Introducing Markov Chains Through Tackling A Fun Chess Problem

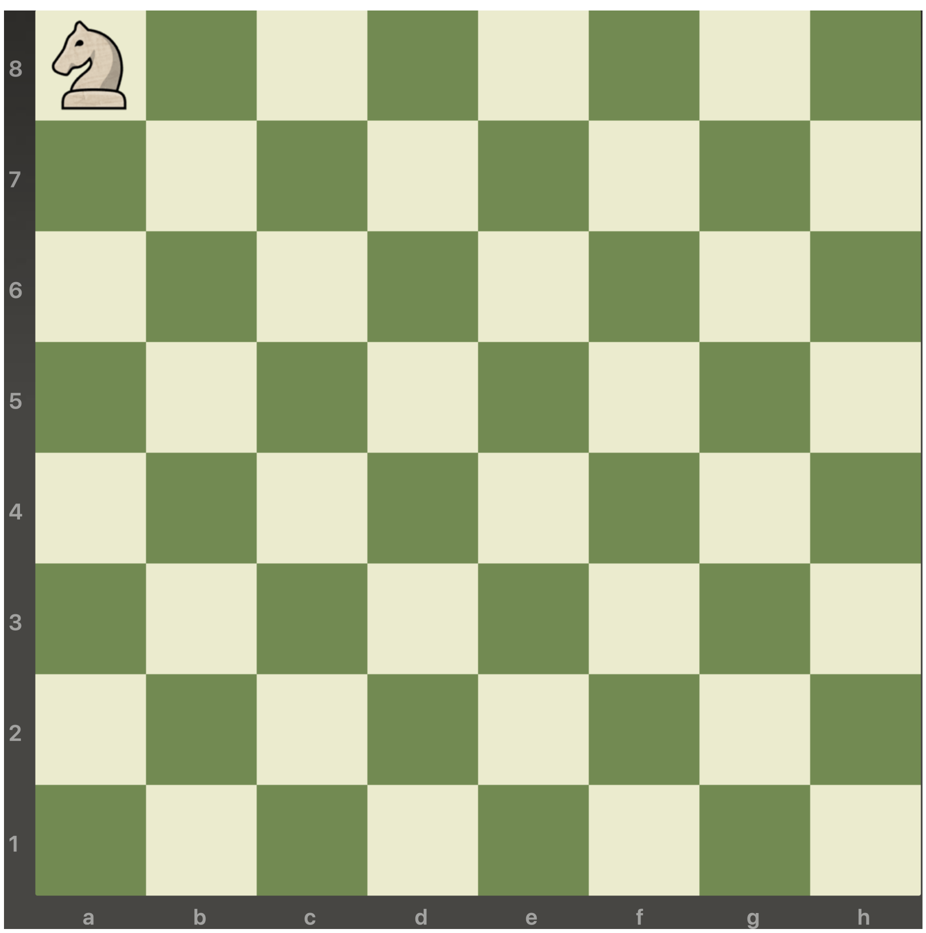
Victor Alvarez
Nick Brinkmann

June 4, 2021

1 Introduction

Many people have heard of a knight's tour in chess: where a knight starts on one square and moves around the board, visiting each square exactly once before returning to its original square. Finding knight's tours is quite a challenge! A lesser known problem involving knights is the motivating question of this paper.

The Problem. Suppose a knight starts on any square of a chessboard, and moves uniformly randomly to any square which constitutes a legal knight's move. The knight continues this process indefinitely. How many moves, on average, will it take the knight to return to its starting square? Does this expected return time depend on the starting square of the knight? Pause for a moment and think about how you might approach this problem!



It turns out that this problem is highly amenable to using *Markov chains*, an extremely useful and widespread method in probability to model stochastic (or random) processes. Markov chains allow us to model systems that change over time, where the 'state' of the system at any given time period depends *only* upon the state in the previous period. We consider only systems with discrete time intervals and finite state spaces, though the concepts of Markov chains can be extended to model continuous-time transitions. Additionally, systems where the state at a given time depends on *multiple* previous states can be studied by expanding the state space. Both of these extensions are beyond the scope of this project.

For some intuition on what Markov chains allow us to do, Section 4.9 of *Linear Algebra and its Applications* (Lay) introduces Markov chains as a way to model population dynamics of citizens moving between a hypothetical city and its suburbs, and to determine what the ratio of people living in the city is to people living in the suburbs over time. Of course, the applications can get much more complex. One very useful characteristic of Markov chains is that they rely on only a few core ideas, and thus can be easily adapted to a huge variety of problems.

In tackling our knight problem, we introduce the core ideas of Markov chains. In particular, we focus on what are known as *time-homogeneous* Markov chains, which just means that the probability of transitioning from one state of the system to another state does not change over time. We build up a number of helpful properties of Markov chains, applying them to our problem of the wandering knight as we go, with relevant sections being split into theory and application.

2 The Core of Markov Chains

2.1 Theory

There are a few core components that constitute a Markov chain. The first is some state space $S = \{S_1, S_2, ..., S_n\}$ where $n \in \mathbb{N}$. The state space just describes all of the possible outcomes which the system could be in at any point.
We then require *probability vectors.*

Definition 2.1.1. Probability vectors are vectors with real, non-negative entries such that the sum of the entries is 1.

The name 'probability vector' follows from the fact that each entry of the vector describes the probability of some event occurring. In particular, we require probability vectors $\mathbf{q_1}, \mathbf{q_2}, ..., \mathbf{q_n} \in \mathbb{R}^n$ for each state in the state space. These vectors are by convention row vectors where $(\mathbf{q_i})_j$, the j^{th} entry in the

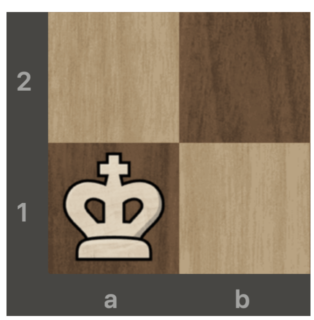
i^{th} probability vector is the probability of transitioning from state i to state j.[1] From these vectors, we construct a *stochastic matrix*.

Definition 2.1.2. A *stochastic matrix* is a square matrix whose rows are probability vectors. [2]

In particular, we construct the stochastic matrix $Q = \begin{bmatrix} \mathbf{q_1} \\ \mathbf{q_2} \\ \vdots \\ \mathbf{q_n} \end{bmatrix}$.[3] The reason that it makes sense that the rows sum to 1 is because, from some state i, the system has to move to some other state j; the system cannot move to a state outside of the state space. Furthermore, the system cannot simultaneously be in both state x and state y for $x \neq y$, so the probabilities must sum to exactly 1. (In probability terms, the events of being in state x and state y for $x \neq y$ are *disjoint*.)
In addition to the stochastic matrix, we use *state vectors* $\mathbf{x_0}, \mathbf{x_1}, \mathbf{x_2}, ... \in \mathbb{R}^n$ where $\mathbf{x_0}$ denotes the initial state of the system (which could be completely determined or itself random), and for all $i \geq 0$, $\mathbf{x_{i+1}} = \mathbf{x_i}Q$. Note here that the $\mathbf{x_i}$ are themselves probability vectors, and are also $1 \times n$ row vectors by convention, so that the matrix product $\mathbf{x_i}Q$ is defined (Recall Q is $n \times n$).

Remark. The equation $\mathbf{x_{i+1}} = \mathbf{x_i}Q$ is known as the *Markov Property*, and it says that at any time period $i + 1$, the state of the system depends only on the state of the system at time i. This assumption provides a useful halfway point between complete independence of events/random variables (which is a simplistic, often very restraining assumption), and complete dependence (which can be incredibly messy and complex to work with).

Example 2.1.3. To solidify these ideas, consider the following problem:



[1] If the $\mathbf{q_i}$ were instead column vectors, the $(\mathbf{q_i})_j$ entry would correspond to the probability of transitioning from state j to state i. Thanks to the teaching staff for pointing out this convention.

[2] If you used the alternative convention, the columns would instead be probability vectors. Other names for stochastic matrices are probability matrices, transition matrices, substitution matrices, or Markov matrices.

[3] We would call the matrix P, and its rows p_i, but for the fact that we want to reserve P for the conventional probability notation to prevent confusion.

A king starts on the a1 square of a 2×2 chessboard, and moves to another square uniformly randomly (i.e. with equal probabilities for all legal moves). Let $S = \{\text{a1, a2, b1, b2}\}$ be our (ordered) state space, corresponding to all the squares that the king can occupy. Here, $\mathbf{x_0} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$ denotes the initial state of the system, as the king occupies $a1$ with probability 1, and occupies a2, b1, and b2 with probability 0.[4]

We now define the 4×4 transition matrix

$$Q = \begin{bmatrix} \mathbf{q_{a1}} \\ \mathbf{q_{a2}} \\ \mathbf{q_{b1}} \\ \mathbf{q_{b2}} \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix}. \tag{1}$$

Recall from above that the i^{th} row of the transition matrix gives the probabilities of transitioning from state i to any other state. Thus, for example, the third row of Q has entries of $\frac{1}{3}$ for the a1, a2 and b2 squares, and 0 for the b1 square, as the king can move to a1, a2, and b2, but cannot move to the b1 square itself. Consequently, x_1, the state vector describing the probabilities of the king occupying any square after one move, is given by

$$\mathbf{x_1} = \mathbf{x_0}Q = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}. \tag{2}$$

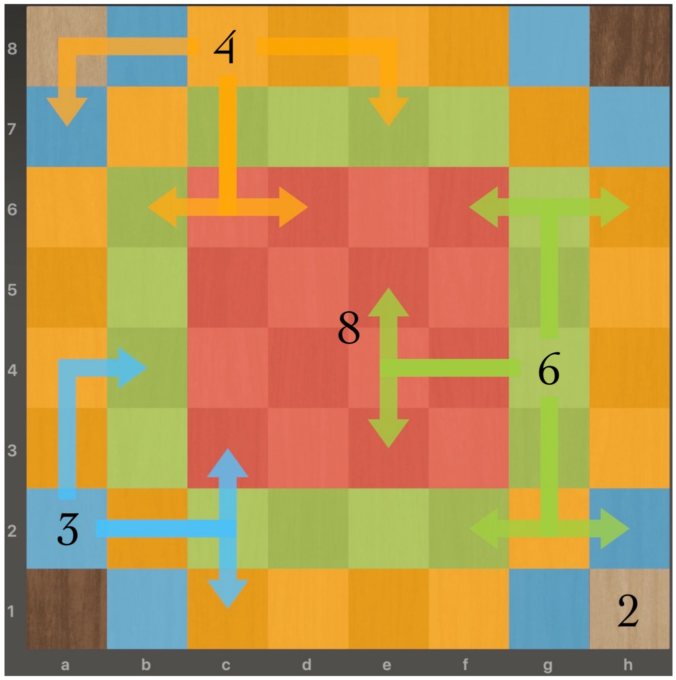
This makes sense, as it tells us that after 1 move, the king is equally likely to be in any square besides the a1 square.

Similarly, we can obtain $\mathbf{x_2}$ by right-multiplying $\mathbf{x_1}$ by Q, $\mathbf{x_3}$ by right-multiplying $\mathbf{x_2}$ by Q, and so on.

2.2 Application

What does this mean for our problem? Analogously to the King example, our state space consists of all the squares that the knight can be on, that is, the 64 squares of the chessboard: $S = \{\text{a1, a2, ..., a8, b1, ..., b8, ..., h1, ..., h8}\}$. Our transition matrix Q is a 64×64 matrix such that each row of Q consists of the transition probabilities from the (letter, number) square of the board. What are these probabilities?

[4] It is conventional to write $\mathbf{x_k} = a_1$ when the state of the system is known, rather than writing out the full vector of probabilities. This notation is more compact than the vector notation, and also more intuitive; consequently we will use it in later examples.



In the image above, the squares are colored according to how many available knight moves there are from that square. When a knight stands on a red square, it can move to 8 other squares; similarly for the other colors. Potential moves are shown for sample orange, green, and blue squares. How does this graphic help us? Well, it basically tells us what Q is! For each row of Q, corresponding to the transition probabilities away from the square denoted by that row, there will be 0 entries for all squares to which a knight cannot move, and $\frac{1}{k}$ entries for all squares to which a knight can move, where k is given by the number associated with the square's color in the image.

Thus, for example, the a2 square's row will have zero entries for all squares except the entries corresponding to b4, c1, and c3, which will each have a value of $\frac{1}{3}$, since a knight on a2 moves to these three squares with equal probability. Similarly, the c8 square's row will have zeros for all squares except a7, b6, d6, and e7, which will have values of $\frac{1}{4}$. Following this logic, we can construct the entire matrix Q, which, for your sake and ours, is not included explicitly.

For any given starting distribution of the knight, we can then find the probabilities of where it will be after one or more moves.

Of course, when the starting square is specified definitely (say, the knight starts on d4), and the number of moves is small, you can just calculate this by hand. But when the starting distribution is more complex (say, starts on all squares with equal probability) or the number of moves is larger than a handful, reason alone becomes impractical.

3 N-Step Transition Probabilities

3.1 Theory

The matrix Q, described in the previous section, contains the 1-step transition probabilities in each entry. That is, each entry q_{ij} represents the probability of transitioning from state i directly to state j.

Definition 3.1.1. The *n-step transition probability*, denoted as $p_{ij}^{(n)}$ is the probability that, starting at state i, we will end at state j after n transitions.

To get to state j from state i will require that we reach $n-1$ intermediate states on the way.

Theorem 3.1.2. *Given that the matrix Q contains the 1-step transition probabilities, then Q^n will contain the n-step transition probabilities.*

Proof. [5]
We prove by mathematical induction on the number of transition steps.
Base Case. We are given matrix Q with entries q_{ij} being the probabilities of transitioning from state i to state j. Now, let's consider Q^2 given by:

$$Q^2 = \begin{bmatrix} q_{11} & \cdots & q_{1n} \\ \vdots & \ddots & \vdots \\ q_{n1} & \cdots & q_{nn} \end{bmatrix} \begin{bmatrix} q_{11} & \cdots & q_{1n} \\ \vdots & \ddots & \vdots \\ q_{n1} & \cdots & q_{nn} \end{bmatrix}, \tag{3}$$

where the (i, j) entry of Q^2 is given by:

$$Q_{ij}^2 = \mathbf{q_i} \cdot \begin{bmatrix} Q_{1j} \\ \vdots \\ Q_{nj} \end{bmatrix}.^6 \tag{4}$$

If we were to calculate the dot product we would get the following:

$$Q_{ij}^2 = \sum_{k=1}^{n} Q_{ik} Q_{kj}. \tag{5}$$

This form of expressing Q_{ij}^2 shows that the (i, j) entry of Q^2 is the sum of probabilities of every 2-step path from i to j, where k would be the intermediate step. Thus, Q_{ij}^2 is the total probability of getting from state i to state j in 2 steps, so $Q_{ij}^2 = p_{ij}^{(2)}$. Since this applies for all $i, j \in [1, n]$ the matrix Q^2 has entries that are the 2-step transition probabilities.
Inductive Hypothesis and Step. Now, assume that Q^k has entries that are

[5] Joseph Blitzstein, Introduction to Probability (Milton: CRC Press LLC, 2019), 500.
[6] Recall that $\mathbf{q_i}$ is a row vector, namely the i^{th} row of matrix Q

the k-step transition probabilities for $1 \leq k \leq n$. Then, if we consider Q^{n+1} we have:

$$Q^{n+1} = (Q^n)Q \tag{6}$$

Following the logic from before, the (i, j) entry of Q^{n+1} is given by:

$$Q_{ij}^{n+1} = (Q^n)_i \cdot \begin{bmatrix} Q_{1j} \\ \vdots \\ Q_{nj} \end{bmatrix}, \tag{7}$$

where $(Q^n)_i$ is the i^{th} row of Q^n. This is a row vector, but we wanted to avoid confusion relating to the exponent.

$$Q_{ij}^{n+1} = \sum_{k=1}^{n} Q_{ik}^{n} Q_{kj} \tag{8}$$

Recall we assumed Q^n will have entries that are the n-step transition probabilities and Q has entries that are the 1-step transition probabilities. Thus, we can rewrite this as:

$$Q_{ij}^{n+1} = \sum_{k=1}^{n} p_{ik}^{(n)} p_{kj}^{(1)} \tag{9}$$

A single term of the sum above represents the product of the probability of getting from state i to state k in n steps, and the probability of going from state k to state j in one step. This will repeat for k being all n possible states in our state space.

For each term, if it's not possible to reach state k from state i in n steps the term will become 0 as $p_{ik}^{(n)} = 0$. Likewise if it's not possible to reach state j from state k in one step the term becomes 0 as $p_{kj} = 0$. Thus, the only remaining terms will be the terms representing transitions where it's possible to go from i to k in n steps, and then from k to j in one step. Since the event of going from state i to state k is independent of the event of going from state k to state j by the Markov Property, their product is the probability of both events occurring. If both events occur, then we can simplify the event to the event of going from state i to state j in $n + 1$ step[7].

This means that the (i, j) entry of Q^{n+1} is the sum of all probabilities of going from state i to state j in $n + 1$ steps. So,

$$Q_{ij}^{n+1} = p_{ij}^{(n+1)} \tag{10}$$

Since this applies for all $i, j \in [1, n]$ the entries of Q^{n+1} are the (n+1)-step transition probabilities.

Therefore, by the Principle of Mathematical Induction, Q^n has entries that are the n-step transition probabilities for all $n \geq 1$. ∎

[7] This idea is explained in a lemma shown in section 4.1.

3.2 Application

How is this useful? Recall that our 64 x 64 matrix Q contains the single step transition probabilities of the knight moving from a square to any other square on the chessboard. Thus, if we wanted the probabilities of going from one square to any other square in n moves, we would look at the corresponding entry in Q^n.

4 Accessibility and Communication [8]

4.1 Theory

Following directly from the previous section, we can now begin to investigate how we are able to move through a Markov chain.

Definition 4.1.1. Given a Markov chain, a state j is *accessible* from state i (denoted $i \to j$) if there is a walk in the Markov chain that goes from state i to state j. In terms of n-step transition probabilities, $i \to j$ if $Q^n_{ij} > 0$ for some $n \in \mathbb{N}$.

An intuitive yet important relationship about accessibility follows.

Lemma 4.1.2. If there is an n-step walk from state i to state j, and an m-step walk from j to k, then there is an $m+n$-step walk from i to k, or equivalently,

$$(Q^n_{ij} > 0) \wedge (Q^m_{jk} > 0) \implies Q^{m+n}_{ik} > 0. \tag{11}$$

Yet another, simpler way to express this relationship is that

$$i \to j \wedge j \to k \implies i \to k. \tag{12}$$

This property is intuitively true, and its proof is left for the curious.

Definition 4.1.3. If states i and j are mutually accessible (that is, $i \to j$ and $j \to i$, for $i \neq j$), we say that they *communicate* and denote this relationship as $i \leftrightarrow j$.

Another intuitive property of accessibility and communication follows.

Lemma 4.1.4. If $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$.
This is quickly shown by noting that $i \leftrightarrow j$ and $j \leftrightarrow k$ imply $i \to j$ and $j \to k$, so $i \to k$ by the previous lemma. We can similarly show that $k \to i$, so $i \leftrightarrow k$.

Since we know our knight can reach any square on the chessboard, the concepts of accessibility and communication are less useful for solving our problem, but they provide useful foundations for the theory in later sections.

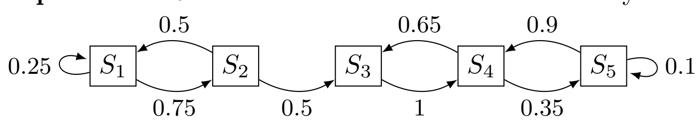
[8] Much of this section closely resembles pg 106 of **this** textbook snippet from MIT 6.262: Discrete Stochastic Processes, 2011. Accessibility is a very helpful way to frame the ideas of recurrent and transient states.

5 Recurrent and Transient States

5.1 Theory

If we leave a Markov chain to transition indefinitely, we would like to develop some tools to describe the behavior of the chain. For example, as n, the number of transition periods, tends to ∞, not all states in a Markov chain will necessarily be visited and revisited infinitely often. The following example illustrates this.

Example 5.1.1. Consider the Markov chain described by the following graph:



This kind of diagram is a common method of visualizing Markov chains. Each of $S_1, \dots, S_5$ represents one of the possible states of the system; each arrow and corresponding label represents a potential transition from one state to another, along with the corresponding transition probabilities. We can imagine starting at S_1 and moving randomly to other states indefinitely. Eventually, since there is an arrow from S_2 to S_3 but no arrow from S_3 to S_2, the system will leave the left half of the chain and remain in the right half (consisting of S_3, S_4, S_5) forever.

This example leads us naturally to characterize our states as those to which the system returns infinitely often (under mild conditions), and those to which the system returns finitely often.

Definition 5.1.2. State i of a Markov chain is *recurrent* if i is accessible from all states that are accessible from i. That is, i is recurrent if $i \to j$ implies $j \to i$ for all states j.[9]

Intuitively, this means that in the long run, if the chain starts at i, it will return to i infinitely many times. This follows from the idea that if the chain *can* return to i from any state after the start state, it eventually *will* return to i, and since this process occurs indefinitely, the chain will return to i without bound. The proof of this requires some more knowledge of probability and is again quite intuitive, and thus is left for the curious.

Once again, since we know that our knight can get anywhere on the board when starting from anywhere else on the board, all states in our problem are recurrent.

[9]Lifted closely from the same source as above: textbook snippet from MIT 6.262:Discrete Stochastic Processes, pg 107. This is only valid for finite-state Markov chains, the focus of this paper.

6 Irreducible Markov Chains and Recurrent States

6.1 Theory

Definition 6.1.1. A Markov chain is *irreducible* if for any two states i and j it is possible to get from state i to state j in a finite number of steps. In terms of linear algebra, this means that for any i and j there exists some positive integer n such that Q^n has entry (i, j) that is non-zero. If this is not the case, then the Markov chain is called *reducible*.

Theorem 6.1.2. *If a Markov chain has a finite state space and it is irreducible, then all states are recurrent.*

Proof. Since the chain is irreducible, this means that it is possible to get from all states i to all states j in a finite number of steps by definition of irreducibility. Thus, $i \leftrightarrow j$ for all states i, j, which means both $i \to j$, and $j \to i$. Thus all states are recurrent by definition. ■

6.2 Application

What does this mean for our knight on the chessboard? It means that, unless it falls off the chessboard, we know it will eventually return to its starting state! This fact is not immediately obvious as it seems possible for the knight to hop around on other squares and not come back to where it started. However, this theorem tells us that if we consider an infinite number of moves we can be certain that the knight will return home. The fact that the chain is irreducible is an important requirement for later theorems.

7 Periodicity; Periodic and Aperiodic Markov Chains

This section is not directly required for our problem. It's a mild detour; feel free to skip to Section 8: Stationary Distributions if you don't have any interest in Markov chains outside of our problem.

7.1 Theory

Definition 7.1.1. The *period* of a state i is the greatest common divisor (gcd) of the lengths of all possible walks that begin at i and return to i. In terms of our transition matrix Q, the period of i is the GCD of all numbers n such that $(Q^n)_{ii} > 0$. A state is called *aperiodic* if its period is 1, and *periodic* otherwise. If all states in a Markov chain are aperiodic, then the Markov chain itself is called *aperiodic*; the chain is *periodic* otherwise.

Theorem 7.1.2. *Given an irreducible Markov chain, all states have the same period.*

Proof. Suppose we have an irreducible Markov chain with state space $S = \{S_1, S_2, ..., S_M\}$. Since the chain is irreducible, we have $i \leftrightarrow j$ for all states $i, j \in S$. Suppose we can reach j from i in x transitions, and i from j in y transitions, for $i \neq j$. Then

$$(Q^x)_{i,j} > 0, (Q^y)_{j,i} > 0 \tag{13}$$

by n-step transition probability. Now define the set J as the set of all numbers z such that $(Q^z)_{j,j} > 0$ - that is, J contains all possible walk lengths to go from state j back to itself. Define the set I identically for i.
Now note that for every element z in J,

$$x + z + y \in I, \tag{14}$$

since we go from i to j, move around and get back to j, then return from j to i. Note that it does not matter if, in the section starting and ending at j, we happen to visit i once again; it only matters that $(Q^{x+y+z})_{i,i} > 0$. Now, since I describes all length walks from i back to itself, I must consist of multiples of the period of i. Call this period $d(i)$. If we take any two elements of J, say, z_1 and z_2, for $z_2 \geq z_1$, note that

$$(x + y + z_2) - (x + y + z_1) = z_2 - z_1 = kd(i) \tag{15}$$

for some nonnegative integer k, since $d(i)$ is defined as the GCD of all elements of I, which both of the terms in (15) are in by (14).
Thus, for any two elements in J, their (absolute) difference is an element in I. Since there is at least one element of J that is also in I [10], it follows that all elements of J are multiples of $d(i)$ (as at least one element is a multiple of $d(i)$, and all pairwise differences are multiples of $d(i)$).
Thus, $d(j) \geq d(i)$, as all elements of J are multiples of $d(i)$ and $d(j)$ is the gcd of all of these elements.
There is nothing special here about i and j. Thus, it follows from symmetry that $d(i) \geq d(j)$, so $d(i) = d(j)$.
We have then shown that any two states in an irreducible Markov chain have the same period. ■

7.2 Application

Periodicity, while a good thing to know in general, is only useful to us because it *rules out* a potential avenue for solving our problem. Consider the way a knight moves: starting on a light square, the knight moves to a dark square, and vice versa. This tells us that to return to its original square, the knight must move an even number of times, and thus all squares have period 2. The reason that this is important is because of a theorem related to *stationary distributions* (the next section), which says that if you run an **aperiodic**, irreducible Markov

[10] A path that goes from i to j then back to i can easily be viewed as a path from j to i and then back to j by swapping the 'forward' and 'backward' paths.

chain for a long time, the distribution of probabilities will converge to the chain's stationary distribution. This would be very helpful in understanding the long term behavior of the knight on the board. However, in our case, since our chain *is* periodic, if we run the chain for a long time, it will *not* converge to the stationary distribution. This makes intuitive sense: for all even-numbered iterations of the chain, the knight will be on a square corresponding to its starting square, so the probabilities of being on any oppositely-colored square are 0, while for any odd number of moves, the probabilities of being on any square of the same color as the starting square are 0.

8 Stationary Distribution

8.1 Theory

As mentioned at the end of the last section, we want to build some tools that allow us to understand the long-term behavior of Markov chains. Stationary distributions are important in doing so.

Definition 8.1.1. A *stationary distribution* of a Markov chain with transition matrix Q is a probability vector $\mathbf{s}$ with entries such that

$$\mathbf{s}Q = \mathbf{s}. \tag{16}$$

For irreducible Markov chains, the stationary distribution is unique. [11]

Thus, $\mathbf{s}$ is a *left eigenvector* of Q, with eigenvalue 1. This tells us that when the system has state according to the entries of the stationary distribution, when we make a transition, the probabilities of the system being in any state remain unchanged.

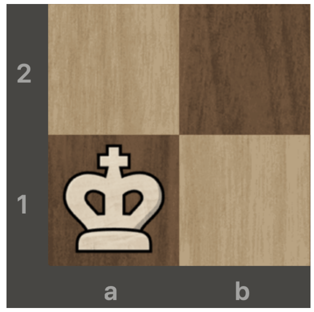
Given a transition matrix Q, you can solve for its left eigenvector of eigenvalue 1. If such an eigenvector exists, then it is the stationary distribution of our state space.
To solve for the left eigenvector, you could use the transpose of the transition matrix, Q^T, and solve for its *right* eigenvector of eigenvalue 1. If the eigenvector exists, then its transpose will be the stationary distribution. In other words, the following equation is equivalent,

$$Q^T\mathbf{s}^T = \mathbf{s}^T \tag{17}$$

but it's important to remember we're solving for the *left* eigenvector $\mathbf{s}$, not its transpose.

Example 8.1.2. It turns out that the stationary distribution tells us what fraction of time the chain will spend at each of its recurrent states. For some intuition as to why, recall our king problem from earlier.

[11] Joseph Blitzstein, Introduction to Probability (Milton: CRC Press LLC, 2019), 500.



The stationary distribution for the transition matrix,

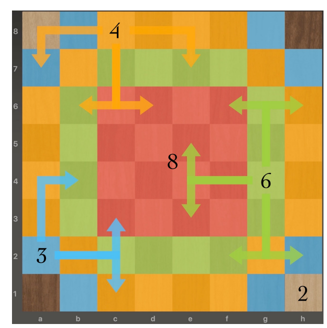
$$Q = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix} \tag{18}$$

is given by

$$\mathbf{s} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}. \tag{19}$$

If we let the king roam his empire for eternity, this is exactly what we would expect! Of course the king will spend a quarter of its time at each square.

8.2 Application



What does the stationary distribution mean for our problem? The entries of the stationary distribution provide very important information. Namely, if we were to consider setting the knight loose and allowing it to roam across the chessboard forever, the entries would tell us the proportion of time spent on the corresponding square.
Now, how do we solve for the stationary distribution of our particular state space? Well, we could solve for the eigenvector of eigenvalue 1 as we have done in class - by subtracting the identity matrix from our matrix Q and finding the basis of its null space. However, with a 64 x 64 matrix this could be a bit tedious. Instead we will use a different approach that, although not as rigorous, follows intuitive logic which reveals the solution rather elegantly.

Consider the diagram from earlier, shown to the right. Now, imagine that for each square you place the number of knights on the square equal to the number of possible moves from said square. So, for example, each red square would contain 8 knights, each green square 6 knights and so on, for a total of 336 knights. Now, consider what would happen if we allowed every single knight on the chessboard to make a single transition, or move, at random. We could expect that, on average, the same number of knights will end up on each square.[12]

[12]For those familiar with chemistry, this is identical to the concept of dynamic chemical equilibrium.

Now, let's consider the vector $\mathbf{k} \in \mathbb{R}^{64}$ whose entries are the number of knights in each corresponding square on the chessboard. If we were to normalize this we would arrive at the stationary distribution. That is, if we scaled this vector so that its entries summed to 1 we'd get:

$$\sum_{i=0}^{64} \mathbf{k}_i = 336 \tag{20}$$

$$\frac{1}{336}\mathbf{k} = \mathbf{s} \tag{21}$$

where s is the stationary distribution. The probability vector s would have $\frac{1}{168}$ in entries corresponding the corner squares, $\frac{1}{112}$ in entries corresponding to the blue squares, $\frac{1}{84}$ in entries corresponding to orange squares, $\frac{1}{56}$ in entries corresponding to green squares and $\frac{1}{42}$ in entries corresponding to red squares.

9 Expected Time of Return

9.1 Theory

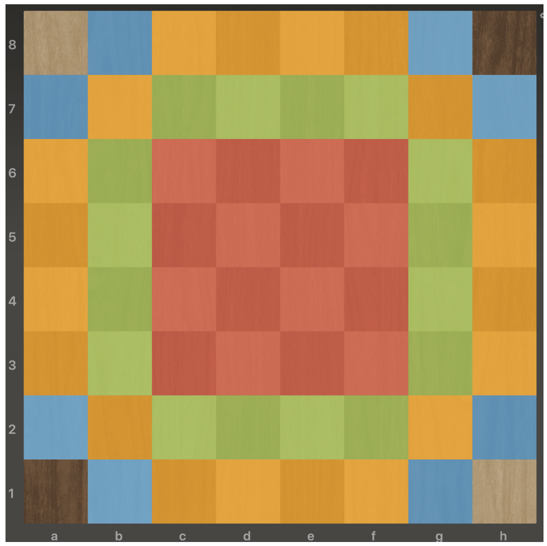
Theorem 9.1.1. *[1] Let $\mathbf{x_0}, \mathbf{x_1}, ... \in \mathbb{R}^M$ be an irreducible Markov chain with state space $S = \{S_1, S_2, ..., S_M\}$ and stationary distribution $\mathbf{s}$, where s_i is the i-th entry. Let r_i be the expected time it takes the chain to return to state i, given that it starts at i. Then $s_i = \frac{1}{r_i}$.*

The proof of this theorem is quite involved, so instead, we will give some intuition for it.[13] Suppose you start at state i of the chain, and do a random walk with probabilities corresponding to the transition matrix Q. The proportion of time you spend at any state j (which you can think of as the limit of the proportion of time spent at j in the first n steps, as n tends to infinity) is by definition s_j, the value of the stationary distribution at state j, so the proportion of time spent at i is s_i. But the proportion of time you spend at i is also the reciprocal of the expected return time to i, and thus these quantities are equal, and

$$s_i = \frac{1}{r_i}. \tag{22}$$

[13] A full proof can be found in [3], pp 264-265. Be warned that the notation is different and may require some back-reading. Alternatively, a less rigorous but more accessible proof is given at [5], on page 5.

9.2 Application



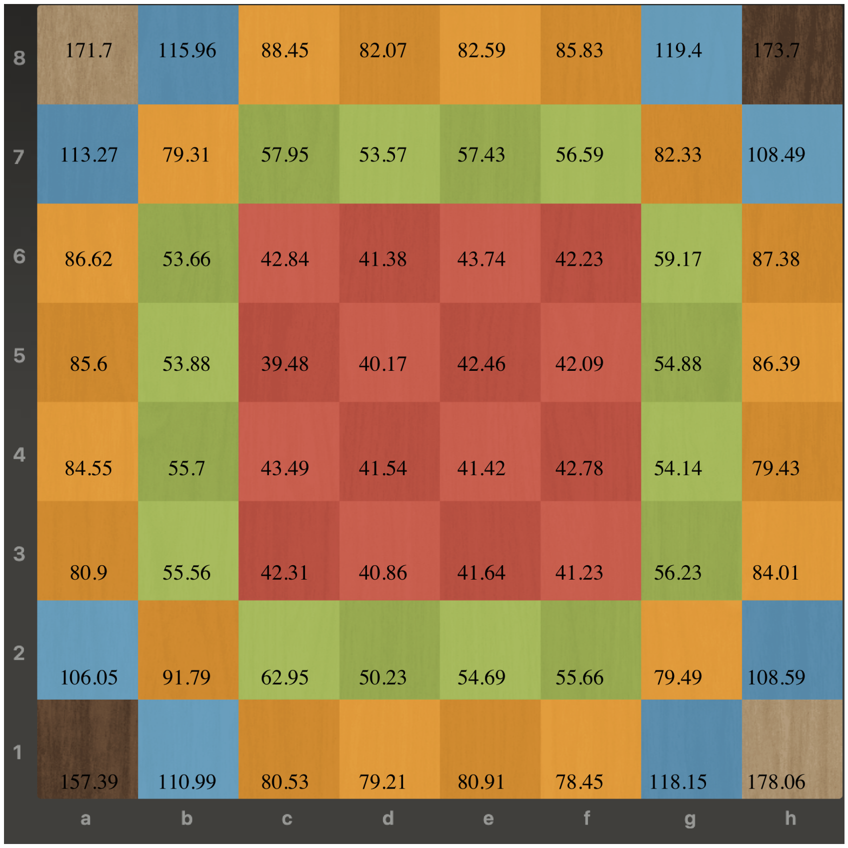
Applying the above theorem to our stationary distribution from Section 8.2, we find that the knight's expected return time is:

- 168 moves if the knight starts on a corner square.
- 112 moves for a blue square.
- 84 moves for an orange square.
- 56 moves for a green square.
- 42 moves for a red square.

Douglas Adams is well-versed in this problem, apparently! With that, we obtain a satisfying conclusion to an interesting problem. ■

10 Simulation

Using C++, we simulated the wandering knight problem for every starting square, conducting 1001 trials on each square. The average number of return moves for each square over the 1001 trials is represented in the image below.



As you can see, most squares returned an average close to our expected values. Interestingly, as we might expect, the squares with a higher expected return time appear to have higher variation in their average times. For a corner square, for example, there is a $\frac{1}{6}$ chance of returning to the corner square after just 2 moves, so there must be cases where the return time is much, much longer. In fact, we ran a million simulations on one of the corner squares, and found that that the longest return time was a whopping 3579! (The exclamation mark does not denote factorial, of course, but rather our excitement.)

An interesting topic for further study may be considering the expected time to go from any square to *any other* square.

References

[1] Blitzstein, Joseph K, and Hwang, Jessica. *Introduction to Probability, Second Edition.* Chapman & Hall/CRC Texts in Statistical Science. Milton: CRC Press LLC, 2019.

[2] Lay, David C., Steven R. McDonald, and McDonald, Judith. *Linear Algebra and Its Applications.* Fifth Edition. Harlow: Pearson Education Limited, 2016.

[3] Durrett, Rick. Probability. Vol. 49. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 2019.

[4] MIT OpenCourseWare; 6.262 Discrete Stochastic Processes. *Finite-State Markov Chains.* https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-262-discrete-stochastic-processes-spring-2011/course-notes/MIT6_262S11_chap03.pdf (accessed December 10, 2020)

[5] Sigman, Karl. "More on Discrete-Time Markov Chains." http://www.columbia.edu/ ww2040/4701Sum07/4701-06-Notes-MCII.pdf (accessed December 10, 2020).