda + m \cdot o(n) \\ m \log n - \frac{rm}{n} &\rightarrow m \log \lambda\end{aligned}$$

If we combine our last two results, we can show

$$\log \left( n^m \left( 1 - \frac{m}{n} \right)^r \right) \geq m \log n - \frac{rm}{n} - \frac{rm}{n^2} \rightarrow \log \lambda^m \quad (36)$$

From here, it follows that

$$\lim_{n \rightarrow \infty} n^m \left( 1 - \frac{m}{n} \right)^r \geq \lambda^m \quad (37)$$

Now we have proven (34) via the squeeze theorem. Now we use (32) and (34) to show that if  $ne^{-r/n} \rightarrow \infty$ , we have

$$p_0(r, n) \rightarrow \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \quad (38)$$

Furthermore, we notice for a fixed  $m$ ,  $(n-m)e^{-r/(n-m)} \rightarrow \lambda$ , so combining this with (33) and (34), we get

$$p_m(r, n) = \frac{\lambda^m}{m!} e^{-\lambda} \quad (39)$$

which is a Poisson distribution. So  $N_n$  follows a Poisson distribution with mean  $\lambda$ , as claimed. □

We can use Theorem 1.1 to learn about our cover time variable  $m_n$  as well. If we set  $r = n \log n + nx$ , we get

$$ne^{-r/n} = ne^{-\log n - x} = \frac{ne^{-x}}{n} = e^{-x} \quad (40)$$

Let's take, for example, a random walk of length  $r$  on a complete graph with  $n$  vertices. Clearly,  $m_n \leq r$  if and only if the particle covers all  $n$  vertices in  $r$  steps, and  $ne^{-r/n} \rightarrow e^{-x}$ , we have

$$P(m_n - n \log n \leq nx) \rightarrow e^{-e^{-x}}. \quad (41)$$

**Example 1.2.** *What is the likelihood that, in a village of 2190 people, all possible birthdays are represented? What about a village of 1825?*

Consider this problem a cover time for a complete graph with 365 vertices, with each step representing a person in the group. With  $n = 365$ , we have  $n \log n = 2153$ . Plugging this into (41), we get

$$\begin{aligned} P(m_{365} \leq 2190) &= P\left(\frac{m_{365} - 2153}{365} \leq \frac{37}{365}\right) \\ &\approx e^{-e^{-0.1014}} = e^{-0.9036} = 0.4051 \\ P(m_{365} \leq 1825) &= P\left(\frac{m_{365} - 2153}{365} \leq \frac{-328}{365}\right) \\ &\approx e^{-e^{-0.8986}} = e^{-2.4562} = 0.085 \end{aligned}$$

So all birthdays being represented in villages of size 2190 and 1825 occurs with a 40% and 8% chance, respectively.

The analysis and results in this section follow from [2].

## 2 Pólya's Recurrence Theorem

In the early 1900s, Hungarian mathematician George Pólya was taking a walk in the park when he noticed he was passing the same couple often, even though it seemed as though they were both taking random paths through the park. Inspired, Pólya investigated the same problem on infinite graphs, eventually presenting his famous theorem in 1921, stating that a random walk on a  $d$ -dimensional lattice is recurrent for  $d = 1, 2$ , and transient for  $d > 2$ . We have introduced 1-dimensional random walks in section 1, but we will explore this further in famous problems Cliff Hanger. We will then look at the 2-dimensional and 3-dimensional cases in the Drunk Man and the Drunk Bird, respectively. The proofs of these problems follow from [4].

## 2.1 One-Dimensional Random Walk: The Cliff-Hanger

Suppose a man stands on the edge of a cliff - one step forward sends him over the edge. He takes random steps either toward or away from the cliff. At each given time, his chance of taking a step backwards is  $\frac{2}{3}$ , and his chance of taking a step forward is  $\frac{1}{3}$ . Will the man escape the cliff?

We can view this problem as a particle walking along the x-axis, starting at  $x = 1$ . The particle visiting the origin is equivalent to the man stepping off the ledge. Clearly, the particle can only reach 0 on an odd-numbered step, so we only consider these cases when computing the probability. At  $t = 1$ , the particle has a  $\frac{1}{3}$  of being at 0. At  $t = 3$ , there is one possible way of reaching 0: one step to the right and two steps to the left. This path occurs with probability  $\frac{2}{27}$ . At  $t = 5$ , the particle has two possible paths of reaching -1, namely

$$1 \rightarrow 2 \rightarrow 1 \rightarrow 2 \rightarrow 1 \rightarrow 0 \text{ and } 1 \rightarrow 2 \rightarrow 1 \rightarrow 2 \rightarrow 1 \rightarrow 0 \quad (42)$$

each with probability  $\frac{4}{243}$ . Together, the particle has probability  $\frac{107}{243}$  of reaching the origin at some point at or before time  $t = 5$ . We can continue adding up these possibilities for  $t \rightarrow \infty$ , but this is tedious - here is a simpler solution.

Let us look at the general case - the probability of a step to the right is equal to  $p$ , not  $\frac{2}{3}$ . Starting at index  $x = n$ , the particle moves to either  $x = n + 1$  with probability  $p$  or to  $x = n - 1$  with probability  $(1 - p)$ . We call  $x = -1$  an *absorbing barrier* - if the particle arrives there, it takes no further steps. We are interested in finding  $P_1$  - the probability that the particle is *absorbed* at  $x = 0$ , given that it starts at  $x = 1$ . At  $t = 1$ , the particle was absorbed with probability  $(1 - p)$ , and is at  $x = 2$  with probability  $p$ . So we have

$$P_1 = (1 - p) + pP_2 \quad (43)$$

Now let's look at  $P_2$ . Paths to absorption from  $x = 2$  can be divided into two parts: the path from  $2 \rightarrow 1$ , and the path from  $1 \rightarrow 0$ . The first part is just  $P_1$ , because this problem is identical to the  $1 \rightarrow 0$  case, just translated one step to the right. The second part is also  $P_1$ , because this is *exactly* the original problem. So we then have  $P_2 = P_1^2$ , because these two parts are independent. So we can rewrite (43) as

$$P_1 = (1 - p) + pP_1^2 \quad (44)$$

Solving this quadratic for  $P_1$ , we get solutions

$$P_1 = 1, \quad P_1 = \frac{1 - p}{p} \quad (45)$$

Now we must determine when each of these solutions is appropriate. Let's look at a

couple values of  $p$  and think about what we'd expect  $P_1$  to be. When  $p = 0$ ,  $P_1$  is obviously 1, and when  $p = 1$ ,  $P_1 = 0$ , because the particle is always moving to the right. When  $p = 0$ ,  $P_1 = 1$ , and both solutions agree. When  $p < \frac{1}{2}$ , the second solution doesn't work because  $(1 - p)/p > 1$ , and we know  $P_1 \leq 1$ . So we use the first equation  $P_1 = 1$  for  $0 \leq p \leq \frac{1}{2}$ . We can connect these two functions at  $p = \frac{1}{2}$  and get the following graph

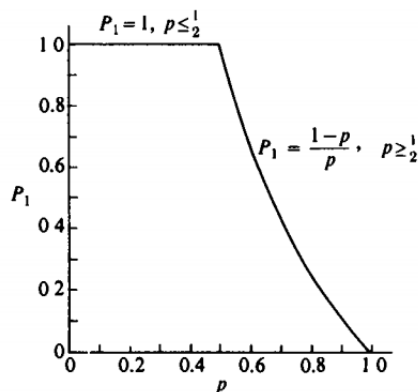


Figure 6: Probabilities of absorption,  $P_1$ , of the Cliff-Hanger [4]

A more thorough proof of this problem would include proving that the second solution,  $P_1 = (1 - p)/p$ , is a continuous as a function of  $p$ . This is a result we will assume but not prove here. To answer our original inquiry, where  $p = \frac{2}{3}$ , we use the second solution, as  $p > \frac{1}{2}$ . We get  $P_1 = \frac{1}{2}$ , which means the Cliff-Hanger's fall and escape from the cliff edge are equally likely. [4]

## 2.2 Stirling's Approximation Proof

As we progress into the 2-dimensional and 3-dimensional components of Pólya's Theorem, we see Stirling's Approximation used to deal with very large values of  $n$ . A very important formula in probability as well as in applied mathematics, it is used to approximate  $n!$  as  $n \rightarrow \infty$ . As I reference it in every section after this one, I saw fit to include its proof here. The work done in this proof closely follows from [3].

We start by taking the log of  $n!$

$$\log(n!) = \log(1) + \log(2) + \dots + \log(n) \quad (46)$$

As the log function monotonically increases on the interval  $(0, \infty)$ , we know



$$\int_{n-1}^n \log(x) dx < \log(n) < \int_n^{n+1} \log(x) dx \quad \text{for all } n > 1 \quad (47)$$

If we take the above inequalities for  $n = 1, 2, \dots, N$ , and add them all together, we get

$$\int_0^N \log(x) dx < \log(N!) < \int_1^{N+1} \log(x) dx \quad (48)$$

This first integral is improper, but convergent. Since the antiderivative of  $\log(x)$  is equal to  $x \log(x) - x$ , we get

$$n \log(n) - n < \log(n!) < (n+1) \log(n+1) - n \quad (49)$$

Now we create a sequence  $d_n$  such that

$$d_n = \log(n!) - \left(n + \frac{1}{2}\right) \log(n) + n \quad (50)$$

Subtracting the  $n+1$  term from the  $n$ th term, we get

$$d_n - d_{n+1} = \left(n + \frac{1}{2}\right) \log\left(\frac{n+1}{n}\right) - 1 \quad (51)$$

Manipulating the algebra gives

$$\frac{n+1}{n} = \frac{1 + \frac{1}{2n+1}}{1 - \frac{1}{2n+1}} \quad (52)$$

Now we implement the Taylor expansion to get

$$\frac{1}{2} \log\left(\frac{1+t}{1-t}\right) = t + \frac{1}{3}t^3 + \frac{1}{5}t^5 + \dots \quad (53)$$

so for  $-1 < t < 1$ , we get

$$d_n - d_{n+1} = \frac{1}{3} \frac{1}{(2n+1)^2} + \frac{1}{5} \frac{1}{(2n+1)^4} + \dots \quad (54)$$

which means

$$0 < d_n - d_{n+1} < \frac{1}{3} \left( \frac{1}{(2n+1)^2} + \frac{1}{(2n+1)^4} + \dots \right) \quad (55)$$

which is a geometric series, so we can solve it to get

$$0 < d_n - d_{n+1} < \frac{1}{3} \frac{1}{(2n+1)^2 - 1} = \frac{1}{12} \left( \frac{1}{n} - \frac{1}{n+1} \right) \quad (56)$$

From this we have determined two things:

1. the sequence  $d_n$  is decreasing
2. the sequence  $d_n - \frac{1}{12n}$  is increasing

Since both series converge, we can infer

$$\lim_{x \rightarrow \infty} d_n = \lim_{x \rightarrow \infty} d_n - \frac{1}{12n} = C \quad \text{for some constant } C \quad (57)$$

We know  $C > d_1 - 1/12 = 1 - 1/12 = 11/12$ . Now we take the exponential of  $d_n$ , from which we get

$$\lim_{n \rightarrow \infty} \frac{n!}{n^{(n+1/2)} e^{-n}} = e^C \quad (58)$$

solving for  $n!$ , we get close to our desired approximation

$$n! \sim n^{(n+1/2)} e^{-n} e^C \quad (59)$$

Now to prove  $e^C = \sqrt{2\pi}$ , we use the Wallis formula (the proof of which will thankfully not be included here), which states

$$\lim_{n \rightarrow \infty} \frac{(2)(2)(4)(4)(6)(6) \cdot \dots \cdot (2n)(2n)}{(1)(1)(3)(3)(5)(5) \cdot \dots \cdot (2n-1)(2n-1)(2n+1)} = \frac{\pi}{2} \quad (60)$$

We can rewrite this to get

$$\frac{(2)(4)(6) \cdot \dots \cdot (2n)}{(1)(3)(5) \cdot \dots \cdot (2n-1)\sqrt{2n}} \sim \sqrt{\frac{\pi}{2}} \quad (61)$$

This can be simplified to

$$\frac{(2^n n!)^2}{(2n)!} \frac{1}{\sqrt{2n}} \sim \sqrt{\frac{\pi}{2}} \quad (62)$$

Plugging in our approximation for  $n!$  given in (59), we get

$$\frac{2^{2n} (n^{2n+1} e^{-2n} e^{2C})}{(2n)^{(2n+1/2)} e^{-2n} e^C} \frac{1}{\sqrt{2n}} \sim \sqrt{\frac{\pi}{2}} \quad (63)$$

From here, we can use very simple algebra to get

$$e^C \sim \sqrt{2\pi} \quad (64)$$

Finally, we can plug this value of  $e^C$  into (59) to get

$$n! \sim (\sqrt{2\pi})n^{n+1/2}e^{-n} \quad (65)$$

which is Stirling's Approximation! We will be using this result in future sections. [3]

## 2.3 Two-Dimensional Random Walk: The Drunk Man

A drunk man stumbles out of his house late one night. With no sense of direction, he travels northeast, northwest, southeast, and southwest with equal probability at every intersection. If the night is infinitely long, will the man become hopelessly lost, or will he eventually find his way back home?

We can again view this problem as a random walk on a graph. In the 1-dimensional version of the random walk presented in the Cliff Hanger, we proved that returning to origin is unity when steps to the left and to the right are equally likely, but this was a special case. If the probability moves (even slightly) from  $\frac{1}{2}$ , this result no longer holds. Considering the set of roads and intersections to be an infinite lattice, and the drunk man a particle starting at the origin, it would seem that wandering off in two dimensions would be likely. To investigate this, we will compute the average number of times the particle returns, and deduce its probability of absorption from this. As we have in previous problems, we will denote a particle's probability of return  $P$ , and the probability of no return  $Q = 1 - P$ . The probability of exactly  $x$  returns is simply  $P^x Q$ , as the walk just restarts every time the particle returns. We could use a geometric series to solve for  $\mu$ , namely

$$\mu = \sum_{x=0}^{\infty} x P^x Q, \quad (66)$$

which would give us our result, except we're still missing the values of  $P$  and  $Q$ . So we are tasked with determining these values.

We know that, in a geometric distribution, the mean number of trials until the first success is the reciprocal of the probability of success. In this example, the particle not returning to the origin is considered to be the 'success', as it terminates the series. So the mean number of trials until a non-return is  $1/Q$ , and the mean number of successes is  $1/Q - 1$ . If we set  $Q = 1$ , the mean number of successes becomes 0, and as  $Q \rightarrow 0$ , mean successes approaches infinity. Furthermore, there exists a bijection between values

of  $\mu$  and values of  $Q$  - for every mean there exists a  $Q$  and vice versa. So we just need to compute one to get the other, and we have chosen to compute  $\mu$ .

We can consider the direction chosen at each vertex a product of two binary decisions. That is, an evenly-weighted coin is flipped to determine the north-south component of the walk, and flipped again to determine the east-west component. If we think about the particle in this way, we can consider the two dimensional walk the product of two independent one dimensional walks, which we have seen before.

Clearly, a return to the origin can only occur on an even-numbered step, so we focus on walks of even length. After two steps, the horizontal component of the walk has the distribution

$$P(X_2 = -2) = \frac{1}{4}, \quad P(X_2 = 0) = \frac{2}{4}, \quad P(X_2 = 2) = \frac{1}{4}. \quad (67)$$

The vertical component of the walk has an identical distribution, and their joint PDF is

		$P(x, y)$			<b>Marginal for <math>Y</math></b>	
					$y$	$P(y)$
		$\bullet \frac{1}{16}$	$\bullet \frac{2}{16}$	$\bullet \frac{1}{16}$	2	$\frac{1}{4}$
		$\bullet \frac{2}{16}$	$\bullet \frac{4}{16}$	$\bullet \frac{2}{16}$	0	$\frac{2}{4}$
		$\bullet \frac{1}{16}$	$\bullet \frac{2}{16}$	$\bullet \frac{1}{16}$	-2	$\frac{1}{4}$
<b>Marginal for <math>X</math></b>	$x$	-2	0	2		
	$P(x)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$		

Figure 7: Joint distribution of  $X$  and  $Y$  after two moves [4]

We determine the probability that the particle rests at the origin after two moves to be  $\frac{4}{16}$  by multiplying  $P(X = 0)$  together with  $P(Y = 0)$ , which is possible because the two events are independent of one another. So far, our contribution to the mean is  $1 \cdot (\frac{4}{16}) + 0 \cdot (\frac{12}{16}) = (\frac{4}{16})$ . We can continue this computation for  $n = 2, 4, 6, \dots$  to get our final value for  $\mu$ , but it gets trickier from here on out.

Firstly, for all 1-dimensional even-length walks greater than 2, we have multiple ways of returning to the origin. For a walk of length 4, for example, we have

$$0 \rightarrow 1 \rightarrow 0 \rightarrow 1 \rightarrow 0 \quad \text{and} \quad 0 \rightarrow 1 \rightarrow 2 \rightarrow 1 \rightarrow 0 \quad (68)$$

which both lead back to the vertex. We can count the number of returning walks of length  $2n$  using binomial probabilities, and we get the overall equation

$$P(\text{return at step } 2n) = P(X_{2n} = 0) \cdot P(Y_{2n} = 0) = \left[ \binom{2n}{n} \left( \frac{1}{2} \right)^{2n} \right]^2. \quad (69)$$

In order to determine this value, we apply Stirling's approximation (65)

$$\binom{2n}{n} \left( \frac{1}{2} \right)^{2n} = \frac{(2n)!}{n!n!} \left( \frac{1}{2} \right)^{2n} \approx \frac{(\sqrt{2\pi})(2n)^{2n+\frac{1}{2}}e^{-2n}}{((\sqrt{2\pi})n^{n+\frac{1}{2}}e^{-n})^2 2^{2n}} \approx \frac{1}{\sqrt{\pi n}} \quad (70)$$

Squaring this result, we get

$$P(X_{2n} = 0) \cdot P(Y_{2n} = 0) \approx \frac{1}{\pi n} \quad (71)$$

Now that we have the approximate probability that a walk will return to the origin at step  $2n$  for all values of  $n$ , we need to calculate the sum of all of these values. We get

$$\sum_{n=1}^{\infty} \frac{1}{\pi n} = \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \quad (72)$$

Since  $\sum_{n=1}^{\infty} \frac{1}{n}$  diverges as  $n$  grows, so does the probability that the particle returns at some even step. If we recall our goal, this is also the mean number of times the particle has returned to the origin after exactly  $2n$  steps. Since this mean is unbounded, we know  $P$ , the probability that the particle will return, is equal to 1. Not only does the particle return, but it returns *infinitely often*. Going back to our original problem, the drunk man will find his way home again. Upon arriving home, he could go for another walk, and infinitely more, and still end up home. [4]

## 2.4 Three-Dimensional Random Walk: The Drunk Bird

*A drunk man will find his way home, but a drunk bird may get lost forever.*

Shizuo Kakutani

We have proven that a particle will return to the origin during both one-dimensional and two-dimensional random walks, so does it follow that it will return on a walk of  $n$  dimensions? We start with the 3-dimensional case. An extension of our work in the

previous section, we know that the chances of position of three independent 1-dimensional walks equalling 0 is

$$P(\text{return at step } 2n) = P(X_{2n} = 0) \cdot P(Y_{2n} = 0) \cdot P(Z_{2n} = 0) = \left[ \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \right]^3. \quad (73)$$

Using Stirling's Approximation, we get

$$P(\text{return at step } 2n) \approx \frac{1}{(\pi n)^{\frac{3}{2}}} \quad (74)$$

If we take the infinite series of these probabilities, we can use the identity  $\sum_{n=1}^{\infty} \frac{1}{n^s}$  diverges for  $s \leq 1$  and converges for  $s > 1$ . The three cases  $d = 1, 2, 3$  considered in this paper correspond to  $s = 1/2, 1, 3/2$ . Thus the series diverge, diverge, and converge, respectively. Hence we have the result that a random walk of a particle is recurrent for  $d = 1, 2$  (which we have proven in previous sections) and transient for  $d = 3$  (a new result). Let us find the exact value of the absorption probability for  $d = 3$ .

Referring back to (74), we can't use a series approximation for the sum of these values, but we can use an integral test to get a bound on the series. Visualize the sum  $\sum_{n=0}^{\infty} 1/n^{3/2}$  as a series of rectangles of length 1 (from  $n \rightarrow (n+1)$ ) and height  $1/(n^{3/2})$ . Using the curve  $f(x) = (x-1)^{-3/2}$ , we have an upper bound that passes through the upper right-hand corners of each of the rectangles. Because this function is an upper bound, the area under the curve is greater than the sum of the areas of the rectangles.

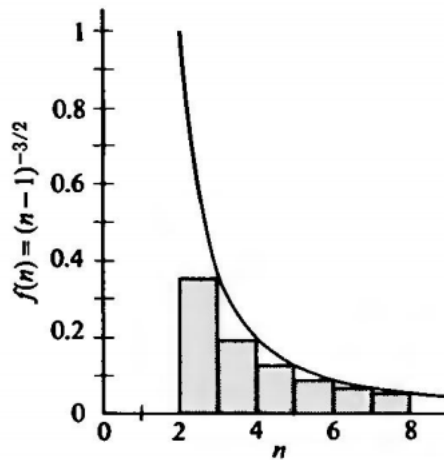


Figure 8: Upper bound for  $\sum_{n=0}^{\infty} \frac{1}{n^{3/2}}$  [4]

Taking the integral, we get

$$\int_n^N (n-1)^{-\frac{3}{2}} dx = -2(x-1)^{-\frac{1}{2}} \Big|_n^N = \frac{2}{(n-1)^{\frac{1}{2}}} - \frac{2}{(N-1)^{\frac{1}{2}}} \quad (75)$$

So as  $N \rightarrow \infty$ , this area approaches  $2/(n-1)^{\frac{1}{2}}$ , which is finite, making the sum of the rectangles finite as well. We can evaluate the number by taking the sum of the series for a small number of  $n$ , and then approximating remainder of the integral using Stirling's Approximation. If we choose to sum the first 18 terms, we get

$$\sum_{n=0}^{18} \left[ \binom{2n}{n} \left( \frac{1}{2} \right)^{2n} \right]^3 \approx 0.315, \quad (76)$$

and Stirling Approximation of the remainder of the integral is so small as to be considered negligible. Thus, 0.315 is approximately the mean number of returns to the origin of a given particle. Since  $1/Q = 1 + \mu$ , we solve for  $Q$  and get  $Q \approx 0.761$ . We then can calculate  $P$ , the probability of a particle returning, to be approximately 0.239.

So a particle is more likely to wander off forever in three dimensions than return to the origin. It should come as no surprise to the reader that a walk in additional dimensions yields an even more improbable return. Using the same summation strategy, we find the probability of returning to the origin in a 4-dimensional walk to be approximately 0.105. These probabilities continue to decrease as the number of dimensions increases. [4]

## References

- [1] Gunnar Blom, Lars Holst, and Dennis Sandell. *Problems and Snapshots from the World of Probability*. Springer Science & Business Media, 2012.
- [2] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.
- [3] Mohamed A Khamsi. Stirling's formula. Dec 1996.
- [4] Frederick Mosteller. *Fifty challenging problems in probability with solutions*. Courier Corporation, 1987.



## A Random Walk on a Chessboard: Source Code

```
public class Main {
    static int dims = 8; // dimensions of the chessboard (dims x dims)
    static int trials = 1000000; // trials

    public static void main(String [] args){
        Random r = new Random();

        /*
         * Random walk of a king
         */

        int [] kingTimes = new int [trials];
        for (int i = 0; i < kingTimes.length; i++) {
            Board b = new Board(dims);
            int randX = r.nextInt(dims);
            int randY = r.nextInt(dims);
            Square initSquare = b.getSquare(randX, randY);

            King myKing = new King(initSquare);
            initSquare.put(myKing);

            int numMoves = 0;
            while (!b.covered()) {

                LinkedList<Square> possibleMoves = myKing.getPossibleMoves(b);
                int ran = r.nextInt(possibleMoves.size());
                Square move = possibleMoves.get(ran);
                myKing.move(move);
                numMoves++;
            }
            kingTimes[i] = numMoves;
            System.out.println(numMoves);
        }

        /*
         * Random walk of a knight
         */
        int [] knightTimes = new int [kingTimes.length];
        for (int i = 0; i < knightTimes.length; i++) {
            Board b = new Board(dims);
            Square [][] array = b.getSquareArray();
            int randX = r.nextInt(dims);
            int randY = r.nextInt(dims);
```

```

    Square initSquare = b.getSquare(randX, randY);

    Knight myKnight = new Knight(initSquare);
    initSquare.put(myKnight);

    int numMoves = 0;
    while (!b.covered()) {
        LinkedList<Square> possibleMoves = myKnight.getPossibleMoves(b);
        int ran = r.nextInt(possibleMoves.size());
        Square move = possibleMoves.get(ran);
        myKnight.move(move);
        numMoves++;
    }
    knightTimes[i] = numMoves;
    System.out.println(numMoves);
}

int kingSum = 0;
int knightSum = 0;
for (int i = 0; i < kingTimes.length; i++){
    kingSum += kingTimes[i];
    knightSum += knightTimes[i];
}

double kingAverage = kingSum/kingTimes.length;
double knightAverage = knightSum/knightTimes.length;

System.out.println("Mean cover time of a king on a chessboard (" +
    kingTimes.length + " trials): " + kingAverage + " moves");
System.out.println("Mean cover time of a knight on a chessboard (" +
    knightTimes.length + " trials): " + knightAverage + " moves");

}
}

```