Stochastic Simulation

APPM 7400

Lesson 13: Markov Chains

(Discrete Time)
October 8, 2018

A Markov chain $\{X_n\}$ is a stochastic process with a sort of "limited memory":

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = j | X_n = i)$$

A Markov chain $\{X_n\}$ is a stochastic process with a sort of "limited memory":

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j | X_n = i)$$

Note:

- This does not mean that X_{n+1} is independent of the earlier $X_0, X_1, \dots, X_{n-1}$.

A Markov chain $\{X_n\}$ is a stochastic process with a sort of "limited memory":

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0)$$
$$= P(X_{n+1} = j | X_n = i)$$

Note:

- This does not mean that X_{n+1} is independent of the earlier $X_0, X_1, \dots, X_{n-1}$.
- Though we can say that X_{n+1} is "conditionally independent" of $X_0, X_1, \dots, X_{n-1}$, "given X_n".

Note that the Markov property not only implies that (for example)

$$P(X_3 = 4|X_2 = 0, X_1 = 2, X_0 = 8) = P(X_3 = 4|X_2 = 0)$$

Note that the Markov property not only implies that (for example)

$$P(X_3 = 4|X_2 = 0, X_1 = 2, X_0 = 8) = P(X_3 = 4|X_2 = 0)$$

but also that $P(X_3 = 4|X_1 = 2, X_0 = 8) = P(X_3 = 4|X_1 = 2)$

Markov Chains

Note that the Markov property not only implies that (for example)

$$P(X_3 = 4 | X_2 = 0, X_1 = 2, X_0 = 8) = P(X_3 = 4 | X_2 = 0)$$

but also that $P(X_3 = 4 | X_1 = 2, X_0 = 8) = P(X_3 = 4 | X_1 = 2)$

Since $P(X_3 = 4 | X_1 = 2, X_0 = 8)$

$$\begin{aligned}
&= \textstyle\sum_j P(X_3 = 4, X_2 = j | X_1 = 2, X_0 = 8) \\
&= \textstyle\sum_j P(X_3 = 4 | X_2 = j, X_1 = 2, X_0 = 8) \cdot P(X_2 = j | X_1 = 2, X_0 = 8) \\
&\stackrel{M.P.}{=} \textstyle\sum_j P(X_3 = 4 | X_2 = j) \cdot P(X_2 = j | X_1 = 2) \\
&\stackrel{M.P.}{=} \textstyle\sum_j P(X_3 = 4 | X_2 = j, X_1 = 2) \cdot P(X_2 = j | X_1 = 2) \\
&= \textstyle\sum_j P(X_3 = 4, X_2 = j, X_1 = 2) = P(X_3 = 4 | X_1 = 2)
\end{aligned}$$

Example 1:

Suppose that a warehouse stocks a certain item to satisfy a continuing demand. The stock is checked at times t_n, $n > 1$.

Example 1:

Suppose that a warehouse stocks a certain item to satisfy a continuing demand. The stock is checked at times t_n, $n > 1$.

At each checking time, if the stock is below some prescribed level a, then the stock is replenished up to some prescribed level b $(a < b)$, otherwise nothing is done.

Example 1:

Suppose that a warehouse stocks a certain item to satisfy a continuing demand. The stock is checked at times t_n, $n > 1$.

At each checking time, if the stock is below some prescribed level a, then the stock is replenished up to some prescribed level b $(a < b)$, otherwise nothing is done.

The demand for the item during the time interval $[t_{n-1}, t_n)$ is random.

Example 1 (continued):

For $n = 0, 1, \ldots$, let

$$X_n = \text{stock level just before time } t_n$$

Example 1 (continued):

For $n = 0, 1, \ldots$, let

$$X_n = \text{stock level just before time } t_n$$

Then $\{X_n\}$ is a discrete time stochastic process with (finite) "state space"

$$S = \{0, 1, 2, \ldots, b\}$$

Example 1 (continued):

For $n = 0, 1, \ldots$, let

$$X_n = \text{stock level just before time } t_n$$

Then $\{X_n\}$ is a discrete time stochastic process with (finite) "state space"

$$S = \{0, 1, 2, \ldots, b\}$$

Furthermore, $\{X_n\}_{n \geq 0}$ is a Markov chain.

Example 2:

Customers arrive at a taxi stand and a taxi arrives every 5 minutes. Assume that a single customer is served during each time period, if there are customers and that the taxi drives away empty if there are no customers.

Example 2:

Customers arrive at a taxi stand and a taxi arrives every 5 minutes. Assume that a single customer is served during each time period, if there are customers and that the taxi drives away empty if there are no customers.

Assume that the customers are arriving at random times, say

$$Y_n = \# \text{ arriving during time period } n$$

(Assumed to be iid and independent of the number of customers waiting.)

Example 2:

Customers arrive at a taxi stand and a taxi arrives every 5 minutes. Assume that a single customer is served during each time period, if there are customers and that the taxi drives away empty if there are no customers.

Assume that the customers are arriving at random times, say

$$Y_n = \# \text{ arriving during time period } n$$

(Assumed to be iid and independent of the number of customers waiting.)

Let

$$X_n = \# \text{ customers waiting at the start of time period } n$$

Example 2 (continued):

Then $\{X_n\}_{n\geq 0}$ is a Markov chain with state space $S = \{0, 1, 2, \ldots\}$

Proof:

$$X_{n+1} = \begin{cases} X_n - 1 + Y_n & , \quad \text{if } X_n \geq 1 \\ Y_n & , \quad \text{if } X_n = 0 \end{cases}$$

So, if $i \geq 1$,

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0)$$

$$= P(X_n - 1 + Y_n = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0)$$

$$= \quad P(Y_n = j - i_n + 1 | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0)$$

$$= \quad P(Y_n = j - i + 1)$$

by assumption of the independence of the $\{Y_n\}$ and $\{X_n\}$ processes.

Example 2 (continued):

Proceeding in the same way, we can get

$$P(X_{n+1} = j | X_n = i) = P(Y_n = j - i + 1)$$

So we see that

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = j | X_n = i)$$

Warning: It would not have been enough to only show that

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0)$$

depends only on i and j.

Note that, since the Y_n are iid,

$$P(X_{n+1} = j | X_n = i) = P(Y_n = j - i + 1)$$

does not depend on time.

Note that, since the Y_n are iid,

$$P(X_{n+1} = j | X_n = i) = P(Y_n = j - i + 1)$$

does not depend on time.

This is a time-homogeneous Markov chain.

In this case we can write

$$p_{ij} = P(X_{n+1} = j | X_n = i)$$

which is the same as

$$P(X_n = j | X_0 = i).$$

For a time-homogeneous Markov chain, say on $S = \{0, 1, 2, \ldots\}$, we can organize transition probabilities into a (one-step) transition probability matrix:

$$\mathcal{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & \cdots \end{array} \\ \begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ p_{20} & p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \end{array}$$

"stochastic matrix"

$$\sum_{j \in S} p_{ij} = 1$$

$$p_{ij} = P(X_{n+1} = j | X_n = i)$$

Example:

Suppose that items produced by a certain worker in a factory are classified as "defective" or "non-defective".

Further suppose that, due to trends in raw material quality, whether or not a particular item is defective depends, in part, on whether the previous item was defective.

Example (continued):

Let

$$X_n = \begin{cases} 0 & , \quad \text{if } n^{th} \text{ item is not defective} \\ 1 & , \quad \text{if } n^{th} \text{ item is defective} \end{cases}$$

Example (continued):

Let

$$X_n = \begin{cases} 0 & , \quad \text{if } n^{th} \text{ item is not defective} \\ 1 & , \quad \text{if } n^{th} \text{ item is defective} \end{cases}$$

Suppose that the probability transition matrix is

$$\mathcal{P} = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} \;\;0 & \;\;\;\;\;1 \end{array} \\ \left[\begin{array}{cc} 0.99 & 0.01 \\ 0.08 & 0.92 \end{array} \right] \end{array}$$

(This suggests that defective items tend to appear in bunches.)

Given the probability transition matrix $\mathcal{P}$, how do we find

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n)?$$

Finite Dimensional Distributions of DTMCs

Given the probability transition matrix $\mathcal{P}$, how do we find

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n)?$$

The transition matrix describes transitions but we still need a starting (initial) distribution:

Definition:

$$\pi_i = P(X_0 = i)$$

$$\left(\sum_{i \in S} \pi_i = 1\right)$$

Finite Dimensional Distributions of DTMCs

Claim: Let $\{X_n\}$ be a DTMC on a state space S with probability transition matrix $\mathcal{P} = [p_{ij}]$.

Then for any $i_0, i_1, \ldots, i_n \in S$,

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \pi_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}$$

Finite Dimensional Distributions of DTMCs

Claim: Let $\{X_n\}$ be a DTMC on a state space S with probability transition matrix $\mathcal{P} = [p_{ij}]$.

Then for any $i_0, i_1, \ldots, i_n \in S$,

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \pi_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}$$

Proof: $P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n)$

$$\begin{aligned}
&= P(X_n = i_n | X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \cdot P(X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \\
&= P(X_n = i_n | X_{n-1} = i_{n-1}) \cdot P(X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \\
&= p_{i_{n-1} i_n} \cdot P(X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \cdots = \\
&= p_{i_{n-1} i_n} \cdot p_{i_{n-2} i_{n-1}} \cdots p_{i_0 i_1} \cdot P(X_0 = i_0) = p_{i_{n-1} i_n} \cdot p_{i_{n-2} i_{n-1}} \cdots p_{i_0 i_1} \cdot \pi_{i_0}
\end{aligned}$$

What about the distribution of $X_{n_1}, X_{n_2}, \ldots, X_{n_k}$ for some $n_1 < n_2 < \cdots < n_k$?

Finite Dimensional Distributions of DTMCs

What about the distribution of $X_{n_1}, X_{n_2}, \ldots, X_{n_k}$ for some $n_1 < n_2 < \cdots < n_k$?

Just sum (integrate) out the ones you don't want.

What about the distribution of $X_{n_1}, X_{n_2}, \ldots, X_{n_k}$ for some $n_1 < n_2 < \cdots < n_k$?

Just sum (integrate) out the ones you don't want.

For the defective factory items example:

Suppose
$$\begin{aligned}\pi_0 &= 0.89\\ \pi_1 &= 0.11\end{aligned}$$

$$\mathcal{P} = \begin{bmatrix} 0.99 & 0.01 \\ 0.08 & 0.92 \end{bmatrix}$$

$$\begin{aligned}
&P(X_1 = 1, X_3 = 0) = \\
&P(X_0 = 0, X_1 = 1, X_2 = 0, X_3 = 0) \\
&+P(X_0 = 1, X_1 = 1, X_2 = 0, X_3 = 0) \\
&+P(X_0 = 0, X_1 = 1, X_2 = 1, X_3 = 0) \\
&+P(X_0 = 1, X_1 = 1, X_2 = 1, X_3 = 0)
\end{aligned}$$

So,

$$\begin{aligned} P(X_1 = 1, X_3 = 0) &= (0.89)(0.01)(0.08)(0.99) \\ &\quad +(0.11)(0.92)(0.08)(0.99) \\ &\quad +(0.89)(0.01)(0.92)(0.08) \\ &\quad +(0.11)(0.92)(0.92)(0.08) \\ &\approx 0.0168 \end{aligned}$$

n-Step Transition Probabilities

Before: $p_{ij} = P(X_{k+1} = j | X_k = i)$

Now define: $p_{ij}^{(n)} = P(X_{k+n} = j | X_{k=i})$

and write the n-step transition matrix:

$$\mathcal{P}^{(n)} = \begin{array}{c|c} & \begin{array}{cccc} 0 & 1 & 2 & \cdots \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \end{array} & \begin{bmatrix} p_{00}^{(n)} & p_{01}^{(n)} & p_{02}^{(n)} & \cdots \\ p_{10}^{(n)} & p_{11}^{(n)} & p_{12}^{(n)} & \cdots \\ p_{20}^{(n)} & p_{21}^{(n)} & p_{22}^{(n)} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \end{array}$$

(Still a time-homogeneous setting.)

n-Step Transition Probabilities

Question:

How does the n-step transition probability matrix relate to the one-step transition probability matrix?

n-Step Transition Probabilities

Question:

How does the n-step transition probability matrix relate to the one-step transition probability matrix?

Example: Return to Defective Factory Items

Let's find $P(X_2 = 1 | X_0 = 0)$.

n-Step Transition Probabilities

Question:

How does the n-step transition probability matrix relate to the one-step transition probability matrix?

Example: Return to Defective Factory Items

Let's find $P(X_2 = 1 | X_0 = 0)$.

Given we started at 0, we either went from

$$0 \to 0 \to 1$$

or

$$0 \to 1 \to 1$$

n-Step Transition Probabilities

$$0 \to 0 \to 1$$
$$0 \to 1 \to 1$$

These paths represent disjoint events.
So

$$P(X_2 = 1 | X_0 = 0) = p_{00}p_{01} + p_{01}p_{11}$$

$$0 \to 0 \to 1$$
$$0 \to 1 \to 1$$

These paths represent disjoint events.
So

$$P(X_2 = 1 | X_0 = 0) = p_{00} p_{01} + p_{01} p_{11}$$

Look familiar?

n-Step Transition Probabilities

$$0 \to 0 \to 1$$
$$0 \to 1 \to 1$$

These paths represent disjoint events.
So

$$P(X_2 = 1 | X_0 = 0) = p_{00}p_{01} + p_{01}p_{11}$$

Look familiar?

$$\begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} p_{00}p_{00} + p_{01}p_{10} & p_{00}p_{01} + p_{01}p_{11} \\ p_{00}p_{10} + p_{11}p_{10} & p_{10}p_{01} + p_{11}p_{11} \end{bmatrix}$$

n-Step Transition Probabilities

Theorem:

The n-step transition probability

$$p_{ij}^{(n)} = P(X_{k+n} = j | X_k = i)$$

is the ij^{th} entry of the n^{th} power of $\mathcal{P}$.

$$\mathcal{P}^{(n)} = \mathcal{P}^n$$

(Though $p_{ij}^{(n)} \neq p_{ij}^n$.)

To prove this, we need the

Chapman-Kolmogorov Equations

$$p_{ij}^{(m+n)} = \sum_{k \in S} p_{ik}^{(m)} \cdot p_{kj}^{(n)}$$

Finite Dimensional Distributions of DTMCs

To prove this, we need the

Chapman-Kolmogorov Equations

$$p_{ij}^{(m+n)} = \sum_{k \in S} p_{ik}^{(m)} \cdot p_{kj}^{(n)}$$

Then the theorem is proven since

$$p_{ij}^{(m+1)} = \sum_{k \in S} p_{ik}^{(m)} \cdot p_{kj}$$

is the ij^{th} entry of $\mathcal{P}^{(m)} \cdot \mathcal{P}$.

Proof of Chapman-Kolmogorov:

$$p_{ij}^{(m+n)} \quad = \quad P(X_{m+n} = j | X_0 = i)$$

Proof of Chapman-Kolmogorov:

$$
\begin{aligned}
p_{ij}^{(m+n)} &= P(X_{m+n} = j | X_0 = i) \\
&= \sum_{k \in S} P(X_{m+n} = j, X_n = k | X_0 = i)
\end{aligned}
$$

Proof of Chapman-Kolmogorov:

$$
\begin{aligned}
p_{ij}^{(m+n)} &= P(X_{m+n} = j | X_0 = i) \\
&= \sum_{k \in S} P(X_{m+n} = j, X_n = k | X_0 = i) \\
&= \sum_{k \in S} P(X_{m+n} = j | X_n = k, X_0 = i) \cdot P(X_n = k | X_0 = i)
\end{aligned}
$$

Proof of Chapman-Kolmogorov:

$$
\begin{aligned}
p_{ij}^{(m+n)} &= P(X_{m+n} = j | X_0 = i) \\
&= \sum_{k \in S} P(X_{m+n} = j, X_n = k | X_0 = i) \\
&= \sum_{k \in S} P(X_{m+n} = j | X_n = k, X_0 = i) \cdot P(X_n = k | X_0 = i) \\
&\overset{M.P.}{=} \sum_{k \in S} P(X_{m+n} = j | X_n = k) \cdot P(X_n = k | X_0 = i)
\end{aligned}
$$

Proof of Chapman-Kolmogorov:

$$
\begin{aligned}
p_{ij}^{(m+n)} &= P(X_{m+n} = j | X_0 = i) \\
&= \sum_{k \in S} P(X_{m+n} = j, X_n = k | X_0 = i) \\
&= \sum_{k \in S} P(X_{m+n} = j | X_n = k, X_0 = i) \cdot P(X_n = k | X_0 = i) \\
&\overset{M.P.}{=} \sum_{k \in S} P(X_{m+n} = j | X_n = k) \cdot P(X_n = k | X_0 = i) \\
&= \sum_{k \in S} p_{ik}^{(m)} \cdot p_{kj}^{(n)}
\end{aligned}
$$

First Step Analysis of a Markov Chain

Consider a MC on $S = \{0, 1, 2, 3\}$ with transition probability matrix

$$\mathcal{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.2 & 0.8 & 0 & 0 \\ 0.5 & 0.2 & 0.1 & 0.2 \\ 0.2 & 0.1 & 0.6 & 0.1 \\ 0 & 0 & 0.7 & 0.3 \end{bmatrix} \end{array}$$

Suppose that we start the chain in state 1 and that we stop observing the chain when it first hits state 3.

First Step Analysis of a Markov Chain

Question 1:

What is the expected number of steps the chain will take before stopping?

Question 1:

What is the expected number of steps the chain will take before stopping?

Let $T = \min\{n \geq 0 : X_n = 3\}$. ($T$ is a "first hitting time".)

Question 1:

What is the expected number of steps the chain will take before stopping?

Let $T = \min\{n \geq 0 : X_n = 3\}$. ($T$ is a "first hitting time".)

We want

$$E[T|X_0 = 1].$$

Question 1:

What is the expected number of steps the chain will take before stopping?

Let $T = \min\{n \geq 0 : X_n = 3\}$. ($T$ is a "first hitting time".)

We want

$$E[T|X_0 = 1].$$

Let's define

$$u_i = E[T|X_0 = i].$$

Question 1:

What is the expected number of steps the chain will take before stopping?

Let $T = \min\{n \geq 0 : X_n = 3\}$. ($T$ is a "first hitting time".)

We want

$$E[T|X_0 = 1].$$

Let's define

$$u_i = E[T|X_0 = i].$$

Then we want to find u_1.

First Step Analysis of a Markov Chain

$u_i = \mathrm{E}[T|X_0 = i]$

$$\begin{bmatrix} 0.2 & 0.8 & 0 & 0 \\ 0.5 & 0.2 & 0.1 & 0.2 \\ 0.2 & 0.1 & 0.6 & 0.1 \\ 0 & 0 & 0.7 & 0.3 \end{bmatrix}$$

$$\begin{aligned} u_1 &= 1 + 0.5u_0 + 0.2u_1 + 0.1u_2 + 0.2(0) \\ u_0 &= 1 + 0.2u_0 + 0.8u_1 \\ u_2 &= 1 + 0.2u_0 + 0.1u_1 + 0.6u_2 + (0.1)(0) \end{aligned}$$

$$\Rightarrow \quad \begin{aligned} u_0 &\approx 9.861 \\ u_1 &\approx 8.610 \\ u_2 &\approx 9.583 \end{aligned}$$

Result by Simulation

Pseudocode:

```
steps = 0
state = 1
while(state is not 3)
   uniform = a random number
   cdf = 0
   do i=1,3
     if(uniform<cdf+P(state,I)
       state = I
       exit I loop
     else
       cdf = cdf + P(state,I)
     end if
   end do
   steps = steps + 1
end while
```

Result by Simulation

I simulated:

- $100,000$ reps starting in state 0
- $100,000$ reps starting in state 1
- $100,000$ reps starting in state 2

Results:

$$\begin{aligned} \hat{u}_0 &\approx 9.866239 \\ \hat{u}_1 &\approx 8.615456 \\ \hat{u}_2 &\approx 9.605806 \end{aligned}$$

A really strong form of stability that one might require of a Markov chain $\{X_n\}$, is that the distribution of X_n does not change in time.

A really strong form of stability that one might require of a Markov chain $\{X_n\}$, is that the distribution of X_n does not change in time.

Definition:

A stochastic process is said to be stationary if, for any $k \in \mathbb{Z}$, the distribution of $(X_n, X_{n+1}, \ldots, X_{n+k})$ does not change as n varies.

It is clear that, for a stationary process, the distribution of X_n does not change in time.

Stationarity and Limit Behavior

It is clear that, for a stationary process, the distribution of X_n does not change in time.

On the other hand, for a time-homogeneous Markov chain, if the distribution of X_n does not change in time, then the process is stationary.

Stationarity and Limit Behavior

We will call this common distribution π.

ie:

$$\pi_i := P(X_n = i), \qquad i \in S$$

Stationarity and Limit Behavior

We will call this common distribution π.

ie:

$$\pi_i := P(X_n = i), \qquad i \in S$$

We will refer to π as the stationary or invariant distribution for the Markov chain.

Stationarity and Limit Behavior

We will call this common distribution π.

ie:

$$\pi_i := P(X_n = i), \qquad i \in S$$

We will refer to π as the stationary or invariant distribution for the Markov chain.

("Invariant" refers to time invariance.)

Stationarity and Limit Behavior

We will call this common distribution π.

ie:

$$\pi_i := P(X_n = i), \qquad i \in S$$

We will refer to π as the stationary or invariant distribution for the Markov chain.

("Invariant" refers to time invariance.)

Note that:

$$\pi_j = \sum_{i \in S} \pi_i \, p_{ij}$$

$$\left(\pi(j) = \sum_{i \in S} \pi(i) \, p(i,j) \right)$$

For a **continuous state space**, if π is the stationary distribution for a MC with "transition law" P, π satisfies:

$$\pi(A) = \int_S \pi(x)\, P(x, A)\, dx$$

Here, $P(x, A)$ is the probability that, starting at x, we move into set A in the next time step.

For a **continuous state space**, if π is the stationary distribution for a MC with "transition law" P, π satisfies:

$$\pi(A) = \int_S \pi(x)\, P(x, A)\, dx$$

Here, $P(x, A)$ is the probability that, starting at x, we move into set A in the next time step.

- For any continuous density $f(x)$, we are using the notation $f(A) = \int_A f(x)\, dx$. So, $\pi(A) = \int_A \pi(x)\, dx$.
- We will assume that $P(x, A) = \int_A p(x, y)\, dy$ for some "transition density" $p(x, y)$.

Stationarity and Limit Behavior

For a **continuous state space**, if π is the stationary distribution for a MC with "transition law" P, π satisfies:

$$\pi(A) = \int_S \pi(x)\, P(x, A)\, dx$$

Here, $P(x, A)$ is the probability that, starting at x, we move into set A in the next time step.

- For any continuous density $f(x)$, we are using the notation $f(A) = \int_A f(x)\, dx$. So, $\pi(A) = \int_A \pi(x)\, dx$.
- We will assume that $P(x, A) = \int_A p(x, y)\, dy$ for some "transition density" $p(x, y)$.

For the moment, we will assume that π exists and is unique...

Stationarity and Limit Behavior

The key thing about the stationary distribution is that, if we start a chain according to a draw from it, and iterate forward according to the transition law P, the chain will maintain that distribution at all fixed time points.

Stationarity and Limit Behavior

The key thing about the stationary distribution is that, if we start a chain according to a draw from it, and iterate forward according to the transition law P, the chain will maintain that distribution at all fixed time points.

$$ie: \quad X_0 \sim \pi \qquad \Rightarrow \qquad X_n \sim \pi \;\; \forall \; n > 0$$

Stationarity and Limit Behavior

The key thing about the stationary distribution is that, if we start a chain according to a draw from it, and iterate forward according to the transition law P, the chain will maintain that distribution at all fixed time points.

$$ie: \quad X_0 \sim \pi \qquad \Rightarrow \qquad X_n \sim \pi \;\; \forall \; n > 0$$

If we don't know π, we can't run a sample path of the Markov chain in "stationary mode" because we don't know how to choose a starting value according to π.

What if we start with some arbitrary X_0 and run a sample path for a really long time until we are convinced that the path couldn't possibly remember where it started...

What if we start with some arbitrary X_0 and run a sample path for a really long time until we are convinced that the path couldn't possibly remember where it started...

We could then think of the end value of the sample path as a draw from π.

Stationarity and Limit Behavior

What if we start with some arbitrary X_0 and run a sample path for a really long time until we are convinced that the path couldn't possibly remember where it started...

We could then think of the end value of the sample path as a draw from π.

In fact, we will get the distribution π in the limit:

$$\lim_{n\to\infty} p^{(n)}(x, y) = \pi(y)$$

(If the limit exists.)

Proof:

- Suppose we start the chain according to some distribution φ.

$$ie: \quad \varphi_i = P(X_0 = i)$$

Proof:

- Suppose we start the chain according to some distribution φ.

$$ie: \quad \varphi_i = P(X_0 = i)$$

... or, in the continuous setting

$$\varphi(A) = P(X_0 \in A) = \int_A \varphi(x)\,dx$$

Proof:

- Suppose we start the chain according to some distribution φ.

$$ie: \quad \varphi_i = P(X_0 = i)$$

... or, in the continuous setting

$$\varphi(A) = P(X_0 \in A) = \int_A \varphi(x)\,dx$$

We continue the proof in the continuous setting since it is more general...

Proof:

- Suppose we start the chain according to some distribution φ.

$$ie: \quad \varphi_i = P(X_0 = i)$$

... or, in the continuous setting

$$\varphi(A) = P(X_0 \in A) = \int_A \varphi(x)\,dx$$

We continue the proof in the continuous setting since it is more general...

(Note that φ could be concentrated at one point.)

Proof (continued):

- Now suppose that there is a limiting distribution:

Proof (continued):

- Now suppose that there is a limiting distribution:

$$P_\varphi(X_n \in A) \to \gamma_\varphi(A)$$

for some probability measure $\gamma_\varphi(\cdot)$.

Proof (continued):
Then,

$$\gamma_\varphi(A) \quad = \quad \lim_{n\to\infty} P_\varphi(X_n \in A)$$

Proof (continued):
Then,

$$\begin{aligned}\gamma_\varphi(A) &= \lim_{n\to\infty} P_\varphi(X_n \in A) \\ &= \lim_{n\to\infty} \int \varphi(x)\, P^{(n)}(x, A)\, dx\end{aligned}$$

Proof (continued):
Then,

$$\begin{aligned}
\gamma_\varphi(A) &= \lim_{n\to\infty} P_\varphi(X_n \in A) \\
&= \lim_{n\to\infty} \int \varphi(x)\, P^{(n)}(x, A)\, dx \\
&= \lim_{n\to\infty} \int \varphi(x) \int p^{(n-1)}(x, w)\, P(w, A)\, dw\, dx
\end{aligned}$$

where $p^{(n-1)}(x, w)$ is the density associated with $P^{(n-1)}(x, \cdot)$.

Proof (continued):
Then,

$$
\begin{aligned}
\gamma_\varphi(A) &= \lim_{n\to\infty} P_\varphi(X_n \in A) \\
&= \lim_{n\to\infty} \int \varphi(x)\, P^{(n)}(x, A)\, dx \\
&= \lim_{n\to\infty} \int \varphi(x) \int p^{(n-1)}(x, w)\, P(w, A)\, dw\, dx
\end{aligned}
$$

where $p^{(n-1)}(x, w)$ is the density associated with $P^{(n-1)}(x, \cdot)$.

$$
ie: \;\; P^{(n-1)}(x, B) = \int_B p^{(n-1)}(x, w)\, dw
$$

Proof (continued):

$$\gamma_\varphi(A) \;=\; \lim_{n\to\infty} \int \varphi(x) \int p^{(n-1)}(x,w)\, P(w,A)\, dw\, dx$$

Proof (continued):

$$\begin{aligned}\gamma_\varphi(A) &= \lim_{n\to\infty} \int \varphi(x) \int p^{(n-1)}(x,w)\, P(w,A)\, dw\, dx \\ &= \lim_{n\to\infty} \int \left[\int \varphi(x) p^{(n-1)}(x,w)\, dx\right] P(w,A)\, dw\end{aligned}$$

Proof (continued):

$$\begin{aligned}
\gamma_\varphi(A) &= \lim_{n\to\infty} \int \varphi(x) \int p^{(n-1)}(x,w)\, P(w,A)\, dw\, dx \\
&= \lim_{n\to\infty} \int \left[\int \varphi(x) p^{(n-1)}(x,w)\, dx\right] P(w,A)\, dw \\
&= \int \lim_{n\to\infty} \left[\int \varphi(x) p^{(n-1)}(x,w)\, dx\right] P(w,A)\, dw
\end{aligned}$$

Proof (continued):

$$
\begin{aligned}
\gamma_\varphi(A) &= \lim_{n\to\infty} \int \varphi(x) \int p^{(n-1)}(x,w)\, P(w,A)\, dw\, dx \\
&= \lim_{n\to\infty} \int \left[\int \varphi(x) p^{(n-1)}(x,w)\, dx\right] P(w,A)\, dw \\
&= \int \lim_{n\to\infty} \left[\int \varphi(x) p^{(n-1)}(x,w)\, dx\right] P(w,A)\, dw \\
&= \int \gamma_\varphi(w)\, P(w,A)\, dw
\end{aligned}
$$

Proof (continued):

$$\begin{aligned}
\gamma_\varphi(A) &= \lim_{n\to\infty} \int \varphi(x) \int p^{(n-1)}(x,w)\, P(w,A)\, dw\, dx \\
&= \lim_{n\to\infty} \int \left[\int \varphi(x) p^{(n-1)}(x,w)\, dx\right] P(w,A)\, dw \\
&= \int \lim_{n\to\infty} \left[\int \varphi(x) p^{(n-1)}(x,w)\, dx\right] P(w,A)\, dw \\
&= \int \gamma_\varphi(w)\, P(w,A)\, dw
\end{aligned}$$

$\Rightarrow \quad \gamma_\varphi$ is stationary

Stationarity and Limit Behavior

- So, a limiting distribution is stationary (invariant)!

- So, a limiting distribution is stationary (invariant)!
- Often the limit is independent of the starting distribution.

$$\lim_{n\to\infty} P^{(n)}(x, A) = \gamma(A)$$

- So, a limiting distribution is stationary (invariant)!
- Often the limit is independent of the starting distribution.

$$\lim_{n\to\infty} P^{(n)}(x, A) = \gamma(A)$$

- So, <u>if</u> the stationary distribution π is unique and if a limiting distribution exists, that limiting distribution <u>is</u> π:

$$\lim_{n\to\infty} P^{(n)}(x, A) = \pi(A)$$

Stationarity and Limit Behavior

- So, a limiting distribution is stationary (invariant)!
- Often the limit is independent of the starting distribution.

$$\lim_{n\to\infty} P^{(n)}(x, A) = \gamma(A)$$

- So, if the stationary distribution π is unique and if a limiting distribution exists, that limiting distribution is π:

$$\lim_{n\to\infty} P^{(n)}(x, A) = \pi(A)$$

or

$$\lim_{n\to\infty} p^{(n)}(x, y) = \pi(y)$$

Stationarity and Limit Behavior

- So, a limiting distribution is stationary (invariant)!
- Often the limit is independent of the starting distribution.

$$\lim_{n\to\infty} P^{(n)}(x, A) = \gamma(A)$$

- So, if the stationary distribution π is unique and if a limiting distribution exists, that limiting distribution is π:

$$\lim_{n\to\infty} P^{(n)}(x, A) = \pi(A)$$

or

$$\lim_{n\to\infty} p^{(n)}(x, y) = \pi(y)$$

or (discrete state space)

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$$

This forms the basis for all "Markov chain Monte Carlo" (MCMC) methods.

This forms the basis for all "Markov chain Monte Carlo" (MCMC) methods.

We either:

- have some sort of Markov process for which we want to understand an equilibrium distribution

This forms the basis for all "Markov chain Monte Carlo" (MCMC) methods.

We either:

- have some sort of Markov process for which we want to understand an equilibrium distribution
- have a (target) distribution we want to draw from for which we will create a Markov chain that will converge in distribution towards the target distribution

A Simple Example:

Let $\{X_n\}$ be a Markov chain on $S = \{0, 1\}$ with probability transition matrix

$$\mathcal{P} = \begin{bmatrix} 0.99 & 0.01 \\ 0.08 & 0.92 \end{bmatrix}$$

Let's find the distribution π.

We need to specify π_0 and π_1.

Stationarity and Limit Behavior

Recall that π satifies

$$\pi_j = \sum_{i \in S} \pi_i p_{ij}$$

for all j.

In other "words":

$$\vec{\pi} = \vec{\pi}\mathcal{P}$$

where $\vec{\pi} = (\pi_0, \pi_1, \ldots)$.

Stationarity and Limit Behavior

Back to the Example:

Let $\{X_n\}$ be a Markov chain on $S = \{0, 1\}$ with probability transition matrix

$$\mathcal{P} = \begin{bmatrix} 0.99 & 0.01 \\ 0.08 & 0.92 \end{bmatrix}$$

In this case,

$$\pi_0 = 8/9, \qquad \pi_1 = 1/9.$$

Seeing stationarity in action:

$$\mathcal{P} = \begin{bmatrix} 0.99 & 0.01 \\ 0.08 & 0.92 \end{bmatrix} \qquad \pi_0 = 8/9, \qquad \pi_1 = 1/9$$

- I started 100,000 sample paths (realizations of $\{X_n\}$).
  - with probability $8/9$, I started from 0
  - with probability $1/9$, I started from 1
- I ran each path for one time step according to $\mathcal{P}$.

Results:

- $89,012$ trials resulted in 0
- $10,988$ trials resulted in 1

ie:

$$\begin{aligned}\hat{\pi}_0 &= 0.89012\\ \hat{\pi}_1 &= 0.10988\end{aligned}$$

Seeing limiting behavior in action:

- I started 100,000 sample paths, all from state 0, and ran for one time step.
  Results
  - 98,955 trials resulted in 0
  - 1,045 trials resulted in 1

ie: after 1 step

$$\begin{aligned} \hat{\pi}_0 &= 0.98955 \\ \hat{\pi}_1 &= 0.01045 \end{aligned}$$

Stationarity and Limit Behavior

Repeated estimates for π_0 for samples of size 100,000:

1 step	2 steps	3 steps
0.98955	0.98099	0.97318
0.98992	0.98104	0.97316
0.98972	0.98102	0.97247

10 steps	25 steps	100 steps
0.93196	0.89821	0.88874
0.93321	0.89855	0.88894
0.93210	0.89920	0.89009

We could have predicted these results since a 3 step simulation, for example, is a single step simulation with transition probability matrix

$$\mathcal{P}_3 = \mathcal{P}^{(3)} = \mathcal{P}^3 = \begin{bmatrix} 0.972619 & 0.027281 \\ 0.219048 & 0.780952 \end{bmatrix}$$

Stationarity and Limit Behavior

We could have predicted these results since a 3 step simulation, for example, is a single step simulation with transition probability matrix

$$\mathcal{P}_3 = \mathcal{P}^{(3)} = \mathcal{P}^3 = \begin{bmatrix} 0.972619 & 0.027281 \\ 0.219048 & 0.780952 \end{bmatrix}$$

The 25 step transition probability matrix is the first one to break the 0.90/0.80 barrier for $0 \to 0$:

$$\mathcal{P}_{25} = \mathcal{P}^{(25)} = \mathcal{P}^{25} = \begin{bmatrix} 0.8994\ldots & 0.1006\ldots \\ 0.8048\ldots & 0.1952\ldots \end{bmatrix}$$

(Still hasn't converged because the columns haven't "stabilized".)

Stationarity and Limit Behavior

At 50, the transitions really start to settle down,

$$\mathcal{P}_{50} = \mathcal{P}^{(50)} = \mathcal{P}^{50} = \begin{bmatrix} 0.8899\ldots & 0.1101\ldots \\ 0.8809\ldots & 0.1191\ldots \end{bmatrix}$$

That is, there is much less change with each additional matrix multiplication.

At 50, the transitions really start to settle down,

$$\mathcal{P}_{50} = \mathcal{P}^{(50)} = \mathcal{P}^{50} = \begin{bmatrix} 0.8899\ldots & 0.1101\ldots \\ 0.8809\ldots & 0.1191\ldots \end{bmatrix}$$

That is, there is much less change with each additional matrix multiplication.

In general, the question of "How far do I need to go?" to see this convergence is a really tough question to answer and is different for each "type" of Markov chain!

Analysis of Convergence Rate via Eigenvalues

Stationarity:

$$\vec{\pi}\mathcal{P} = \vec{\pi}$$

Analysis of Convergence Rate via Eigenvalues

Stationarity:

$$\vec{\pi}\mathcal{P} = \vec{\pi}$$

Note that $\vec{\pi}$ is a (left) eigenvector of the transition matrix corresponding to the eigenvalue 1.

Analysis of Convergence Rate via Eigenvalues

Stationarity:

$$\vec{\pi}\mathcal{P} = \vec{\pi}$$

Note that $\vec{\pi}$ is a (left) eigenvector of the transition matrix corresponding to the eigenvalue 1.

Facts:

- The determinant of a stochastic matrix is at most 1.
- The largest eigenvalue of a stochastic matrix is 1.

(Perron-Frobenius Theorem)

Analysis of Convergence Rate via Eigenvalues

We consider the case of a unique stationary distribution. (ie: only one eigenvector associated with eigenvalue 1.

We consider the case of a unique stationary distribution. (ie: only one eigenvector associated with eigenvalue 1.

Denote all eigenvalues

$$1 = \lambda_1 > |\lambda_2| \geq |\lambda_3| \geq \cdots$$

and associated eigenvectors

$$\pi, \vec{v}_2, \vec{v}_3, \cdots$$

Analysis of Convergence Rate via Eigenvalues

We consider the case of a unique stationary distribution. (ie: only one eigenvector associated with eigenvalue 1.

Denote all eigenvalues

$$1 = \lambda_1 > |\lambda_2| \geq |\lambda_3| \geq \cdots$$

and associated eigenvectors

$$\pi, \vec{v}_2, \vec{v}_3, \cdots$$

We can represent the intial distribution

$$\vec{\varphi} = \vec{\pi} + a_2 \vec{v}_2 + a_3 \vec{v}_3 + \cdots$$

Analysis of Convergence Rate via Eigenvalues

$$\vec{\varphi} = \vec{\pi} + a_2\vec{v}_2 + a_3\vec{v}_3 + \cdots$$

$$\vec{\varphi} = \vec{\pi} + a_2\vec{v}_2 + a_3\vec{v}_3 + \cdots$$

Then, starting with a draw from $\vec{\varphi}$, the distribution of the chain after n steps is

$$P_\varphi(X_n = i) = i^{th} \text{ component of } \vec{\varphi}\mathcal{P}^n$$

$$\vec{\varphi} = \vec{\pi} + a_2\vec{v}_2 + a_3\vec{v}_3 + \cdots$$

Then, starting with a draw from $\vec{\varphi}$, the distribution of the chain after n steps is

$$P_{\varphi}(X_n = i) = i^{th} \text{ component of } \vec{\varphi}\mathcal{P}^n$$

But,

$$\begin{aligned} \varphi\mathcal{P}^n &= \vec{\pi}\mathcal{P}^n + a_2\vec{v}_2\mathcal{P}^n + a_3\vec{v}_3\mathcal{P}^n + \cdots \\ &\overset{stat.\&evals}{=} \vec{\pi} + a_2\lambda_2^n\vec{v}_2 + a_3\lambda_3^n\vec{v}_3 + \cdots \end{aligned}$$

Analysis of Convergence Rate via Eigenvalues

$$\vec{\varphi} = \vec{\pi} + a_2\vec{v}_2 + a_3\vec{v}_3 + \cdots$$

Then, starting with a draw from $\vec{\varphi}$, the distribution of the chain after n steps is

$$P_\varphi(X_n = i) = i^{th} \text{ component of } \vec{\varphi}\mathcal{P}^n$$

But,

$$\begin{aligned} \varphi\mathcal{P}^n &= \vec{\pi}\mathcal{P}^n + a_2\vec{v}_2\mathcal{P}^n + a_3\vec{v}_3\mathcal{P}^n + \cdots \\ &\overset{stat.\&evals}{=} \vec{\pi} + a_2\lambda_2^n\vec{v}_2 + a_3\lambda_3^n\vec{v}_3 + \cdots \end{aligned}$$

Since the λ's have magnitude smaller than 1, as n increases $\varphi\mathcal{P}^n$ will converge to $\vec{\pi}$ with a rate of convergence governed by the magnitude of the second largest eigenvalue (λ_2) of $\mathcal{P}$.

In particular

$$||P^{(n)}(x, \cdot) - \pi||_{TV} \leq C\lambda_2^n$$

where $||\mu - \nu||_{TV}$ is the **total variation norm distance** defined by

$$||\mu - \nu||_{TV} = \max_{A \subseteq S} |\mu(A) - \nu(A)|.$$

One can show that if μ and ν are discrete densities,

$$||\mu - \nu||_{TV} = \frac{1}{2} \sum_{x \in S} |\mu(x) - \nu(x)|.$$

Existence and Uniqueness of π

To explore existence and uniqueness, we need a different classification of states based on the mean recurrence time:

$$\mu_i = \mathsf{E}_i[T_i]$$

where $T_i = \min\{n \geq 1 : X_n = i\}$

Existence and Uniqueness of π

To explore existence and uniqueness, we need a different classification of states based on the mean recurrence time:

$$\mu_i = \mathsf{E}_i[T_i]$$

where $T_i = \min\{n \geq 1 : X_n = i\}$

A recurrent state for a Markov chain is one that you will eventually return to with probability 1. It is

- positive recurrent if $\mu_i < \infty$
- null recurrent if $\mu_i = \infty$

(If there is some positive probability to get away, never to return again, the state is called transient).

Existence and Uniqueness of π

Theorem: Let $\{X_n\}_{n\geq 0}$ be an irreducible Markov chain.

(irreducible= all states can be reached from all states)

A) If the chain has stationary distribution π, then π is given by

$$\pi(i) = \frac{1}{\mathsf{E}_i[T_i]}$$

(Hence π is unique!)
(Moreover, all states are positive recurrent!)

Existence and Uniqueness of π

Theorem:

B) Conversely, if the chain is positive recurrent (all states positive recurrent), then π, defined by

$$\pi(i) = \frac{1}{\mathrm{E}_i[T_i]}$$

is the unique stationary distribution.

Simulation Through First Hitting Time

- I started $100,000$ sample paths from 0
- I ran each path until 0 was hit at time $n > 0$
- Results:

  The average number of steps it took to hit 0 starting from 0:

$$1.1251 \quad \Rightarrow \quad \hat{\pi}(0) = \frac{1}{1.1251} \approx 0.88881$$

Similarly,

- I started $100,000$ sample paths from 1
- I ran each path until 1 was hit at time $n > 0$
  item Results:
  The average number of steps it took to hit 1 starting from 1:

$$8.9978 \quad \Rightarrow \quad \hat{\pi}(1) = \frac{1}{8.9978} \approx 0.11138$$

Existence and Uniqueness of π

Theorem:
An irreducible aperiodic Markov chain belongs to one of the following classes:

I. All states are transient or all are null recurrent. In this case, $\lim_{n\to\infty} P^{(n)}(i,j) = 0$ for all i, j and there is no stationary distribution.

II. All states are positive recurrent. In this case

$$\pi_j = \lim_{n\to\infty} P^{(n)}(i,j) > 0$$

is the unique stationary distribution.