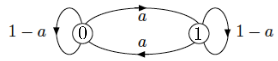
## CS70: Markov Chains.

Markov Chains

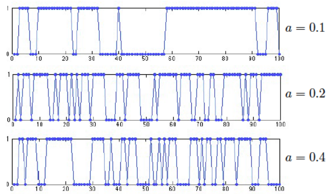
1. Examples
2. Definition
3. Hitting Time.
4. Here before there.
5. Stationary Distribution
6. Peridoicity.

## Two-State Markov Chain

Here is a symmetric two-state Markov chain. It describes a random motion in $\{0,1\}$. Here, $a$ is the probability that the state changes in the next step.

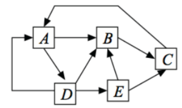



Let's simulate the Markov chain:

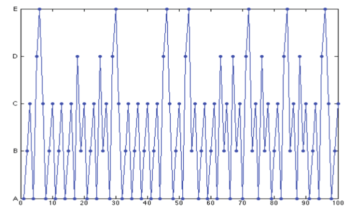



## Five-State Markov Chain

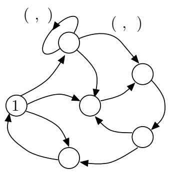
At each step, the MC follows one of the outgoing arrows of the current state, with equal probabilities.




Let's simulate the Markov chain:



## Finite Markov Chain: Definition




- A finite set of states: $\mathscr{X} = \{1,2,\dots,K\}$
- A probability distribution $\pi_0$ on $\mathscr{X}$ : $\pi_0(i) \geq 0, \sum_i \pi_0(i) = 1$
- Transition probabilities: $P(i,j)$ for $i,j \in \mathscr{X}$

  $P(i,j) \geq 0, \forall i,j; \sum_j P(i,j) = 1, \forall i$
- $\{X_n, n \geq 0\}$ is defined so that

  $Pr[X_0 = i] = \pi_0(i), i \in \mathscr{X}$ (initial distribution)

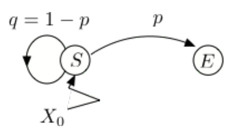
  $Pr[X_{n+1} = j \mid X_0,\dots,X_n = i] = P(i,j), i,j \in \mathscr{X}$.

## Hitting Time - Example 1

Let's flip a coin with $Pr[H] = p$ until we get $H$. How many flips, on average?

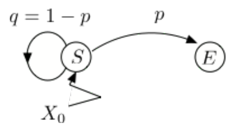
Let's define a Markov chain:

- $X_0 = S$ (start)
- $X_n = S$ for $n \geq 1$, if last flip was $T$ and no $H$ yet
- $X_n = E$ for $n \geq 1$, if we already got $H$ (end)




## First Passage Time - Example 1. Poll

Let's flip a coin with $Pr[H] = p$ until we get $H$. How many flips, on average?




Let $\beta(S)$ be the average time until $E$, starting from $S$.

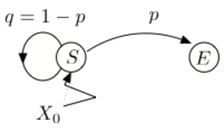
What is correct?

(A) $\beta(S)$ is at least 1.
(B) From $S$, in one step, go to $S$ with prob. $q = 1-p$
(C) From $S$, in one step, go to $E$ with prob. $p$.
(D) If you go back to $S$, you are back at $S$.
(D) $\beta(S) = 1 + q\beta(S) + p0$.

All are correct. (D) is the "Markov property." Only know where you are.

## Hitting Time - Example 1

Let's flip a coin with $Pr[H] = p$ until we get $H$. How many flips, on average?




Let $\beta(S)$ be the average time until $E$, starting from $S$.

Then,

$$\beta(S) = 1 + q\beta(S) + p0.$$

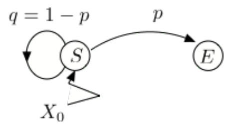
(See next slide.) Hence,

$$\beta(S) = 1 + (1-p)\beta(S) \implies \beta(S) = 1, \text{ so that } \beta(S) = 1/p.$$

Note: Time until $E$ is $G(p)$.
The mean of $G(p)$ is $1/p$!!!

## First Passage Time - Example 1

Let's flip a coin with $Pr[H] = p$ until we get $H$. How many flips, on average?




Let $\beta(S)$ be the average time until $E$.
Then,

$$\beta(S) = 1 + q\beta(S) + p0.$$

**Justification:** $N$ – number of steps until $E$, starting from $S$.
$N'$ – number of steps until $E$, after the second visit to $S$.
And $Z = 1\{\text{first flip } = H\}$. Then,

$$N = 1 + (1-Z) \times N' + Z \times 0.$$

$Z$ and $N'$ are "independent." $E[N'] = E[N] = \beta(S)$.
Hence, taking expectation,

$$\beta(S) = E[N] = 1 + (1-p)E[N'] + p0 = 1 + q\beta(S) + p0.$$

## Hitting Time - Example 2

Let's flip a coin with $Pr[H] = p$ until we get two consecutive $H$s. How many flips, on average?

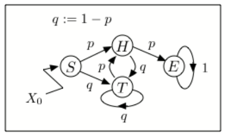
*H T H T T T H T H T H T T T H T H H*

Let's define a Markov chain:

- $X_0 = S$ (start)
- $X_n = E$, if we already got two consecutive $H$s (end)
- $X_n = T$, if last flip was $T$ and we are not done
- $X_n = H$, if last flip was $H$ and we are not done

## Hitting Time - Example 2

Let's flip a coin with $Pr[H] = p$ until we get two consecutive $H$s. How many flips, on average? Here is a picture:



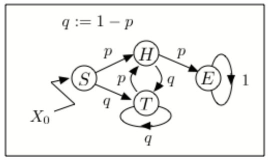

$S$: Start
$H$: Last flip $= H$
$T$: Last flip $= T$
$E$: Done

Which one is correct?
(A) $\beta(S) = 1 + p\beta(H) + q\beta(T)$
(B) $\beta(S) = p\beta(H) + q\beta(T)$
(C) $\beta(S) = \beta(S) + q\beta(T) + p\beta(H)$.

(A) Expected time from $S$ to $E$.
$\beta(S) = Pr[H]E[\beta(S)|H] + Pr[T]E[\beta(S)|T]$
$\beta(S) = p(1 + \beta(H)) + q(1 + \beta(T))$
$\beta(S) = 1 + p\beta(H) + q\beta(T)$

## Hitting Time - Example 2

Let's flip a coin with $Pr[H] = p$ until we get two consecutive $H$s. How many flips, on average? Here is a picture:




$S$: Start
$H$: Last flip $= H$
$T$: Last flip $= T$
$E$: Done

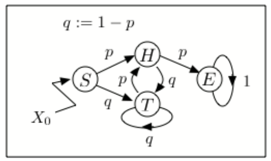
Let $\beta(i)$ be the average time from state $i$ until the MC hits state $E$.

We claim that (these are called the first step equations)

$$\beta(S) = 1 + p\beta(H) + q\beta(T)$$
$$\beta(H) = 1 + p0 + q\beta(T)$$
$$\beta(T) = 1 + p\beta(H) + q\beta(T).$$

Solving, we find $\beta(S) = 2 + 3qp^{-1} + q^2p^{-2}$. (E.g., $\beta(S) = 6$ if $p = 1/2$.)

## Hitting Time - Example 2




$S$: Start
$H$: Last flip $= H$
$T$: Last flip $= T$
$E$: Done

Let us justify the first step equation for $\beta(T)$. The others are similar.

$N(T)$ – number of steps, starting from $T$ until the MC hits $E$.
$N(H)$ – be defined similarly.
$N'(T)$ – number of steps after the second visit to $T$ until MC hits $E$.

$$N(T) = 1 + Z \times N(H) + (1-Z) \times N'(T)$$

where $Z = 1\{\text{first flip in } T \text{ is } H\}$. Since $Z$ and $N(H)$ are independent, and $Z$ and $N'(T)$ are independent, taking expectations, we get

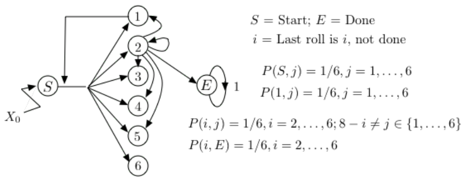
$$E[N(T)] = 1 + pE[N(H)] + qE[N'(T)],$$

i.e.,

$$\beta(T) = 1 + p\beta(H) + q\beta(T).$$

## Hitting Time - Example 3

You roll a balanced six-sided die until the sum of the last two rolls is 8.
How many times do you have to roll the die, on average?




The arrows out of $3, \ldots, 6$ (not shown) are similar to those out of 2.

$$\beta(S) = 1 + \frac{1}{6}\sum_{j=1}^{6}\beta(j); \beta(1) = 1 + \frac{1}{6}\sum_{j=1}^{6}\beta(j); \beta(i) = 1 + \frac{1}{6}\sum_{j=1,\ldots,6; j\neq 8-i}\beta(j), i = 2, \ldots, 6.$$

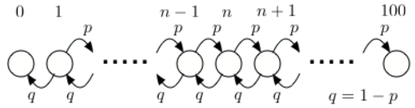
Symmetry: $\beta(2) = \cdots = \beta(6) =: \gamma$. Also, $\beta(1) = \beta(S)$. Thus,

$$\beta(S) = 1 + (5/6)\gamma + \beta(S)/6; \quad \gamma = 1 + (4/6)\gamma + (1/6)\beta(S).$$

$$\Rightarrow \cdots \beta(S) = 8.4.$$

## Here before There - A before B

Game of "heads or tails" using coin with 'heads' probability $p < 0.5$.
Start with \$10.
Each step, flip yields 'heads', earn \$1. Otherwise, lose \$1.
What is the probability that you reach \$100 before \$0?

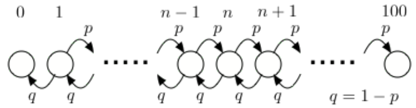



Let $\alpha(n)$ be the probability of reaching 100 before 0, starting from $n$, for $n = 0, 1, \ldots, 100$.
Which equations are correct?
(A) $\alpha(0) = 0$
(B) $\alpha(0) = 1$.
(C) $\alpha(100) = 1$.
(D) $\alpha(n) = 1 + p\alpha(n+1) + q\alpha(n-1), 0 < n < 100$.
(E) $\alpha(n) = p\alpha(n+1) + q\alpha(n-1), 0 < n < 100$.

(B) is incorrect, 0 is bad.
(D) is incorrect. Confuses expected hitting time with $A$ before $B$.

## Here before There - A before B

Game of "heads or tails" using coin with 'heads' probability $p < 0.5$.
Start with \$10.
Each step, flip yields 'heads', earn \$1. Otherwise, lose \$1.
What is the probability that you reach \$100 before \$0?



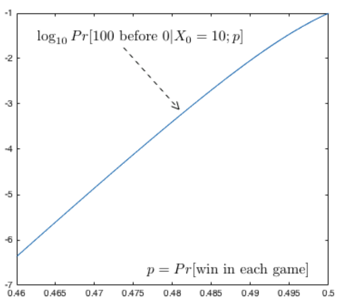

Let $\alpha(n)$ be the probability of reaching 100 before 0, starting from $n$, for $n = 0, 1, \ldots, 100$.

$$\alpha(0) = 0; \alpha(100) = 1.$$
$$\alpha(n) = p\alpha(n+1) + q\alpha(n-1), 0 < n < 100.$$

$$\Rightarrow \alpha(n) = \frac{1 - \rho^n}{1 - \rho^{100}} \text{ with } \rho = qp^{-1}. \text{ (See LN 22)}$$

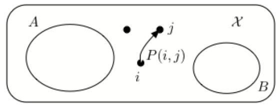
## Here before There - A before B

Game of "heads or tails" using coin with 'heads' probability $p = .48$.
Start with \$10.
Each step, flip yields 'heads', earn \$1. Otherwise, lose \$1.
What is the probability that you reach \$100 before \$0?




Less than 1 in a 1000. Moral of example: Money in Vegas stays in Vegas.

## First Step Equations




Let $X_n$ be a MC on $\mathcal{X}$ and $A, B \subset \mathcal{X}$ with $A \cap B = \emptyset$. Define

$$T_A = \min\{n \geq 0 \mid X_n \in A\} \text{ and } T_B = \min\{n \geq 0 \mid X_n \in B\}.$$

For $\beta(i) = E[T_A \mid X_0 = i]$, first step equations are:

$$\beta(i) = 0, i \in A$$
$$\beta(i) = 1 + \sum_j P(i, j)\beta(j), i \notin A$$

For $\alpha(i) = Pr[T_A < T_B \mid X_0 = i], i \in \mathcal{X},$, first step equations are:

$$\alpha(i) = 1, i \in A$$
$$\alpha(i) = 0, i \in B$$
$$\alpha(i) = \sum_j P(i, j)\alpha(j), i \notin A \cup B.$$

## Accumulating Rewards

Let $X_n$ be a Markov chain on $\mathcal{X}$ with $P$. Let $A \subset \mathcal{X}$

Let also $g : \mathcal{X} \to \Re$ be some function.

Define

$$\gamma(i) = E[\sum_{n=0}^{T_A} g(X_n) | X_0 = i], i \in \mathcal{X}.$$

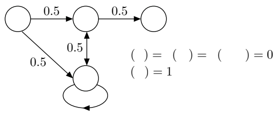
Then

$$\gamma(i) = \begin{cases} g(i), & \text{if } i \in A \\ g(i) + \sum_j P(i, j)\gamma(j), & \text{otherwise.} \end{cases}$$

## Example

Flip a fair coin until you get two consecutive $H$s.

What is the expected number of $T$s that you see?




FSE:

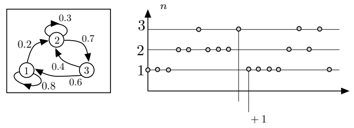
$$\begin{aligned}
\gamma(S) &= 0 + 0.5\gamma(H) + 0.5\gamma(T) \\
\gamma(H) &= 0 + 0.5\gamma(HH) + 0.5\gamma(T) \\
\gamma(T) &= 1 + 0.5\gamma(H) + 0.5\gamma(T) \\
\gamma(HH) &= 0.
\end{aligned}$$

Solving, we find $\gamma(S) = 2.5$.

## Recap

- Markov Chain:
  - Finite set $\mathscr{X}$; $\pi_0$; $P = \{P(i,j), i,j \in \mathscr{X}\}$;
  - $Pr[X_0 = i] = \pi_0(i), i \in \mathscr{X}$
  - $Pr[X_{n+1} = j \mid X_0, \ldots, X_n = i] = P(i,j), i,j \in \mathscr{X}, n \geq 0.$
  - Note:
    $Pr[X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n] = \pi_0(i_0)P(i_0,i_1) \cdots P(i_{n-1},i_n).$
- First Passage Time:
  - $A \cap B = \emptyset; \beta(i) = E[T_A | X_0 = i]; \alpha(i) = P[T_A < T_B | X_0 = i]$
  - $\beta(i) = 1 + \sum_j P(i,j)\beta(j);$
  - $\alpha(i) = \sum_j P(i,j)\alpha(j).\ \alpha(A) = 1, \alpha(B) = 0.$

## Distribution of $X_n$

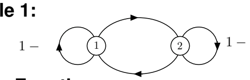



Recall $\pi_n$ is a distribution over states for $X_n$.

Stationary distribution: $\pi = \pi P$.
Distribution over states is the same before/after transition.
probability entering $i$: $\sum_{i,j} P(j,i)\pi(j)$.
probability leaving $i$: $\pi_i$.
are Equal!
Distribution same after one step.
Questions? Does one exist? Is it unique?
If it exists and is unique. Then what?
Sometimes the distribution as $n \to \infty$

## Stationary: Example

**Example 1:**




$= \begin{bmatrix} 1- & \\ & 1- \end{bmatrix}$

**Balance Equations.**

$$\begin{aligned}
\pi P = \pi \quad &\Leftrightarrow \quad [\pi(1),\pi(2)] \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} = [\pi(1),\pi(2)] \\
&\Leftrightarrow \quad \pi(1)(1-a) + \pi(2)b = \pi(1) \text{ and } \pi(1)a + \pi(2)(1-b) = \pi(2) \\
&\Leftrightarrow \quad \pi(1)a = \pi(2)b.
\end{aligned}$$

These equations are redundant! We have to add an equation:
$\pi(1) + \pi(2) = 1$. Then we find

$$\pi = [\frac{b}{a+b}, \frac{a}{a+b}].$$

## Stationary distributions: Example 2


$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$\pi P = \pi \Leftrightarrow [\pi(1),\pi(2)] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = [\pi(1),\pi(2)] \Leftrightarrow \pi(1) = \pi(1) \text{ and } \pi(2) = \pi(2).$$

Every distribution is invariant for this Markov chain. Since $X_n = X_0$ for all $n$. Hence, $Pr[X_n = i] = Pr[X_0 = i], \forall (i,n)$.

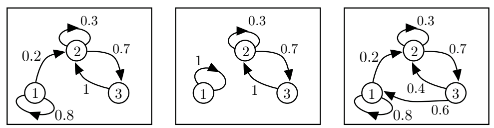
**Discussion.**
We have seen a chain with one stationary,
and a chain with many.

When is here just one?

## Irreducibility.

**Definition** A Markov chain is irreducible if it can go from every state $i$ to every state $j$ (possibly in multiple steps).

**Examples:**




[A] [B] [C]

[A] is not irreducible. It cannot go from (2) to (1).

[B] is not irreducible. It cannot go from (2) to (1).

[C] is irreducible. It can go from every $i$ to every $j$.

If you consider the graph with arrows when $P(i,j) > 0$, irreducible means that there is a single connected component.

## Existence and uniqueness of Invariant Distribution

**Theorem** A finite irreducible Markov chain has one and only one invariant distribution.

That is, there is a unique positive vector $\pi = [\pi(1), \ldots, \pi(K)]$ such that $\pi P = \pi$ and $\sum_k \pi(k) = 1$.

Ok. Now.
Only one stationary distribution if irreducible (or connected.)

## Long Term Fraction of Time in States

**Theorem** Let $X_n$ be an irreducible Markov chain with invariant distribution $\pi$.

Then, for all $i$,

$$\frac{1}{n}\sum_{m=0}^{n-1} 1\{X_m = i\} \to \pi(i), \text{ as } n \to \infty.$$

The left-hand side is the fraction of time that $X_m = i$ during steps $0, 1, \ldots, n-1$. Thus, this fraction of time approaches $\pi(i)$.

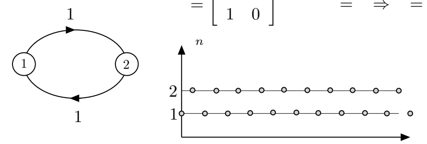
**Proof:** Lecture note 21 gives a plausibility argument. $\square$

## Long Term Fraction of Time in States

**Theorem** Let $X_n$ be an irreducible Markov chain with invariant distribution $\pi$. Then, for all $i$, $\frac{1}{n}\sum_{m=0}^{n-1} 1\{X_m = i\} \to \pi(i)$, as $n \to \infty$.

**Example 1:**

$= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad = \quad \Rightarrow \quad = [1/2, 1/2]$

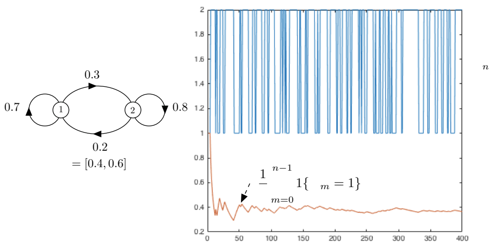



The fraction of time in state 1 converges to $1/2$, which is $\pi(1)$.

## Long Term Fraction of Time in States

**Theorem** Let $X_n$ be an irreducible Markov chain with invariant distribution $\pi$. Then, for all $i$, $\frac{1}{n}\sum_{m=0}^{n-1} 1\{X_m = i\} \to \pi(i)$, as $n \to \infty$.

**Example 2:**

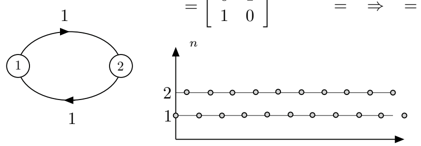



## Convergence to Invariant Distribution

**Question:** Assume that the MC is irreducible. Does $\pi_n$ approach the unique invariant distribution $\pi$?

**Answer:** Not necessarily. Here is an example:

$= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad = \quad \Rightarrow \quad = [1/2, 1/2]$




Assume $X_0 = 1$. Then $X_1 = 2, X_2 = 1, X_3 = 2, \ldots$.

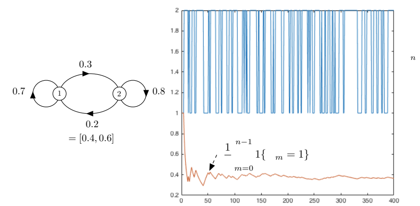
Thus, if $\pi_0 = [1,0]$, $\pi_1 = [0,1]$, $\pi_2 = [1,0]$, $\pi_3 = [0,1]$, etc.

Hence, $\pi_n$ does not converge to $\pi = [1/2, 1/2]$.
Notice, all cycles or closed walks have even length.

## Convergence to stationary distribution.

**Theorem** Let $X_n$ be an irreducible Markov chain with invariant distribution $\pi$. Then, for all $i$, $\frac{1}{n}\sum_{m=0}^{n-1} 1\{X_m = i\} \to \pi(i)$, as $n \to \infty$.

**Example 2:**



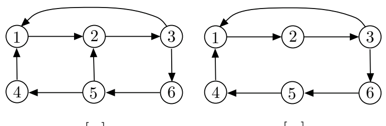

As $n$ gets large the probability of being in either state approaches $1/2$. (The stationary distribution.) Notice cycles of length 1 and 2.

## Periodicity

**Definition:** Periodicity is gcd of the lengths of all closed walks in irreducible chain. Previous example: 2.

**Definition** If periodicity is 1, Markov chain is said to be aperiodic. Otherwise, it is periodic.

**Example**




Which one is converges to stationary?

(A) [A]

(B) [B]

(C) both

(D) neither.

(A).

[A]: Closed walks of length 3 and length 4 $\implies$ periodicity = 1.

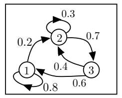
[B]: All closed walks multiple of 3 $\implies$ periodicity =2.

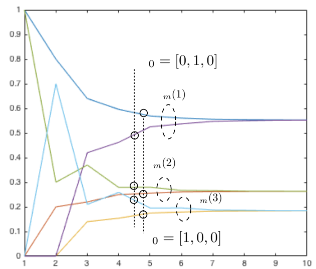
## Convergence of $\pi_n$

**Theorem** Let $X_n$ be an irreducible and aperiodic Markov chain with invariant distribution $\pi$. Then, for all $i \in \mathscr{X}$,

$$\pi_n(i) \to \pi(i), \text{ as } n \to \infty.$$

**Example**



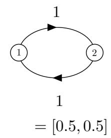



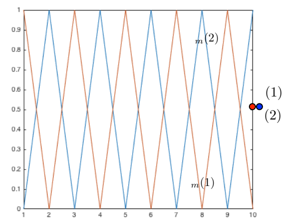

## Convergence of $\pi_n$

**Theorem** Let $X_n$ be an irreducible and aperiodic Markov chain with invariant distribution $\pi$. Then, for all $i \in \mathscr{X}$,

$$\pi_n(i) \to \pi(i), \text{ as } n \to \infty.$$

**Example**




## Summary

Markov Chains

- Markov Chain: $Pr[X_{n+1} = j | X_0, \ldots, X_n = i] = P(i,j)$
- FSE: $\beta(i) = 1 + \sum_j P(i,j)\beta(j); \alpha(i) = \sum_j P(i,j)\alpha(j)$.
- $\pi_n = \pi_0 P^n$
- $\pi$ is invariant iff $\pi P = \pi$
- Irreducible $\Rightarrow$ one and only one invariant distribution $\pi$
- Irreducible $\Rightarrow$ fraction of time in state $i$ approaches $\pi(i)$
- Irreducible + Aperiodic $\Rightarrow$ $\pi_n \to \pi$.
- Calculating $\pi$: One finds $\pi = [0, 0 \ldots, 1]Q^{-1}$ where $Q = \cdots$.