Stats 102C - Lecture 5-2: Properties of Markov Chains

Miles Chen, acknowledgements Michael Tsiang

Week 5 Wednesday

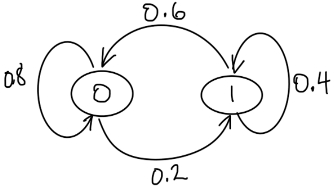
Section 1

Limiting Distributions

Two-state Markov Chain example

Let's say we have a two-state Markov chain, with state space $\{0, 1\}$ and transition matrix

$$\mathbb{P} = \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix}$$



Let's see what happens if we start the chain with X at time 0, $X_0 = 0$.

We'll let it run for a long time (n = 1000).

A nice visualization from “Explained Visually”

https://setosa.io/markov/index.html#%7B%22tm%22%3A%5B%5B0.8%2C0.2%5D%2C%5B0.6%2C0.4%5D%5D%7D


```r
P = rbind(c(0.8, 0.2),
          c(0.6, 0.4))
n <- 10^5
X <- c(0, rep(NA, n-1))

set.seed(1)

for(i in 2:n){
  # specify which row of P to use for the probability
  # If the previous x = 0, use row 1
  # If the previous x = 1, use row 2
  row <- X[i - 1] + 1

  X[i] <- sample(c(0, 1), size = 1, prob = P[row, ])
}
```

```r
# Proportion of steps in state 0
sum(X == 0) / n
```

```
## [1] 0.7497
```

```r
# Proportion of steps in state 1
sum(X == 1) / n
```

```
## [1] 0.2503
```

Let's try it again with a different starting state

```r
n <- 10^5
X <- c(1, rep(NA, n-1))
set.seed(2)
for(i in 2:n){
  row <- X[i - 1] + 1
  X[i] <- sample(c(0, 1), size = 1, prob = P[row, ])
}
# Proportion of steps in state 0
sum(X == 0) / n
```

```
## [1] 0.74685
```

```r
# Proportion of steps in state 1
sum(X == 1) / n
```

```
## [1] 0.25315
```

Results

Both simulations end up with approximately 75% in state 0 and 25% in state 1.

- Let π_0 be the long-run relative frequency the chain is at state 0
- Let π_1 be the long-run relative frequency the chain is at state 1
- Then $\boldsymbol{\pi} = (\pi_0, \pi_1)$ is a distribution on the state space {0, 1}

From our simulations, we may guess that $\boldsymbol{\pi} = (\pi_0, \pi_1) = (0.75, 0.25)$

The limiting distribution

- The fraction of time in state j is π_j
  - If π_0 is 0.75, then the Markov chain is in state 0 about 75% of the time.
- The probability that X_n is in state j is π_j
  - $\lim_{n\to\infty} \Pr(X_n = 0 | X_0 = 0) = \pi_0 = 0.75$
  - $\lim_{n\to\infty} \Pr(X_n = 0 | X_0 = 1) = \pi_0 = 0.75$
  - The distribution is independent of the initial state X_0

The limiting distribution

Definition:

Let $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2, \ldots, \pi_N)$ be a probability distribution on the state space $\{0, 1, 2, \ldots, N\}$.

We say that $\boldsymbol{\pi}$ is the **limiting distribution** of a Markov chain $\{X_0, X_1, X_2, \ldots\}$ if

$$\lim_{n\to\infty} \Pr(X_n = j | X_0 = i) = \pi_j$$

for all $i, j \in \{0, 1, 2, \ldots, N\}$

If the limiting distribution $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2, \ldots, \pi_N)$ exists, then π_j represents

- The probability that X_n is in state j, independent of the initial state X_0.
- The long run expected fraction of time that the Markov chain spends in state j.

The stationary distribution

The **stationary distribution** is a distribution over states, such that the expected distribution of the Markov chain at the next time is the same. The stationary distribution is the distribution of the system at equilibrium.

If the limiting distribution of a Markov chain exists, it will be the stationary distribution.

Theorem: If a Markov chain is **ergodic** it will eventually converge to a stationary distribution.

Section 2

Informal definitions of Markov Chains

Some Informal definitions

A Markov Chain is said to be **ergodic** if:

- we can start the chain in any state i and
- at some time in the future t
- the probability of being in any of the other states j is > 0

(This must be true for all t greater than some finite integer T_0.)

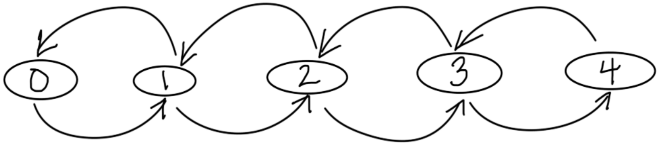
For a Markov Chain to be ergodic, it must be:

- irreducible and
- aperiodic

Irreducible Chains

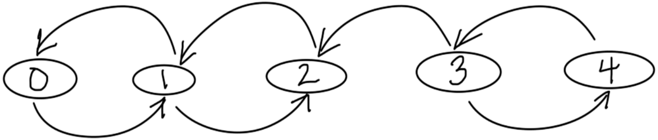
A Markov Chain is **irreducible** if all states *communicate* with each other. A Markov Chain is reducible if it is not irreducible.

Two states i and j **communicate** with each other if there is a path (even if travels through other states) to go from i to j and a path that goes from j to i.



This Markov Chain is irreducible. All of the states communicate with each other. If you start at any of the states, there is a path to any of the other states.

Reducible Chains

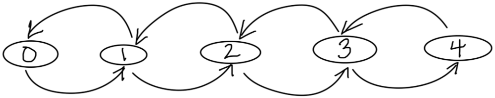


This Markov Chain reducible. All the states do not communicate with each other. If you start at state 0, 1, or 2, there is no path to get to states 3 or 4.

For any chain that begins in states 0, 1, or 2, the chain diagram can be reduced to just those three states.

State Periods

The **period** of a state is the greatest common divisor of n the number of steps you can take and end up back at the state itself.



Look at state 3. You can start at state 3 and take many paths and finish back at state 3. There are many possible options, each with a different number of steps:

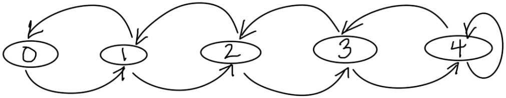
- $3 \to 2 \to 3$ (2 steps)
- $3 \to 4 \to 3 \to 2 \to 3$ (4 steps visiting the state before ending at the state is allowed)
- $3 \to 2 \to 1 \to 2 \to 3$ (4 steps)
- $3 \to 2 \to 1 \to 0 \to 1 \to 2 \to 3$ (6 steps)
- etc.

The greatest common divisor of the number of steps $\{2, 4, 6, 8, \ldots\}$ is 2. State 3 has a period of 2.

A further study will show that all states have a period of 2.

State Periods

We add a pathway for state 4 to repeat and return to itself.



We look at state 3 again.

- $3 \rightarrow 4 \rightarrow 3$ (2 steps)
- $3 \rightarrow 4 \rightarrow 4 \rightarrow 3$ (3 steps)
- $3 \rightarrow 2 \rightarrow 1 \rightarrow 2 \rightarrow 3$ (4 steps)
- $3 \rightarrow 4 \rightarrow 4 \rightarrow 4 \rightarrow 4 \rightarrow 3$ (5 steps)
- etc.

The greatest common divisor of the number of steps $\{2, 3, 4, 5, \ldots\}$ is 1. State 3 has a period of 1.

Because we can "abuse" visiting state 4 as many times as we want (even or odd), we can see that all of the states will have a period of 1.

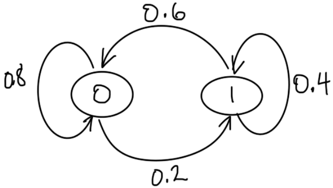
Aperiodic Markov Chains

A Markov Chain is **aperiodic** if all states have a period of 1.

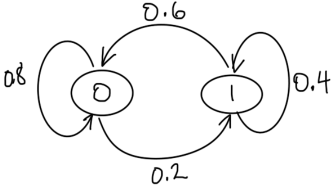
If a Markov Chain is irreducible and at least one state can go back to itself, the chain will be aperiodic.

Let's revisit our two-state Markov chain.

$$\mathbb{P} = \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix}$$



The stationary distribution will be achieved when there is equilibrium between states: the probability of transitioning from state 0 to state 1 is equal to the probability of transitioning from state 1 to state 0.



If you are already at state 0, the prob of going to state 1 is:

$$P_{01} = \Pr(x_{n+1} = 1 | x_n = 0) = 0.2$$

The total probability of transitioning from state 0 to state 1 is:

$$\Pr(X_n = 0) \cdot \Pr(x_{n+1} = 1 | x_n = 0) = p_0 P_{01}$$

If you are already at state 1, the prob of going to state 0 is:

$$P_{10} = \Pr(x_{n+1} = 0 | x_n = 1) = 0.6$$

The total probability of transitioning from state 1 to state 0 is:

$$\Pr(X_n = 1) \cdot \Pr(x_{n+1} = 0 | x_n = 1) = p_1 P_{10}$$

We want to find a distribution $\boldsymbol{\pi} = (\pi_0, \pi_1)$ that satisfies:

$$\pi_0 P_{01} = \pi_1 P_{10}$$

$$0.2\pi_0 = 0.6\pi_1$$

and $\pi_0 + \pi_1 = 1$

We find that $\frac{\pi_0}{\pi_1} = 3$, which yields the stationary distribution: $\pi_0 = 0.75$ and $\pi_1 = 0.25$

We found our stationary distribution.

$\boldsymbol{\pi} = [0.75, 0.25]$

We can see that when we apply the transition matrix to the stationary distribution, we get the stationary distribution back.

$$
\begin{aligned}
\boldsymbol{\pi} \times \mathbb{P} &= [0.75, 0.25] \times \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} \\
&= [0.75(0.8) + 0.25(0.6), 0.75(0.2) + 0.25(0.4)] \\
&= [0.6 + 0.15, 0.15 + 0.1] \\
&= [0.75, 0.25]
\end{aligned}
$$

Stationary distributions

The stationary distribution is the vector $\boldsymbol{\pi}$, so that it has the property:

$\boldsymbol{\pi}\mathbb{P} = \boldsymbol{\pi}$

The stationary distribution $\boldsymbol{\pi}$ is the left-eigenvector of the matrix $\mathbb{P}$.

(Normally, when we talk about eigenvectors of a matrix, we refer to the right eigenvector where the vector is multiplied on the right side of the matrix. The right eigenvector is the solution to $\mathbb{A}\mathbf{v} = \lambda\mathbf{v}$)

$$\begin{aligned}
\boldsymbol{\pi}\mathbb{P} &= \boldsymbol{\pi} \\
(\boldsymbol{\pi}\mathbb{P})^T &= \boldsymbol{\pi}^T \\
\mathbb{P}^T\boldsymbol{\pi}^T &= \boldsymbol{\pi}^T
\end{aligned}$$

The left eigenvector (a row vector) of $\mathbb{P}$ is equal to the transpose of right eigenvector of $\mathbb{P}^T$.

Function eigen() finds the right eigenvectors.

```r
print(P) # transition matrix P
```

```
##      [,1] [,2]
## [1,]  0.8  0.2
## [2,]  0.6  0.4
```

```r
eig_vectors <- eigen(t(P))$vectors # find the right eigenvectors of P transpose
eig_vectors # the first column is all positive, so we keep that
```

```
##           [,1]       [,2]
## [1,] 0.9486833 -0.7071068
## [2,] 0.3162278  0.7071068
```

```r
stationary <- t(eig_vectors[,1]) # left eigenvector is transpose of the right eigenvector
stationary / (sum(stationary)) # we normalize the vector, so the sum is 1
```

```
##      [,1] [,2]
## [1,] 0.75 0.25
```

Some nice visualizations of a stationary distribution

https://setosa.io/ev/eigenvectors-and-eigenvalues/

The limiting distribution

Multiplying the current distribution at time n $\boldsymbol{\pi}^{(n)}$ by the transition matrix will give you the probability distribution at the next step at time $n+1$, $\boldsymbol{\pi}^{(n+1)}$.

If you continue multiplying the the current distribution by the transition matrix $\mathbb{P}$ it will converge to the stationary distribution if the chain is ergodic.

We'll take a look at the following transition matrix.

$$\mathbb{P} = \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix}$$

The matrix has stationary distribution $\boldsymbol{\pi} = [0.75, 0.25]$.

We can see that $\boldsymbol{\pi}^{(n)} \to \boldsymbol{\pi}$ as $n \to \infty$

$$\boldsymbol{\pi}^{(n)}\mathbb{P} = [1, 0] \times \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} = [0.8, 0.2] = \boldsymbol{\pi}^{(n+1)}$$

$$[0.8, 0.2] \times \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} = [0.76, 0.24]$$

$$[0.76, 0.24] \times \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} = [0.752, 0.248]$$

$$\vdots \quad = \quad \vdots$$

```r
P <-  rbind(c(0.8, 0.2),
            c(0.6, 0.4))
x <- matrix(c(1, 0), nrow = 1)
x
```

```
##      [,1] [,2]
## [1,]    1    0
```

```r
x <- x %*% P; print(x)
```

```
##      [,1] [,2]
## [1,]  0.8  0.2
```

```r
x <- x %*% P; print(x)
```

```
##      [,1] [,2]
## [1,] 0.76 0.24
```

```r
x <- x %*% P; print(x)
```

```
##       [,1]  [,2]
## [1,] 0.752 0.248
```

```r
P <-  rbind(c(0.8, 0.2),
            c(0.6, 0.4))

x <- matrix(c(1, 0), nrow = 1)

for(i in 1:1000) { # apply the transition matrix 1000 times
  x <- x %*% P
}

x
```

```
##      [,1] [,2]
## [1,] 0.75 0.25
```

A different starting distribution

```
x <- matrix(c(0, 1), nrow = 1) # different starting distribution
x
```

```
##      [,1] [,2]
## [1,]    0    1
```

```
x <- x %*% P; print(x)
```

```
##      [,1] [,2]
## [1,]  0.6  0.4
```

```
x <- x %*% P; print(x)
```

```
##      [,1] [,2]
## [1,] 0.72 0.28
```

```
x <- x %*% P; print(x)
```

```
##       [,1]  [,2]
## [1,] 0.744 0.256
```

```r
x <- matrix(c(0, 1), nrow = 1) # different starting distribution

for(i in 1:1000) { # apply the transition matrix 1000 times
  x <- x %*% P
}

x
```

```
##      [,1] [,2]
## [1,] 0.75 0.25
```

Monopoly Flashbacks

Some of you took my class for Stats 102A and one of the homework assignments was to find the probability of landing on each space.

An alternative to writing code that follows the rules is to create a large transition matrix showing the probability of going from one space to another space.

There are 40 spaces on the board, so the transition probability matrix would be 40 x 40.

The first row would represent the probabilities of moving from "Go" to another space.

If you don't implementing any rules other than rolling dice, the probabilities would be:

$[0, 0, 1/36, 2/36, 3/36, 4/36, 5/36, 6/36, 5/36, 4/36, 3/36, 2/36, 1/36, 0, 0, 0, 0, \ldots]$

But in reality, the probailities are a bit different because it is possible to end up at Boardwalk if you land on the Chance space and draw the card that moves you to Boardwalk.

A few links:

http://carlabernard.ch/beni/downloads/bernard_monopoly.pdf

https://conf.math.illinois.edu/~bishop/monopoly.pdf

Next steps

What we saw today was that an ergodic Markov chain with transition matrix $\mathbb{P}$ has a stationary distribution $\boldsymbol{\pi}$

If we let the Markov chain run long enough, it converges to the stationary distribution, and from that point on, the values that it generates will come from the stationary distribution.

Next idea: Let's say there is some target probability distribution that we want to generate samples from. Can we design a Markov chain (i.e. figure out transition probabilities) so that the stationary distribution of the chain is equal to the desired target distribution?

If we can do this, then we can let the chain run for a long time and the values it generates will come from our desired distribution.