Markov Chain Case Study

EA 1

1 Introduction

In this case study we will learn about a mathematical modeling tool called Markov chains.

A Markov chain is a very common tool in probabilistic modeling that you will undoubtedly see again and again. Within the Markov chain framework, a system is described by a collection of states (e.g. pages in the world wide web or the state of the weather) with probabilities of transitioning from one state to another at any given time. The essential property of a Markov chain is the *Markov Property* or *Memorylessness*, meaning that the next state depends only on the current state and not any previous states. For instance, a spring will expand only if it is currently compressed; it is irrelevant if it has been compressed in the past. This allows us to predict the future of the system we are modeling with some certainty, based on only what we know right now.

Markov chains have applications ranging from text generation and speech recognition to statistical mechanics and chemical reaction networks. You'll find them used by neuroscientists and economists, among researchers in just about any field that leverages mathematical modeling. In this case study, we will see how Markov chains can be applied to the familiar problem of PageRank as well as to epidemic population models.

1.1 The Markov Chain

In a Markov chain, we model the *states* of a system as a network with edge weights that correspond to the probability of transitioning from one state to another at a given time step. The states of a system can be anything; for example, in PageRank the states will be particular websites, but if we were modelling weather the states could be conditions such as sunny, cloudy, or raining. The *transition probabilities* given by the weights of the links in the Markov chain tell you the probability of going from one state to another (e.g. if I am on Reddit now, what is the probability the next website I visit will be Imgur? Or if it is raining today, what is the probability it will be sunny tomorrow?). Figure 1 shows an abstract Markov chain with two states. Markov chains are allowed to have self loops since, in many systems, you do not have to leave your current state (e.g. it can be raining today and it can keep raining tomorrow).

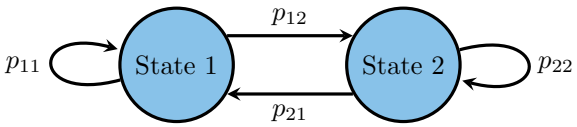


Figure 1: A general Markov chain.

1.2 Transition Probabilities

So far we have used the word probability without defining it; let us be more precise. A transition probability, p_{ij}, from state i to state j is a nonnegative value that is less than or equal to one, i.e. $0 \leq p_{ij} \leq 1$. Furthermore, the transition probabilities out of any state i add up to one, i.e. $\sum_j p_{ij} = 1$. When these conditions are satisfied for all p_{ij}, we call the transition probabilities *valid* and we can think of the probabilities as a percentage chance of a transition event happening. For example, in figure 1, if we are currently in state 1 and $p_{11} = 0.4$ and $p_{12} = 0.6$, we can say that at the next time step there is a 40% chance that we stay in state 1 and there is a 60% chance that we transition from state 1 to state 2. If any of the probabilities were negative or greater than 1, this interpretation wouldn't make sense; having a negative percentage chance or a percentage chance over 100% doesn't have a real meaning.

We can collect the transition probabilities into a *transition matrix*, P, where the elements of P are simply the transition probabilities, i.e. $P_{ij} = p_{ij}$. When the transition probabilities are valid, we say that the matrix P is *stochastic*. A stochastic matrix is one whose entries are nonnegative and whose row sums are one, which is the same as saying that a stochastic matrix is a matrix whose elements correspond to valid transition probabilities. An example of such a matrix would be

$$\begin{bmatrix} 0.3 & 0 & 0.4 & 0.3 \\ 0 & 0.1 & 0.2 & 0.7 \\ 0.2 & 0.2 & 0.2 & 0.4 \\ 0 & 0 & 0 & 1 \end{bmatrix};$$

we can verify that this matrix has nonnegative entries and rows that sum to one by inspection.

1.3 State Probability Update

To complete our probabilistic model, our states also have probabilities associated with them and an update rule similar to the iterative algorithms we have studied previously. We will define the probability of being in state i at time k as $s_i(k)$. Again we require that $0 \leq s_i(k) \leq 1$ and that $\sum_i s_i(k) = 1$ so that the state probabilities are valid at every time k. Since we have valid probabilities, we can interpret them as a percentage chance of being in each state at a given time. For example, if we were modelling weather and the state probabilities were $s_1(10) = 0.7$, $s_2(10) = 0.3$, where state 1 was sunny and state 2 was raining,

we would say there is a 70% chance that it is sunny at time $k = 10$ and a 30% chance that it is raining.

We can now use the state probabilities and the transition probabilities to define the update

$$s_j(k+1) = \sum_i p_{ij} s_i(k). \tag{1}$$

In the setting of our two state example, we would have $s_1(k+1) = p_{11}s_1(k) + p_{21}s_2(k)$, meaning that the probability of being in state s_1 at time $k+1$ is the probability that the system was in state s_1 at time k and stays there, plus the probability that the system was in state s_2 at time k and transitions from state s_2 to state s_1.

If the transition probabilities are valid and the state probabilities at time k are valid, then the state probabilities at time $k+1$ will also be valid. Since the update is a sum of nonnegative values, we can be sure that the result is also nonnegative. If we sum the state probabilities at time $k+1$ we get

$$\begin{aligned}
\sum_j s_j(k+1) &= \sum_j \sum_i p_{ij} s_i(k) && (2)\\
&= \sum_i s_i(k) \sum_j p_{ij} && (3)\\
&= \sum_i s_i(k) && (4)\\
&= 1 && (5)
\end{aligned}$$

and thus the state probabilities at time $k+1$ are valid.

If we collect the state probabilities into a vector such that

$$\boldsymbol{s}(k) = [s_1(k), s_2(k), ..., s_n(k)]^\intercal,$$

we can redefine equation (1) using the transition matrix as[1]

$$\boldsymbol{s}(k+1) = P^\intercal \boldsymbol{s}(k). \tag{6}$$

The Markov chain update equation in (6) makes the memorylessness property of Markov chains readily apparent; $\boldsymbol{s}(k+1)$ only depends on the current states $\boldsymbol{s}(k)$ and not any of the previous states such as $\boldsymbol{s}(k-1)$.

1.4 Steady State Behaviour

One of the advantages of using Markov chains is the ability to predict how the system will behave in the long run. If a system has valid transition probabilities (and the Markov chain is well behaved), the system will always converge to a set

[1] In the analysis of Markov chains, it is convention to write (6) as $\boldsymbol{s}^\intercal(k+1) = \boldsymbol{s}^\intercal(k)P$. We have chosen to keep it in terms of $\boldsymbol{s}$ to make the appearance of equations more similar to what we have already done in class.

of steady state probabilities for each state. Meaning, if we wait a long enough time, the probability that we are in state i stops changing with time.

Conveniently, we can compute these steady state probabilities using only the transition probabilities. To do so, we use the fact that in steady state $\boldsymbol{s}(k+1) = \boldsymbol{s}(k)$. If we call this steady state value $\boldsymbol{s}^\star$ and plug it into (6) we get

$$\boldsymbol{s}^\star = P^\intercal \boldsymbol{s}^\star \tag{7}$$

$$(I - P^\intercal)\boldsymbol{s}^\star = \mathbf{0} \tag{8}$$

Equation (8) is now in the form $Ax = b$ where $A = (I - P^\intercal)$ and $b = \mathbf{0}$. We can use row reduction to solve this, which we will see in the following examples.

1.5 Illustrative Examples

1.5.1 Finding Steady State Probabilities

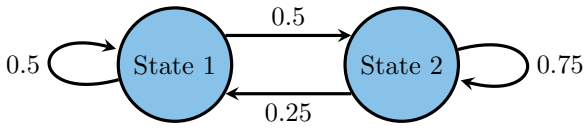


Figure 2: A two state Markov chain.

Figure 2 shows a basic two state Markov chain. If we collect the edge weights, or transition probabilities, into a matrix we get

$$P = \begin{bmatrix} 0.5 & 0.5 \\ 0.25 & 0.75 \end{bmatrix}$$

We can then solve the equation $(I - P^\intercal)\boldsymbol{s}^\star = \mathbf{0}$ in order to determine the steady state probabilities of each state. To solve this equation, we put the augmented matrix $[I - P^\intercal \,|\, \mathbf{0}]$ in reduced row echelon form. In this example the matrix $[I - P^\intercal \,|\, \mathbf{0}]$ is

$$\left[\begin{array}{cc|c} 0.5 & -0.25 & 0 \\ -0.5 & 0.25 & 0 \end{array}\right] \tag{9}$$

After performing row reduction we get,

$$\left[\begin{array}{cc|c} 1 & -0.5 & 0 \\ 0 & 0 & 0 \end{array}\right] \tag{10}$$

Since there is 1 pivot, $s_2^\star$ is a free variable and $s_1^\star$ is basic. Since we have one free variable, the solutions form a line given by the equations

$$\begin{cases} s_1^\star = 0.5t \\ s_2^\star = t \end{cases} \tag{11}$$

There are infinite solutions to the equation for the steady state values but we are only interested in the solution that corresponds to valid probabilities, i.e. $s_1^\star + s_2^\star = 1$ with both values nonnegative. If we plug in equation (11), we find $t = \frac{2}{3}$ which implies that

$$\begin{cases} s_1^\star = \frac{1}{3} \\ s_2^\star = \frac{2}{3} \end{cases} . \tag{12}$$

In words, this means that, regardless of our initial probabilities, after a long time we have a $\frac{1}{3}$ chance of being in state 1 and a $\frac{2}{3}$ chance of being in state 2. This can be confirmed by running the update rule (6) and seeing that the state probabilities converge to the values calculated above. Figure 3 shows that we reach the predicted steady state probabilities after only 5 iterations when starting from the initial probabilities $s_1(0) = 0.9$ and $s_2(0) = 0.1$. Being an example, these initial probabilities are arbitrary and chosen purely for illustration; initial probabilities are typically determined by the application or system under study.

1.5.2 A More Compact Approach

The requirement that $\sum_i s_i^\star = 1$ is just another linear equation that we can add to the system $(I - P^\intercal)\boldsymbol{s}^\star = \mathbf{0}$ then solve in a compact fashion. If we combine these equations for the previous example, instead of Eq (9), we get

$$\left[\begin{array}{cc|c} 0.5 & -0.25 & 0 \\ -0.5 & 0.25 & 0 \\ 1 & 1 & 1 \end{array}\right] . \tag{13}$$

After performing row reduction we get,

$$\left[\begin{array}{cc|c} 1 & 0 & \frac{1}{3} \\ 0 & 1 & \frac{2}{3} \\ 0 & 0 & 0 \end{array}\right] , \tag{14}$$

which highlights the fact that we have two pivots meaning that we no longer have a free variable to determine. Thus the unique solution is determined to be

$$\begin{cases} s_1^\star = \frac{1}{3} \\ s_2^\star = \frac{2}{3} \end{cases} , \tag{15}$$

which is the same solution found in the last section. Although this approach and that used in the previous section both yield the same solution, when solved by hand the compact approach can be unwieldy for larger matrices.

1.5.3 Absorbing States

Figure 4 shows a special case of a Markov chain with an *absorbing state*; a state that, once the system enters it, the system cannot leave. Notice that state 2

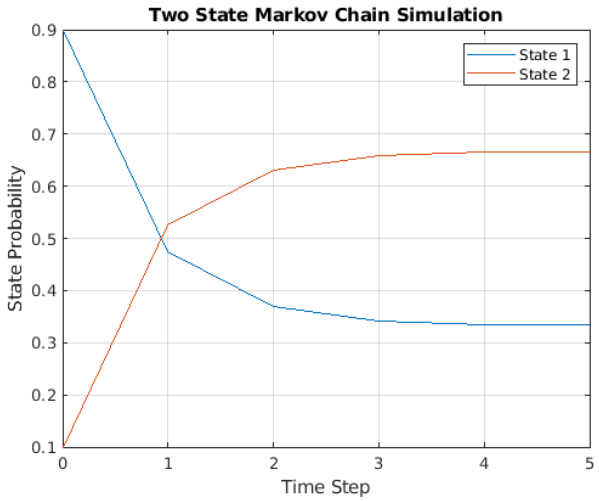


Figure 3: A simulation of the Markov chain shown in figure 2 with $s_1(0) = 0.9$ and $s_2(0) = 0.1$.

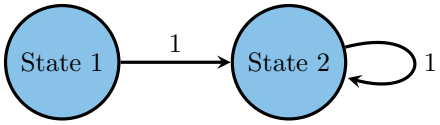


Figure 4: A Markov chain with an absorbing state.

only has an incoming link and a self loop, there is no way for the system to leave state 2. This chain will always end up with a 100% chance of being in state 2. We can confirm this by following the same steps as above. First we construct P and the matrix $[I - P^\intercal \,|\, \mathbf{0}]$:

$$P = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \tag{16}$$

$$[I - P^\intercal \,|\, \mathbf{0}] = \left[\begin{array}{cc|c} 1 & 0 & 0 \\ -1 & 0 & 0 \end{array}\right] \tag{17}$$

We then put $[I - P^\intercal \,|\, \mathbf{0}]$ in reduced row echelon form:

$$\left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right] \tag{18}$$

As in the previous example, we have 1 pivot and $s_2^\star$ is a free variable. The

solutions form a line given by the equations

$$\begin{cases} s_1^\star = 0 \\ s_2^\star = t \end{cases} \tag{19}$$

In order to have valid probabilities, we pick $t = 1$ which shows that there will be a 100% chance of being in state 2 after a long enough time.

2 PageRank as a Random Walk

To refresh your memory, the rank of a page in the PageRank algorithm is given by the solution to

$$\mathbf{r} = (1-d)\mathbf{1} + dH\mathbf{r} \tag{20}$$

Where $\mathbf{r}$ are is the vector of ranks, $\mathbf{1}$ is the ones vector which is the same as `ones(n,1)` (where n is the number of pages in the network), d is the damping factor, and H encodes the network structure. Using this original formulation, it can be shown that the sum of the ranks is always equal to n. Therefore, the vector of ranks $\mathbf{r}$ is *unnormalized*. If we divide both sides of (20) by n and define $\boldsymbol{s} = \frac{1}{n}\mathbf{r}$, we get

$$\boldsymbol{s} = \frac{1-d}{n}\mathbf{1} + dH\boldsymbol{s} \tag{21}$$

Now the elements of $\boldsymbol{s}$ sum to one (they represent valid probabilities) and $\boldsymbol{s}$ represents the normalized ranks.

With the normalized ranks, we can view the PageRank Algorithm as a random walk: Imagine you are bored on the internet and flip a biased coin. The coin is heads with probability d and tails with probability $1-d$. Whenever the coin is heads, you click on a random outgoing link on the page you are currently on; whenever it is tails, you go to a random page (including the one you are already on). If you repeat this process for long enough, the probability that you are on any given page is its (normalized) rank given by $\boldsymbol{s}$.

2.1 The Transition Matrix and the Markov Chain

Taking this view, the transition probability from one site to another is

$$p_{ij} = \frac{1-d}{n} + dH_{ji}. \tag{22}$$

The first term is the case where the coin is tails with probability $1-d$. Since we go to a random page, the chance we arrive at page j is $\frac{1}{n}$. The second term is the case where the coin is heads with probability d. Here we follow the links on the current page and the chance we go to page j corresponds to the network structure encoded in H. Using (22) and (6), the update for the normalized ranks (i.e. the probability that we are on a given page) is then

$$\boldsymbol{s}(k+1) = P^\intercal \boldsymbol{s}(k) \tag{23}$$

Where

$$P^\intercal = \frac{1-d}{n}\mathbf{1}\mathbf{1}^\intercal + dH, \tag{24}$$

and $\mathbf{1}\mathbf{1}^\intercal$ denotes the ones matrix which is the same as `ones(n,n)`.

Since all of the p_{ij} are nonnegative and for every website i

$$\sum_j p_{ij} = 1 - d + d\sum_j H_{ji} = 1, \tag{25}$$

the transition probabilities are valid and the transition matrix, P, is stochastic.

From here we can see that this formulation is the same as that introduced in equation (2) of the PageRank case study (equation (20) here).

$$\begin{aligned}
\boldsymbol{s}(k+1) &= P^\intercal \boldsymbol{s}(k) && (26)\\
&= \frac{1-d}{n}\mathbf{1}\mathbf{1}^\intercal \boldsymbol{s}(k) + dH\boldsymbol{s}(k) && (27)\\
&= \frac{1-d}{n}\mathbf{1} + dH\boldsymbol{s}(t) && (28)\\
\frac{1}{n}\mathbf{r}(t+1) &= \frac{1-d}{n}\mathbf{1} + \frac{d}{n}H\mathbf{r}(t) && (29)\\
\mathbf{r}(t+1) &= (1-d)\mathbf{1} + dH\mathbf{r}(t) && (30)
\end{aligned}$$

Since we have defined the page rank in terms of the steady state probabilities, we must solve (8) with the above P to recover the ranks.

2.2 Example

Here we will use the same example from the PageRank case study. Using $d = 0.9$, we have

$$H = \begin{bmatrix} 0 & 0 & 1/2 & 1/3 & 1/2 \\ 1/2 & 0 & 0 & 1/3 & 1/2 \\ 0 & 1 & 0 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{31}$$

$$P = \begin{bmatrix} 0.02 & 0.47 & 0.02 & 0.02 & 0.47 \\ 0.02 & 0.02 & 0.92 & 0.02 & 0.02 \\ 0.47 & 0.02 & 0.02 & 0.47 & 0.02 \\ 0.32 & 0.32 & 0.32 & 0.02 & 0.02 \\ 0.47 & 0.47 & 0.02 & 0.02 & 0.02 \end{bmatrix} \tag{32}$$

Figure 5 depicts the Markov chain corresponding to the transition matrix given by equation (32). Notice that all the elements of P are nonnegative and that its row sums are 1; thus P is stochastic and the transition probabilities are valid. Also notice that this graph is somewhat different than the underlying internet topology shown in the PageRank case study; a damping factor, $d < 1$ adds extra links to the graph connecting every node to every other node and

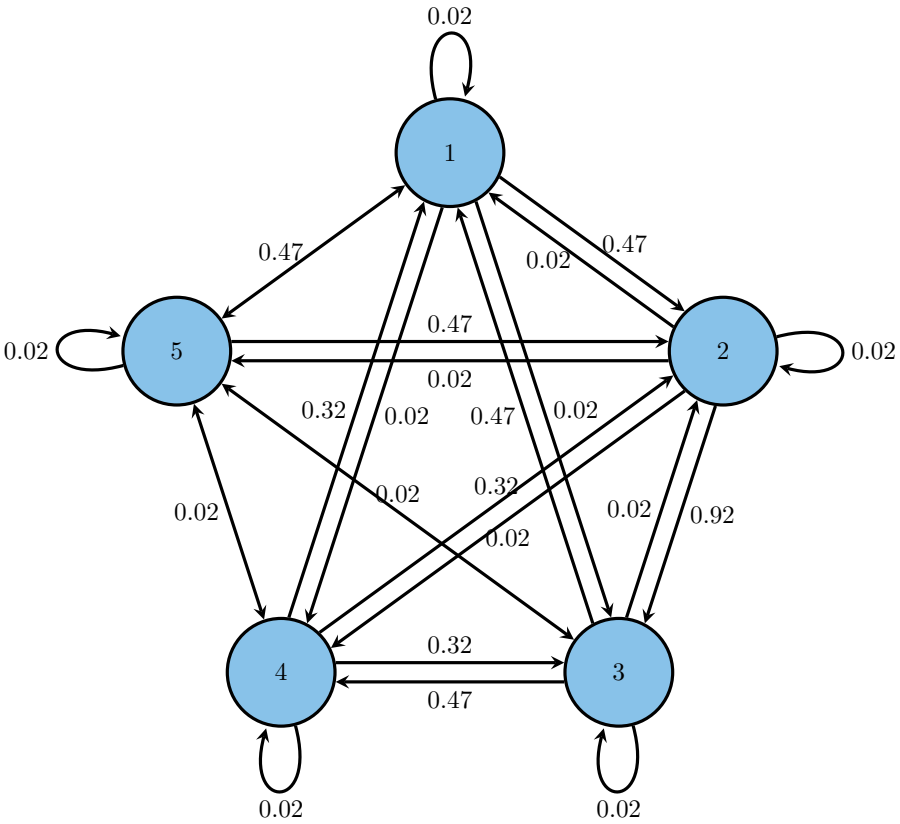


Figure 5: Markov chain for a PageRank based random walk on the internet.

introduces self-loops (i.e. there is a chance to stay on a page). The damping factor ensures that the Markov chain will be well behaved and that there will be no absorbing states, thus giving every page in the network a meaningful rank. Now we wish to solve for the steady state ranks, which we do so by following the same procedure from the earlier examples. First we write down the augmented matrix corresponding to $[I - P^\intercal \,|\, 0]$,

$$\left[\begin{array}{ccccc|c} 0.98 & -0.02 & -0.47 & -0.32 & -0.47 & 0 \\ -0.47 & 0.98 & -0.02 & -0.32 & -0.47 & 0 \\ -0.02 & -0.92 & 0.98 & -0.32 & -0.02 & 0 \\ -0.02 & -0.02 & -0.47 & 0.98 & -0.02 & 0 \\ -0.47 & -0.02 & -0.02 & -0.02 & 0.98 & 0 \end{array}\right] \tag{33}$$

After performing row reduction we get,

$$\left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & -1.8740 & 0 \\ 0 & 1 & 0 & 0 & -1.7786 & 0 \\ 0 & 0 & 1 & 0 & -2.0861 & 0 \\ 0 & 0 & 0 & 1 & -1.0954 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \tag{34}$$

Since there are 4 pivots, $s_5^\star$ is a free variable and all other variables are basic. Since we have one free variable, the solutions form a line given by the equations

$$\begin{cases} s_1^\star = 1.8740t \\ s_2^\star = 1.7786t \\ s_3^\star = 2.0861t \\ s_4^\star = 1.0954t \\ s_5^\star = t \end{cases} \tag{35}$$

Thus our steady state ranks are given by

$$\boldsymbol{s}^\star = [1.8740t\ 1.7786t\ 2.0861t\ 1.0954t\ t]^\intercal,$$

where t is our free parameter. Since we require that the elements of $\boldsymbol{s}^\star$ sum to one, we can choose t accordingly. Doing so yields $t = 0.1276$ and $\boldsymbol{s}^\star = [0.2392\ 0.2270\ 0.2663\ 0.1398\ 0.1276]^\intercal$. Upon normalizing the answer given in the PageRank case study, we verify that we have arrived at the same answer.

3 Nonlinear Models and Epidemic Spreading

Markov chains are a very broad tool and are not limited to the linear case where we can find a constant matrix of transition probabilities. Sometimes the transition probabilities depend on states of the system. We call these situations *nonlinear*. Take for example the chemical reaction $CH_4 + 2O_2 \to CO_2 + 2H_2O$, where methane reacts with oxygen to produce carbon dioxide and water. The transition probability or reaction rate from the products to the reactions depends nonlinearly on the concentrations of the reactants. In this case the rate is proportional to $[CH_4][O_2]^2$, where $[O_2]$ is the concentration of oxygen in the system. Markov Chains are still used to study chemical reaction networks, but a whole host of nonlinear techniques are used to analyze them.

A simple yet nonlinear model that we can analyze with Markov chains is a population model for epidemic spreading (remember that whole COVID-19 thing?). The model we will look at is the Susceptible-Infected-Recovered-Susceptible (SIRS) model and its simplifications.

3.1 SI Model

Possibly the most basic epidemic spreading model is the Susceptible-Infected model. There are two states, Susceptible, and Infected, and the probability of

transitioning from one state to the other naturally depends on the percentage of the population in each state. This model is typically posed as a set of differential equations but we have discretized them into iterative update equations like we have seen previously (see the optional derivation at the end of this section.)

$$\begin{aligned} S(k+1) &= -\beta S(k)I(k) + S(k) \\ I(k+1) &= \beta S(k)I(k) + I(k) \\ 1 &= S(k) + I(k) \end{aligned} \tag{36}$$

The corresponding chain is shown in figure 6. The transition probabilities depend on the product of susceptible and infected populations and captures the idea that if there is a small infected population or a small susceptible population, then the rate of new infections will be low. The rate parameter β captures how infectious the disease is and how easily it spreads.

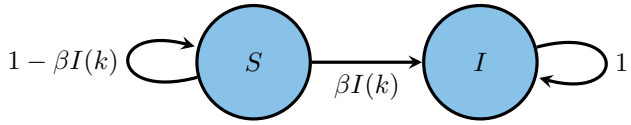


Figure 6: Markov Chain for the SI model with parameter β.

Since the transition probabilities depend on the product of states, this Markov chain is non-linear and we cannot represent it with a constant matrix. But we can still think of it similarly by introducing a matrix that depends on the states of the system:

$$P(I(k)) = \begin{bmatrix} -\beta I(k) + 1 & \beta I(k) \\ 0 & 1 \end{bmatrix}. \tag{37}$$

$P(I(k))$ is still stochastic for all valid values of $I(k)$ as long as $\beta \leq 1$. This matrix can also be made to depend only on $S(k)$, since $I(k) = 1 - S(k)$.

Even though we lack the mathematical tools to rigorously analyze this model, we can still tell from the Markov chain that "Infected" is an absorbing state since it has no outgoing link back to "Susceptible". The absorbing nature of the "Infected" state is a critical failing of the SI model. There is no way for the infected population to recover; once they are infected, they are infected forever. An easy remedy to this is to introduce a transition probability from infected back to susceptible.

3.2 SIS Model

The SIS model is the straightforward extension to the SI model. By adding a transition from "Infected" back to "Susceptible" that corresponds to recovery without immunity, we can achieve a more interesting equilibrium where some fraction of the population is still susceptible and not everyone is infected. The new equations are

$$\begin{aligned} S(k+1) &= -\beta S(k)I(k) + \gamma I(k) + S(k) \\ I(k+1) &= \beta S(k)I(k) - \gamma I(k) + I(k) \\ 1 &= S(k) + I(k) \end{aligned} \tag{38}$$

Where β is still the "infectiousness" of the epidemic being modelled and γ is the "recovery rate".

$$P(I(k)) = \begin{bmatrix} -\beta I(k) + 1 & \beta I(k) \\ \gamma & 1 - \gamma \end{bmatrix} \tag{39}$$

Again, $P(I(k))$ is still stochastic for all valid values of $I(k)$ as long as $\beta \leq 1$ and $\gamma \leq 1$. This Markov chain is depicted in figure 7.

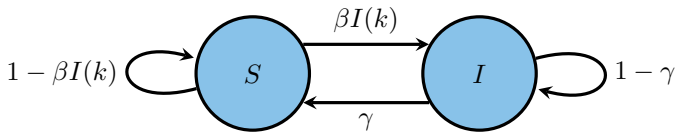


Figure 7: Markov Chain for the SIS model with parameters β and γ.

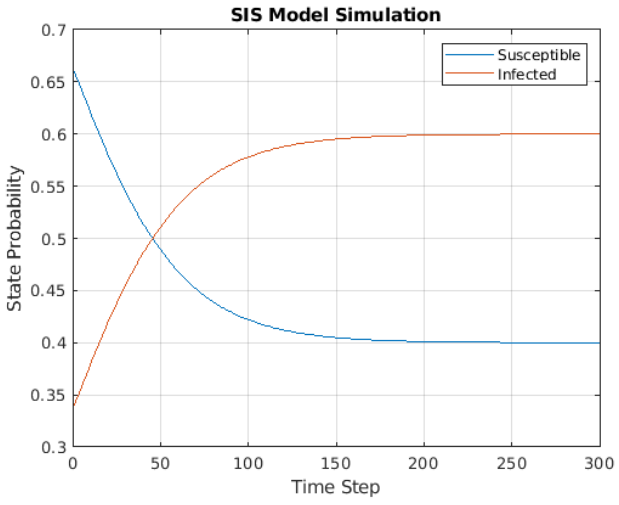
Even though we do not have the tools to analyze this Markov chain and the long term behaviour is no longer immediately evident, we can still simulate the equations given by (38) to determine both the transient behavior and the long term behavior. As the COVID-19 pandemic has demonstrated, the transient behavior of a system can be as important or even more important than the long term behavior (see efforts to "flatten the curve"). Figure 8 shows the result of simulating an SIS Markov chain for some parameter choices. Even though the system is nonlinear, it still gives us behaviour that is remarkably similar to the linear two state example we simulated earlier.

While the SIS model is much more useful than the SI model, it still does not capture the behaviour of a real disease; in a real disease, people who recover typically have immunity from the disease for some period of time. We can rectify this by adding another state.

3.3 SIRS Model

The SIRS Model stands for Susceptible-Infected-Recovered-Susceptible and accounts for immunity by adding a new state, recovered. The equations are

$$\begin{aligned} S(k+1) &= -\beta S(k)I(k) + \xi R(k) + S(k) \\ I(k+1) &= \beta S(k)I(k) - \gamma I(k) + I(k) \\ R(k+1) &= \gamma I(k) - \xi R(k) + R(k) \\ 1 &= S(k) + I(k) + R(k) \end{aligned} \tag{40}$$



[h]

Figure 8: An SIS epidemic using $\beta = 0.05$ and $\gamma = 0.02$ with initial conditions $S(0) = 0.66$ and $I(0) = 0.34$.

The new parameter ξ can be thought of as how quickly a virus mutates and renders prior immunity useless.

While this is a much more useful model than SI or SIS, the SIRS model is not the end of the story. For a given epidemic, other quantities might be of interest, such as deaths and death rate. Additionally, there are dynamics not captured here: What happens when the rate parameters vary with time? What if we model interactions in social networks instead of mixing all of the "susceptible" with all of the "infected"? What about vaccines and social distancing mandates? These kind of inquiries motivate the extension of these models in different directions but the basic principles and analysis remain largely the same.

3.4 (Optional) Discretizing the SI Model

The SI model is usually used in continuous time and represented as a differential equation given by

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI \\ 1 &= S + I \end{aligned}$$

In order to discretize a differential equation, we take inspiration from the definition of the derivative

$$\frac{dS}{dt} = \lim_{\Delta t \to \infty} \frac{S(t + \Delta t) - S(t)}{\Delta t}$$

If we ignore the limit and pick a small value of Δt, we get approximate equality, we can then say

$$\begin{aligned} \frac{dS}{dt} &\approx \frac{S(t+\Delta t) - S(t)}{\Delta t} \\ \frac{S(t+\Delta t) - S(t)}{\Delta t} &\approx -\beta S(t) I(t) \\ S(t+\Delta t) &= -\beta \Delta t S(t) I(t) + S(t) \end{aligned}$$

If we choose new variables $k = \frac{t}{\Delta t}$, $\beta' = \beta \Delta t$ and assign a new function $S(k) = S(k\Delta t)$ then we can rewrite the above as

$$S(k+1) = \beta' S(k) I(k) + S(k)$$

The derivation for I(k) is the same. As with the definition of the derivative, the smaller Δt is, the more accurate the discritization.

4 Mendelian Genetics as a Markov Chain (Optional)

Mendelian Genetics is a basic theory of genetics proposed by Gregor Mendel in 1865. The tenets of Mendelian Genetics are

1. Some alleles are dominant during expression, and some are recessive; an organism with at least one dominant allele will display that trait.
2. Each parent passes down one of its two alleles randomly.
3. Genes of different traits are independent of each other during inheritance.

As a quick refresher on biology, organisms have two copies of every allele and Mendel hypothesized that one copy gets passed down randomly during reproduction. He further hypothesized that some alleles are dominant and that others are recessive, where dominant alleles are expressed over recessive alleles. An organism is called a hybrid or heterozygous when it has both a dominant and recessive allele; otherwise it is called homozygous dominant or homozygous recessive if it has two copies of the dominant or recessive gene respectively. Both homozygous dominant and heterozygous organisms express the dominant allele. A tool for assessing the potential genetic makeup of the offspring of two parents is a Punnett square. Figure 9 shows a Punnett square for two hybrid pea flowers being mated. As one can see, there is a 25% chance that a homozygous dominant offspring is produced, a 50% chance another hybrid is produced, and a 25% chance a homozygous recessive offspring is produced.

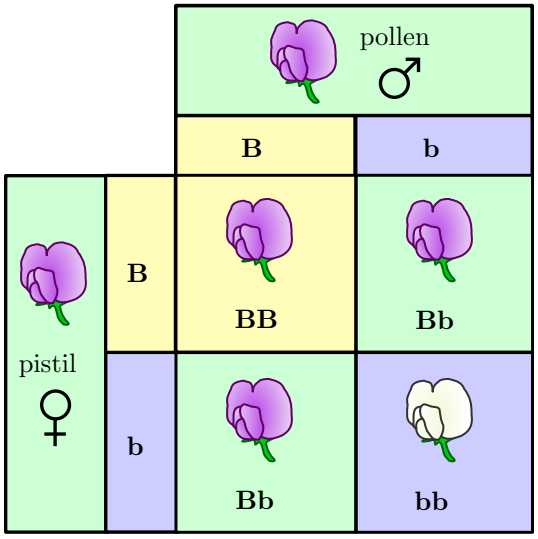


Figure 9: A Punnett square for the color of a pea plant flower. Courtesy of Wikimedia Commons.

4.1 Problem Setup and Details

We begin by mating two hybrids (Bb) together, we then take the offspring and mate it with another hybrid, and repeat. Thus at every step we are mating an offspring with a hybrid and looking at the genetic makeup of the next generation.

Let's derive the transition probabilities for each of the 3 cases:

4.1.1 Current Generation is Homozygous Dominant

The process of deriving the transition probabilities is very simple. First create the Punnett square for the current generation mated with a hybrid, then read off the probabilities of the offspring.

	B	b
B	BB	Bb
B	BB	Bb

Our probabilities are then $p_{dd} = 0.5, p_{dh} = 0.5, p_{dr} = 0$, where p_{dd} is the probability that mating a dominant with a hybrid yields another dominant and likewise for the other notations.

4.1.2 Current Generation is Hybrid

	B	b
B	BB	Bb
b	Bb	bb

Our probabilities are then $p_{hd} = 0.25, p_{hh} = 0.5, p_{hr} = 0.25$.

4.1.3 Current Generation is Homozygous Recessive

	B	b
b	Bb	bb
b	Bb	bb

Our probabilities are then $p_{rd} = 0, p_{rh} = 0.5, p_{rr} = 0.5$.

4.1.4 Markov Chain

If we collect the transition probabilities from the previous sections into a transition matrix, we have

$$P = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 0.5 & 0.5 \end{bmatrix} \tag{41}$$

This matrix is stochastic, and the corresponding Markov chain is shown in figure 10.

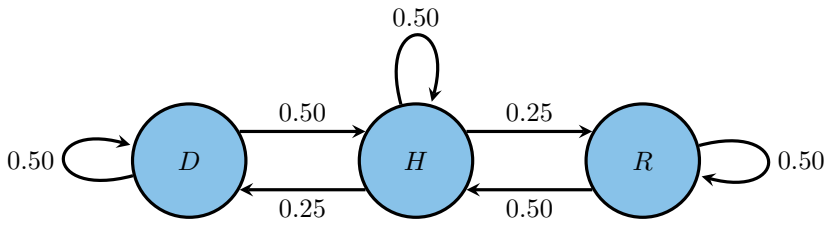


Figure 10: Markov Chain for breeding with a hybrid

If we find the reduced row echelon form of the augmented matrix $[I - P^{\intercal}|0]$ we arrive at

$$\left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \tag{42}$$

The solution is then given by a line defined by the parametric equations

$$\begin{cases} s_1^\star = t \\ s_2^\star = 2t \\ s_3^\star = t \end{cases} \tag{43}$$

or, equivalently, the vector $\boldsymbol{s}^\star = [t\ 2t\ t]^\intercal$. Since we are interested in probabilities, we choose t such that the vector sums to one and we arrive at $\boldsymbol{s}^\star = [0.25\ 0.5\ 0.25]^\intercal$. Thus, after a long time, we can expect the offspring to be dominant with probability 1/4, hybrid with probability 1/2, and recessive with probability 1/4.

4.1.5 Mating with Homozygous Dominant

What if at every step we mated with a Homozygous Dominant individual at every step instead of a hybrid? We get a transition matrix and Markov chain that look like

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 0.5 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{44}$$

From the Markov chain alone, it is easy to see that if we wait long enough, we

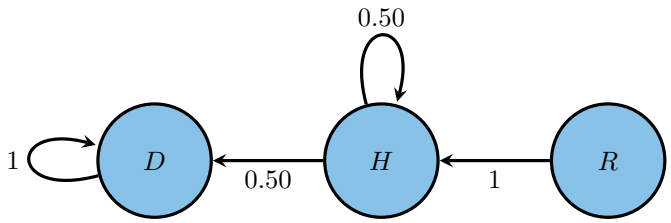


Figure 11: Markov Chain for breeding with a dominant

will end up with only dominant offspring. To verify this, we can compute the steady state population distribution the same way as above. Doing so yields $\boldsymbol{s}^\star = [1\ 0\ 0]^\intercal$