Markov Chains

Adam Kelly (`ak2316@cam.ac.uk`)

October 19, 2021

A stochastic process is said to have the 'Markov property' if, conditional on its present value, the future is independent of the past. This is a *restrictive* assumption, but we do end up with a useful model with a rich mathematical theory, which we shall study in this course.

This article constitutes my notes for the 'Markov Chains' course, held in Michaelmas 2021 at Cambridge. These notes are *not a transcription of the lectures*, and differ significantly in quite a few areas. Still, all lectured material should be covered.

Contents

§1 The Markov Property

§1.1 What is a Markov Chain?

Let S be a countable set (the set of possible 'states'), and let X_n be a sequence of random variables taking values in S.

> **Definition 1.1** (Markov Chain)
>
> The sequence of random variables X_n is a **Markov chain** if it satisfies the **Markov property**
>
> $$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, X_0 = x_0) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n).$$
>
> The Markov chain is said to be **homogeneous** if for all $i, j \in S$ the conditional probability $\mathbb{P}(X_{n+1} = j \mid X_n = i)$ is independent of n.

In this course we are only going to study homogeneous Markov chains.

Markov chains are often best described by diagrams[1] which show the probability of moving from one state to another. For example, the Markov chain in the diagram below has three states which we label $\{1, 2, 3\}$, and the probability of moving from state 1 to state 2 is $1/2$, and the probability of moving from state 2 to state 3 is $1/3$, and so on.
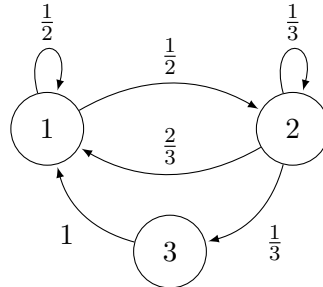
In general, to calculate the probabilities associated with a Markov chain, we need to know two quantities.

- *The initial distribution.* We first need to know about the starting state of a Markov chain. This is described by the initial distribution $\lambda = (\lambda_i \mid i \in S)$, where $\lambda_i = \mathbb{P}(X_0 = i)$.

- *The transition probabilities.* We also need to know the probability of moving from a state $i \in S$ to a state $j \in S$. This is typically given by a *transition matrix* $P = (P_{i,j} \mid i, j \in S)$ with $p_{i,j} = \mathbb{P}(X_1 = j \mid X_0 = i)$.

These quantities are of course subject to some constraints, in that we require $\sum_{i \in S} \lambda_i = 1$ and the transition matrix P must be **stochastic**, in that $\sum_{j \in S} P_{i,j} = 1$ for all $i \in S$.

If a Markov chain X_n has initial distribution λ and transition matrix P, we say that it is Markov(λ, P).

Once we have these quantities, we can begin to actually establish various properties about the Markov chain, for example the probability that it goes through a given sequence of states.

Theorem 1.2 (Probability of a Sequence of States)

The sequence of random variables X_n is Markov(λ, P) if and only if

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \lambda_{i_0} P_{i_0, i_1} \cdots P_{i_{n-1}, i_n},$$

for all $n \geq 0$ and $i_0, \ldots, i_n \in S$.

Proof. Let A_k denote the event $\{X_k = i_k\}$. First suppose that X_n is Markov(λ, P). We prove the result holds by induction. For $n = 0$ this is true by definition. Then if it holds up to n, we have

$$\mathbb{P}(A_0 \cap \cdots \cap A_n) = \mathbb{P}(A_0 \cap \cdots \cap A_{n-1})\mathbb{P}(A_n \mid A_0 \cap \cdots \cap A_{n-1})$$
$$= \mathbb{P}(A_0 \cap \cdots \cap A_{n-1})\mathbb{P}(A_n \mid A_{n-1})$$

---

[1]You might notice that these diagrams are labelled directed graphs, and indeed you will see concepts from graph theory such as connectivity reoccur when we later talk about communicating classes.

$$= \left(\lambda_{i_0} P_{i_0,i_1} \cdots P_{i_{n-2},i_{n-1}}\right) P_{i_{n-1},i_n},$$

which completes our induction.

Conversely, suppose that the result holds. Then with $n = 0$ we get that the initial distribution of X_n is λ. Then

$$\mathbb{P}(A_{n+1} \mid A_0 \cap \cdots \cap A_n) = \frac{\mathbb{P}(A_0 \cap \cdots \cap A_{n+1})}{\mathbb{P}(A_0 \cap \cdots \cap A_n)} = P_{i_n,i_{n+1}}.$$

Since this does not depend on $i_0, \ldots, i_{n-1}$, X_n is a homogeneous Markov chain with transition matrix P, as required.   □

§1.2 Simple Markov Property

An important aspect of Markov chains is that they are *memoryless*, in the future is independent of the past, conditional on the present. This is encapsulated in the **simple Markov property**.

Theorem 1.3 (Simple Markov Property)

Let X_n be a Markov chain. Then conditional on $X_m = i$, the sequence of random variables $(X_{m+n})_{n \geq 0}$ is[a] Markov(δ_i, P), and is independent of $X_0, \ldots, X_m$.

---
[a]Here δ_{ij} is 1 if $i = j$ and 0 otherwise.

Proof. Given any event H determined by $X_0, \ldots, X_m$, we want to show that for an event $F = \{X_m = i_m, \ldots X_{m+n} = i_{m+n}\}$ we have

$$\mathbb{P}(H \cap F \mid X_m = i) = \delta_{ii_m} P_{i_m,i_{m+1}} \cdots P_{i_{m+n-1},i_{m+n}} \mathbb{P}(H \mid X_m = i).$$

Indeed, considering the case of $H = \{X_0 = i_0, \ldots, X_m = i_m\}$ we have

$$\mathbb{P}(H \cap F \mid X_m = i) = \frac{\lambda_{i_0} P_{i_0,i_1} \cdots P_{i_{m-1},i} P_{i,i_{m+1}} \cdots P_{i_{m+n-1},i_{m+n}}}{\mathbb{P}(X_m = i)}$$

$$= \delta_{ii_m} P_{i,i_{m+1}} \cdots P_{i_{m+n-1},i_{m+n}} \mathbb{P}(H \mid X_m = i),$$

as required.

Then for a general H, we can write it as the disjoint union $H = \bigcup_{k=1}^{\infty} H_k$, and then the overall result follows by summing the above result for relevant H_k.   □

§1.3 Transition Probabilities

We are now going to address how to find the probability that a Markov chain is in a given state after n steps. The core idea of this section is that we will be able to reduce such questions into questions about the transition matrix that we introduced earlier.

Recall that if X_n is a Markov chain with transition matrix P, then $P_{i,j}$ was the probability of moving from the state i to the state j. We call these the **one-step transition probabilities**. We can generalize this slightly.

> **Definition 1.4** (n-Step Transition Probabilities)
>
> For a Markov chain X_n, the n**-step transition probabilities** are given by
>
> $$p_{i,j}(n) = \mathbb{P}(X_n = j \mid X_0 = i).$$

These n-step transition probabilities naturally form the n**-step transition matrix** $P(n) = (p_{i,j}(n) \mid i, j \in S)$. The nice thing about writing these transition probabilities as a matrix is that they satisfy a lovely set of equations that relate extremely well to matrix algebra.

> **Theorem 1.5** (Chapman-Kolmogorov Equations)
>
> We have that
> $$p_{i,j}(n + m) = \sum_{k \in S} p_{i,k}(n) p_{k,j}(m),$$
> where $i, j \in S$ and $m, n \geq 0$. In particular, $P(m + n) = P(m)P(n)$.

> *Proof.* Using the partition theorem and simple Markov property,
>
> $$\begin{aligned} p_{i,j}(n + m) &= \mathbb{P}(X_{n+m} = j \mid X_0 = i) \\ &= \sum_{k \in S} \mathbb{P}(X_{m+n} = j \mid X_n = k, X_0 = i)\mathbb{P}(X_n = k \mid X_0 = i) \\ &= \sum_{k \in S} \mathbb{P}(X_{m+n} = j \mid X_n = k)\mathbb{P}(X_n = k \mid X_0 = i) \\ &= \sum_{k \in S} p_{i,k}(n) p_{k,j}(m). \end{aligned}$$
>
> $\square$

So, if we have some n-step transition matrix $P(n)$, the above result shows us that it satisfies $P(n) = P^n$. This reduces our problem to just this:

<p align="center" style="color:blue">To compute $p_{i,j}(n)$, we can compute powers of
the transition matrix, and take $(P^n)_{i,j}$.</p>
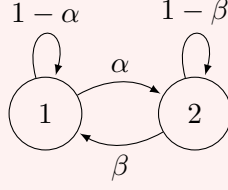
In general, this makes our problem significantly easier, and if the state space is finite, we can use tools from linear algebra such as diagonalisation.

> **Example 1.6** (Computing Transition Probabilities)
>
> Let $\alpha, \beta \in (0, 1)$. Consider the Markov chain X_n with states $S = \{1, 2\}$, with transition matrix
> $$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}.$$
> A diagram of this markov chain is shown below.

We want to find the n-step transition probabilities.

Method 1 (Difference Equations). We are first going to find $p_{1,1}(n)$ using a 'bare-hands' method. By the Chapman-Kolmogorov equations (that is, conditioning on X_n) we have

$$
\begin{aligned}
p_{1,1}(n+1) &= p_{1,1}(n)p_{1,1}(1) + p_{1,2}(n)p_{2,1}(1) \\
&= (1-\alpha)p_{1,1}(n) + \beta p_{1,2}(n) \\
&= (1-\alpha)p_{1,1}(n) + \beta(1 - p_{1,1}(n)) \\
&= p_{1,1}(n)(1 - \alpha - \beta) + \beta.
\end{aligned}
$$

This is a recurrence relation which we can then solve using the boundary condition $p_{1,1}(0) = 1$ to get

$$
p_{1,1}(n) = \begin{cases} \frac{\alpha}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta} \cdot (1-\alpha-\beta)^n & \text{if } \alpha + \beta > 0, \\ 1 & \text{if } \alpha + \beta = 0. \end{cases}
$$

Method 2 (Diagonalisation). An alternative solution uses some tools from matrix algebra. To calculate the n-step transition matrix P^n, we are going to diagonalise P.

The eigenvalues of P are given by the solutions to $\det(P - \mu I) = 0$, which are $\{1, 1 - \alpha - \beta\}$. Thus for some invertible matrix U we have

$$
P = U^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 1-\alpha-\beta \end{pmatrix} U \quad \text{and} \quad P^n = U^{-1} \begin{pmatrix} 1 & 0 \\ 0 & (1-\alpha-\beta)^n \end{pmatrix} U.
$$

Thus $p_{1,1}(n) = A + B(1-\alpha-\beta)^n$, for constants A and B. We can find these by noting the boundary conditions $p_{1,1}(0) = 1$ and $p_{1,1}(1) = 1 - \alpha$, giving

$$
p_{1,1}(n) = \begin{cases} \frac{\alpha}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta} \cdot (1-\alpha-\beta)^n & \text{if } \alpha + \beta > 0, \\ 1 & \text{if } \alpha + \beta = 0. \end{cases}
$$

In general, if the state space of a Markov chain is finite with $|S| = k$, then P is a $k \times k$ matrix with eigenvalues $\mu_1, \ldots, \mu_k$.

If *all of the eigenvalues are distinct*, then P is diagonalisable, and we can write

$$
P = U^{-1} \begin{pmatrix} \mu_1 & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \mu_k \end{pmatrix} U \quad \text{and} \quad P^n = U^{-1} \begin{pmatrix} \mu_1^n & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \mu_k^n \end{pmatrix} U,
$$

and $p_{i,j} = a_1 \mu_1^n + \cdots + a_k \mu_k^n$, for some constants $a_1, \ldots, a_k$, which are determined by the boundary conditions.

If some eigenvalue μ_k is complex, then it's conjugate is also an eigenvalue, so if $\mu_k = re^{i\theta}$

we have also the eigenvalue $\overline{\mu_k} = re^{-i\theta}$, and we can write

$$p_{i,j} = a_1\mu_1^n + \cdots + a_{k-2}\mu_{k-2}^n + a_{k-1}r^n\cos(n\theta) + a_k r^n\sin(n\theta),$$

since $p_{i,j}$ is real and so all of the imaginary parts must cancel out.

If *some eigenvalues repeat*, then the situation is slightly more complicated. If an eigenvalue μ_k has multiplicity m, we can simply replace a_k by a degree m polynomial in n^2.[2]

Example 1.7 (Transition Matrix with Complex Eigenvalues)

Consider the markov chain X_s with states $S = \{1, 2, 3\}$ as shown in the diagram below. We want to find the n-step transition probability $p_{i,i}(n)$.
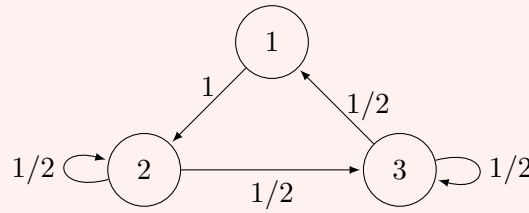
This Markov chain has the transition matrix

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix},$$

which has distinct eigenvalues $\{1, i/2, -i/2\}$.

We can rewrite the complex eigenvalues using trigonometric functions as

$$\frac{i}{2} = \cos\frac{\pi}{2} + i\sin\frac{\pi}{2},$$
$$-\frac{i}{2} = \cos\frac{\pi}{2} - i\sin\frac{\pi}{2}.$$

The general form for $p_{1,1}(n)$ is then given by

$$p_{1n1}(n) = A + B \cdot \left(\frac{1}{2}\right)^n \cos\left(\frac{n\pi}{2}\right) + C \cdot \left(\frac{1}{2}\right)^n \sin\left(\frac{n\pi}{2}\right).$$

The boundary conditions can be computed by hand as $p_{1,1}(0) = 1$, $p_{1,1}(1) = 0$ and $p_{1,1}(2) = 0$. This allows us to solve for A, B and C to get

$$p_{1,1}(n) = \frac{1}{5} + \left(\frac{1}{2}\right)^n \left(\frac{4}{5}\cos\left(\frac{n\pi}{2}\right) - \frac{2}{5}\sin\left(\frac{n\pi}{2}\right)\right).$$

---

[2]This comes from considering the Jordan Normal form of the transition matrix.

§2 Class Structure

§2.1 Communicating Classes

In Graph Theory, one can often make a problem easier by looking at the *connected components* of the graph, effectively splitting it up into smaller disconnected chunks which are easier to understand on their own. This has a natural analogue in the study of Markov chains,[3] where we consider which states can interact.

> **Definition 2.1** (Communication)
>
> Let X_n be a Markov chain. For $i, j \in S$, we say that i **leads to** j, written $i \to j$, if $p_{i,j}(n) > 0$ for some $n \geq 0$.
>
> If $i \to j$ and $j \to i$, we say that i and j **communicate**, and write $i \leftrightarrow j$.

We can use this idea of communication to break up our state space by noting that communication is an equivalence relation.

> **Theorem 2.2** (Communication is an Equivalence Relation)
>
> The relation $\leftrightarrow$ is an equivalence relation on S.

Proof. For any $i \in S$, $p_{i,i}(0) = 1$ and thus $i \leftrightarrow i$, so the relation is reflexive. The relation is symmetric by definition, so we are just left to check transitivity.

Let $i, j, k \in S$, and suppose that $i \leftrightarrow j$ and $k \leftrightarrow k$. Then there exists $m, n \geq 0$ such that $p_{i,j}(n) > 0$ and $p_{j,k}(m) > 0$. Then by the Chapman-Kolmogorov equations,

$$p_{i,k}(m+n) = \sum_{l \in S} p_{i,l}(m) p_{l,k}(n) \geq p_{i,j}(n) p_{j,k}(m) > 0,$$

so $i \to k$. By the same argument $k \to i$, and thus $\leftrightarrow$ is transitive. $\square$

Because communication is an equivalence relation, it partitions the state space.

> **Definition 2.3** (Communicating Classes)
>
> The equivalence classes induced by $\longleftrightarrow$ on S are called **communicating classes**.

Because communication is really a notion about directed graphs (and not *really* about individual probabilities), it's usually relatively easy to identify the communicating classes from the diagram of a Markov chain.

For example, the Markov chain below has its communicating classes coloured (and the transition probabilities are not labelled, as they are not needed to infer the communicating classes, aside from knowing they are non-zero along each edge).

---

[3]Indeed, one can view a Markov chain as a labelled directed graph, and then apply the same connected components argument to obtain the theory shown here.
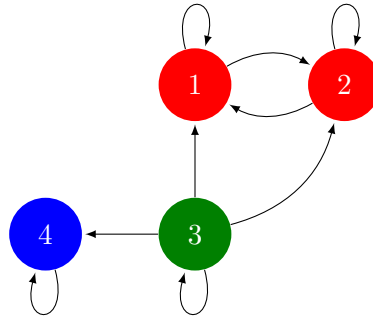
In the Markov chain above, we can see that there are a few different ways that communicating classes can appear.

A communicating class C is **closed** if there is no way to leave it, in that if $i \in C$ and $i \to j$, then j in C. For example, in the Markov chain above, the communicating class $\{1, 2\}$ is closed.

A state $i \in S$ is **absorbing** if there is no way to leave that state, that is, $\{i\}$ is a communicating class. In the Markov chain above, 4 is absorbing.

If a Markov chain has only one communicating class, we call is **irreducible**, as we cannot split it into smaller parts in this way.

§3 Hitting Times and Absorption Probabilities

§3.1 Hitting Times

When studying Markov chains, a frequently occurring question is '*how long does it take to reach a given state $i \in S$?*', a concept which we formulate as the *hitting time*.

> **Definition 3.1** (Hitting Time)
>
> Let X_n be a Markov chain. The **hitting time** of some subset of states $A \subseteq S$ is the random variable T_A, which is the least n for which $T_n \in A$. In particular,
>
> $$T_A = \inf\{n \geq 0 \mid X_n \in A\},$$
>
> where this may be ∞ if $X_n \notin A$ for all $n \geq 0$.

In this section, we will also look some related quantities, such as the *hitting probability* and *mean hitting time*.

> **Definition 3.2** (Hitting Probability)
>
> Let X_n be a Markov chain, and let $A \subseteq S$ be some subset of states. The **hitting probability** h_i^A of A is the probability that X_n eventually reaches a state in A, given $X_0 = i$. That is,
> $$h_i^A = \mathbb{P}(T_A < \infty \mid X_0 = i),$$
> where $i \in S$.

> **Definition 3.3** (Mean Hitting Time)
>
> Let X_n be a Markov chain, and let $A \subseteq S$ be some subset of states. The **mean hitting time** k_i^A of A is the expected number of steps taken for X_n to reach a state in A, given $X_0 = i$. That is,
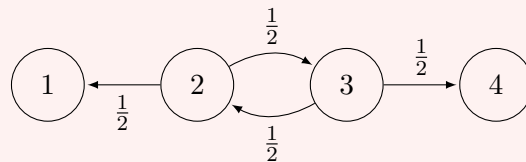>
> $$k_i^A = \mathbb{E}[T_A \mid X_0 = i],$$
>
> where $i \in S$. We note that this may be ∞.

So how do we go about calculating these quantities? An intuitive way is to try and consider what happens as we move from one state to another, as we will see in the example below.

> **Example 3.4** (Computing Hitting Probabilities and Mean Hitting Times)
>
> Consider the Markov chain in the diagram below.
>
>
>
> Suppose we took $A = \{4\}$ and wanted to find the hitting probability $h_2^A = \mathbb{P}(T_A < \infty \mid X_0 = 2)$. One might intuitively suppose that
>
> $$h_2^A = P_{2,3} h_3^A = \frac{1}{2} h_3^A,$$
> $$h_3^A = P_{3,2} \cdot h_2^A + P_{3,4} \cdot h_4^A = \frac{1}{2} \cdot h_2^A + \frac{1}{2}.$$
>
> which when solved gives us $h_2^A = 1/3$.
>
> Alternative, if we took $B = \{1, 4\}$ and wanted to find the mean hitting time k_2^B, we could again try the intuitive approach of writing
>
> $$k_2^B = 1 + \frac{1}{2} k_3^B,$$
>
> and noting that $k_3^B = k_2^B$ by symmetry, giving $k_2^B = 2$.

In the computations above, we really should check that these are valid methods (though it *really is* quite intuitive).

> **Theorem 3.5** (Computing Hitting Probabilities)
>
> Let $A \subseteq S$. Then the set of hitting probabilities $h^A = (h_i^A \mid i \in S)$ is the minimal[a] non-negative solution to the equations
>
> $$h_i^A = \begin{cases} 1 & \text{if } i \in A, \\ \sum_{j \in S} P_{i,j} h_j^A & \text{if } i \notin A. \end{cases}$$
>
> ------------
> [a]Minimal in that if x is another non-negative solution, then $x_i \geq h_i^A$ for all $i \in S$.

Proof. We first check that h_i solves this system. Clearly if $i \in A$, then $h_i^A = 1$, so take $i \in S \backslash A$. Then conditioning on the first step in the chain,

$$h_i^A = \sum_{j \in S} P_{i,j} \mathbb{P}(T_A < \infty \mid X_0 = i, X_1 = j)$$
$$= \sum_{j \in S} P_{i,j} h_j^A,$$

by the simple Markov property. So h_i does indeed solve this system.

We now check minimality. Let $x = (x_i \mid i \in S)$ be another non-negative solution. For $i \in A$, we have $x_i = h_i^A = 1$, so that holds. Then for $i \in S \backslash A$, since x satisfies the system of equations we can write

$$x_i = \sum_{j \in S} P_{i,j} x_j = \sum_{j \in A} P_{i,j} x_j + \sum_{j \in S \backslash A} P_{i,j} x_j. \tag{†}$$

Then with $x_j = 1$ for $j \in A$ and $x \geq 0$, we have

$$x_i \geq \sum_{j \in A} P_{i,j} = \mathbb{P}(X_1 \in A \mid X_0 = i) = \mathbb{P}(T_A < 2 \mid X_0 = i).$$

Similarly, expanding (†), we have

$$x_i = \mathbb{P}(X_1 \in A \mid X_0 = i) + \sum_{j \in S \backslash A} P_{i,j} \left(\sum_{k \in A} P_{j,k} x_k + \sum_{k \in S \backslash A} P_{j,k} x_k \right)$$
$$\geq \mathbb{P}(X_1 \in A \mid X_0 = i) + \mathbb{P}(X_1 \notin A, X_2 \in A \mid X_0 = i)$$
$$= \mathbb{P}(T_A < 3 \mid X_0 = 1).$$

We can repeat this argument to eventually establish that $x_i \geq \mathbb{P}(T_A < n \mid X_0 = i)$, for all $n \geq 0$. Then as $n \to \infty$, we have $x_i \geq \mathbb{P}(T_A < \infty \mid X_0 = i)$, as required. $\quad\square$
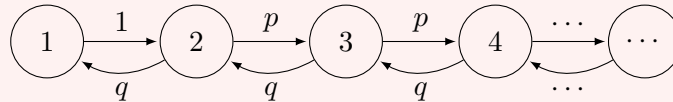
This method of computing hitting probabilities is perfectly valid on Markov chains with infinite state spaces.

Example 3.6 (Hitting Probabilities on an Infinite State Space)

Consider the Markov chain with states $S = \{0, 1, \dots\}$ such that $P_{0,1} = 1$, and

$$P_{i,i+1} = p, \quad P_{i,i-1} = q \quad \text{for all } i \geq 1,$$

where $p, q \in (0, 1)$ with $p + q = 1$.

Let $h_i = \mathbb{P}(T_0 < \infty \mid X_0 = i)$, and suppose that $p \neq q$. Then we clearly have $h_0 = 1$, and generally we can write down the recurrence

$$h_i = p h_{i+1} + q h_{i-1}.$$

This then implies $p(h_{i+1} - h_i) = q(h_i - h_{i-1})$, so that

$$h_{i+1} - h_i = \left(\frac{q}{p}\right)^i (h_1 - 1).$$

This inspires us to write h_i as a telescoping sum

$$h_i = \sum_{j=1}^{i} (h_j - h_{j-1}) + 1$$

$$= (h_1 - 1) \sum_{j=1}^{i} \left(\frac{q}{p}\right)^j + 1,$$

which gives us our general solution $h_i = a + b(q/p)^i$, which we can then solve.

In the case that $q > p$, then in order for h_i to be the minimal solution, we need to have $a = 1$ and $b = 0$, so $h_i = 1$.

We also need to check that our method for calculating the mean hitting time was valid.

Theorem 3.7 (Computing Mean Hitting Times)

Let $A \subseteq S$. Then the set of mean hitting times $k^A = (k_i^A \mid i \in S)$ is the minimal non-negative solution to the equations

$$k_i^A = \begin{cases} 0 & \text{if } i \in A, \\ 1 + \sum_{j \in S \setminus A} P_{i,j} k_j^A & \text{if } i \notin A. \end{cases}$$

Proof. To be completed next lecture. $\quad\square$