Discrete Time Markov Chains and Random Walks

Discrete Time Markov Chains

- An extremely pervasive probability model

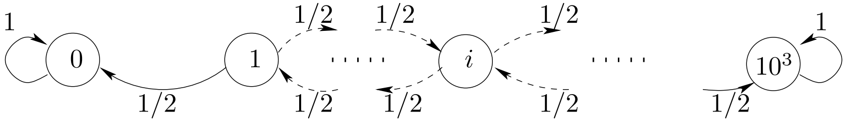
Random Walks on Graphs

- One of the most important special cases of DTMC

Why cover them in this course?

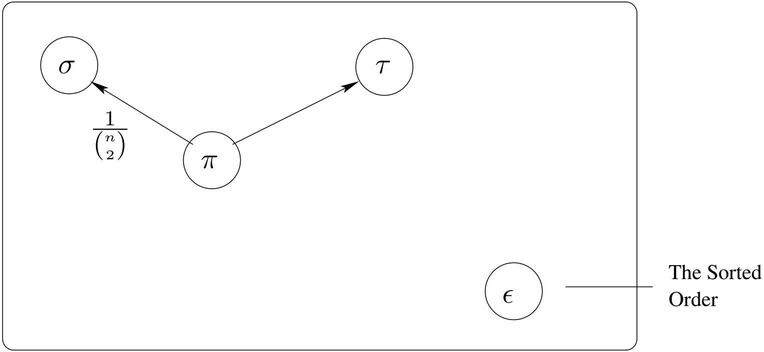
- DTMC is the first non-trivial probability model we discuss
- Lots of fundamental probabilistic reasoning
- Lots of applications in Computer Science

- A gambler has \$100, bets \$1 each game, wins with probability $1/2$.
- He stops if he gets broke or wins \$1000
- Questions: what's the probability that he gets broke? On average how many games are played?



Example 2: Slow Sort

- Given an array $A = (a_1, \ldots, a_n)$ of numbers
- Slow-Sort: as long as A is not sorted, swap two random elements
- Question: what's the running time?



$n!$ orderings of $a_1, \ldots, a_n$

Examples 3 & 4: Card Shuffling and Single Server Queue

Shuffling a Deck of Cards

- What's the number of shuffles until the ordering is "almost" uniform?

Single Server Queue

- At each time slot, an Internet router's buffer gets an additional packet with probability p, or releases one packet with probability q, or remains the same with probability r.
- Starting from an empty buffer, what is the distribution of the number of packets after n slots?
- As $n \to \infty$, will the buffer be overflowed?
- As $n \to \infty$, what's the typical buffer size?

Examples 5 & 6: 2-SAT and P2P Networks

2-SAT: find a satisfying assignment if there's any

1. Assign $x_i = \text{TRUE}/\text{FALSE}$ with probability $1/2$
2. If there's an unsatisfied clause, flip value of one of the two literals randomly
3. If there is a satisfying assignment, how long does it take?

Searching on P2P Networks

- The initial node sends k independent queries out
- Each query sent to a random neighbor
- Upon receiving a query, an intermediate node forwards it to its random neighbor
- Questions: how long does it take until all nodes receive one of the queries? What's the trade-off between k and this speed?

Stochastic Process

- Stochastic process: a collection of random variables (on some probability space) indexed by some set T:

$$\{X_t, t \in T\}$$

  When $T \subseteq \mathbb{R}$, think of T as set of points in time
- State space, denoted by I, is the set of all possible values of the X_t
- When T is countable: discrete-time stochastic process
- When T is an interval of the real line: continuous-time stochastic process

Examples of Stochastic Processes

Bernoulli process: a sequence $\{X_0, X_1, X_2, \dots\}$ of *independent* Bernoulli random variables with parameter p.

- In assignment 1, we have derived statistics on

$$\begin{aligned} S_n &= X_1 + \cdots + X_n \\ T_n &= \text{number of slots from the } (n-1)\text{th } 1 \text{ to the } n\text{th } 1 \\ Y_n &= T_1 + \cdots + T_n \end{aligned}$$

In practice, things are not that easy. We often see processes whose variables are correlated.

- Stock market and exchange rate fluctuations
- Signals (speech, audio and video)
- Daily temperatures
- Brownian motion or random walks
- etc.

Discrete-Time Markov Chain

- A DTMC is a discrete-time stochastic process $\{X_n\}_{n\geq 0}$
- State space I is countable (often labeled with a subset of $\mathbb{N}$)
- For all states i, j there is a given probability p_{ij} such that

$$P\big[X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots X_0 = i_0\big] = p_{ij},$$

for all $i_0, \dots, i_{n-1} \in I$ and all $n \geq 0$.

$$\begin{aligned} p_{ij} &\geq 0, \quad \forall i, j \in I, \\ \sum_{j \in I} p_{ij} &= 1, \quad \forall i \geq 0. \end{aligned}$$

- $\mathbf{P} = (p_{ij})$ is called the *transition probability matrix* of the chain

Typical Questions We Want Answers To

Given a DTMC $\mathbf{P}$ with state space I, $A \subset I$

- Starting from $X_0 = i \notin A$
  - What's the probability A is ever "hit"?
  - What's the probability $X_n \in A$ for a given n?
  - What's the expected number of steps until A is hit
  - What's the probability we'll come back to i?
  - What's the expected number of steps until we come back?
  - What's the expected number of steps until all states are visited?
  - As $n \to \infty$, what's the distribution of where we are? Does the "limit distribution" even exist? If it does, how fast is the convergence rate?
- Given an initial distribution λ of X_0, repeat the above questions
- And many more, depending on the application

Some Definitions

- A measure on I is a vector λ where $\lambda_i \geq 0$, for all $i \in I$.
- A measure is a distribution if $\sum_i \lambda_i = 1$.
- For any event F, let

$$\mathsf{Pr}_i[F] = \mathsf{Pr}[F \mid X_0 = i]$$

- For any random variable Z, let

$$\mathsf{E}_i[Z] = \mathsf{E}[Z \mid X_0 = i]$$

- If we know λ is the distribution of X_0, then we also write

$$(X_n)_{n \geq 0} = \text{Markov}(\mathbf{P}, \lambda).$$

Multistep Transition Probabilities and Matrices

- Define the probability of going from i to j in n steps $p_{ij}^{(n)} = \mathsf{Pr}_i[X_n = j]$
- And the n-step transition probability matrix $\mathbf{P}^{(n)} = (p_{ij}^{(n)})$

Chapman-Kolmogorov Equations

$$p_{ij}^{(m+n)} = \sum_{k \in I} p_{ik}^{(n)} p_{kj}^{(m)}, \quad \forall n, m \geq 0, i, j \in I.$$

It follows that

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

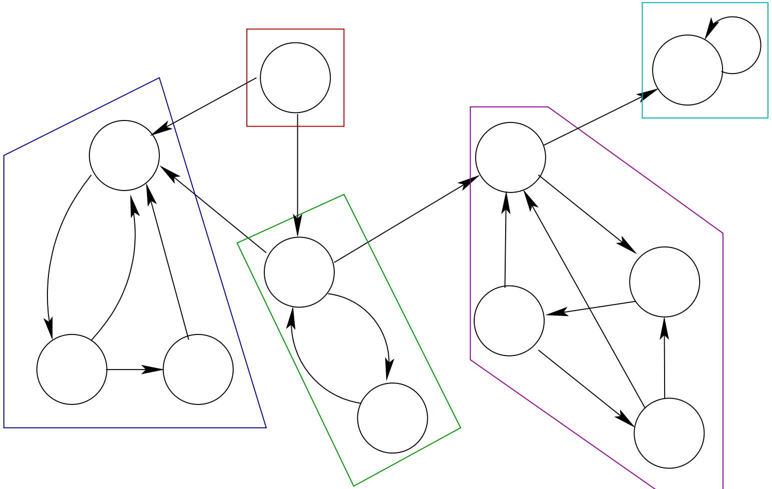
Corollary

If λ (a row vector) is the distribution of X_0, then $\lambda \mathbf{P}^n$ is the distribution of X_n

Communication Classes

- j is *reachable* from i if $p_{ij}^{(n)} > 0$ for some $n \geq 0$. We write $i \rightsquigarrow j$.
- i and j *communicate* if $i \rightsquigarrow j$ and $j \rightsquigarrow i$. We write $i \leftrightarrow j$.
- Communication is an *equivalence relation*, partitioning I into *communication classes*
- Communication classes are *strongly connected components* of the directed graph corresponding to for $\mathbf{P}$
- A chain is *irreducible* if there is only one class
- A *closed* class C is a class where $i \in C$ and $i \rightsquigarrow j$ imply $j \in C$ (no escape!)
- A state i is *absorbing* if $\{i\}$ is a closed class

Illustration of Communication Classes



Recurrent and Transient States

- First passage time

$$T_i = \min\{n \geq 1 \mid X_n = i\} \quad (X_0 = i \text{ doesn't count!})$$

- Define

$$\begin{aligned} f_{ij}^{(n)} &= \mathsf{Pr}_i[X_n = j \ \wedge \ X_s \neq j, \forall s = 1, .., n-1] = \mathsf{Pr}_i[T_j = n] \\ f_{ij} &= \sum_{n=1}^{\infty} f_{ij}^{(n)} \end{aligned}$$

- Note that $f_{ij} = \mathsf{Pr}_i[T_j < \infty]$
- State i is
  - recurrent (also called persistent) if $f_{ii} = \mathsf{Pr}_i[T_i < \infty] = 1$
  - transient if $f_{ii} = \mathsf{Pr}_i[T_i < \infty] < 1$

When is a State Recurrent?

- Let V_i be the number of visits to i, namely $V_i := \sum_{n=0}^{\infty} 1_{\{X_n=i\}}$.

Theorem

Given a DTMC $\mathbf{P}$ and a state i, the following are equivalent

1. i *is recurrent*
2. $f_{ii} = \mathsf{Pr}_i[T_i < \infty] = 1$
3. $\mathsf{Pr}_i[X_n = i$ *for infinitely many* $n] = 1$
4. $\mathsf{E}_i[V_i] = \infty$
5. $\sum_{n\geq 0} p_{ii}^{(n)} = \infty$

Theorem

Given a DTMC $\mathbf{P}$ *and a state* i, *the following are equivalent*

1. i *is transient*
2. $f_{ii} = \mathsf{Pr}_i[T_i < \infty] < 1$
3. $\mathsf{Pr}_i[X_n = i \text{ for infinitely many } n] = 0$
4. $\mathsf{E}_i[V_i] < \infty$
5. $\sum_{n \geq 0} p_{ii}^{(n)} < \infty$

To prove the last two theorems, we need the *strong Markov property*

Strong Markov Property

- A r.v. T taking values in $\mathbb{N}$ is called a stopping time of a DTMC $(X_n)_{n\geq 0}$ if the event $\{T = n\}$ can be determined by looking at $X_0, \cdots, X_n$
  (Need measure theory to be rigorous on this definition.)
- Examples:
  - First passage time $T_i = \min\{n \geq 1 : X_n = i\}$ is a stopping time
  - Last exit time $L_A = \max\{n : X_n \in A\}$ is not a stopping time

Theorem (Strong Markov Property)

Suppose T is a stopping time of a DTMC $(X_n)_{n\geq 0}$. Then, conditioned on $T < \infty$ and $X_T = i$, the sequence $(X_{T+n})_{n\geq 0}$ behaves exactly like the Markov chain with initial state i.

Proofs of Previous Two Theorems

Proof.

By strong Markovian:

$$\mathsf{E}_i[V_i] = \sum_{n=1}^{\infty} n f_{ii}^{n-1}(1 - f_{ii}) = \frac{1}{1 - f_{ii}}.$$

On the other hand,

$$\mathsf{E}_i[V_i] = \mathsf{E}_i\left[\sum_{n=0}^{\infty} 1_{\{X_n = i\}}\right] = \sum_{n=0}^{\infty} \mathsf{E}_i[1_{\{X_n=i\}}] = \sum_{n=0}^{\infty} p_{ii}^{(n)}.$$

□

Recurrence and Transience are Class Properties

Theorem

Recurrence and transience are class properties, i.e. in a communication class C either all states are recurrent or all states are transient

Why is it true? Suppose i and j belong to the same class. If i is recurrent and j is transient, each time the process returns to i there's a positive chance of going to j. Thus, the process cannot avoid j forever.

Recurrent and Closed Classes

Theorem

(i) *Every recurrent class is closed*

(ii) *Every finite, closed class is positive recurrent*

- *Why does* (i) *hold?*

If the class is not closed, there's an escape route, and thus the class cannot be recurrent.

- *Why does* (ii) *hold?*

In a finite and closed class, it cannot be the case that every state is visited a finite number of times. So, the chain is recurrent.

Infinite Closed Class Could be Transient or Recurrent

- Consider a random walk on $\mathbb{Z}$, where $p_{i,i+1} = p$ and $p_{i+1,i} = 1 - p$, for all $i \in \mathbb{Z}$, $0 < p < 1$
- The chain is an infinite and closed class.
- For any state i, we have

$$\begin{aligned} p_{ii}^{(2n+1)} &= 0 \\ p_{ii}^{(2n)} &= \binom{2n}{n} p^n (1-p)^n \end{aligned}$$

Hence,

$$\begin{aligned} \sum_{n=0}^{\infty} p_{ii}^{(2n)} &= \sum_{n=0}^{\infty} \binom{2n}{n} p^n (1-p)^n \\ &\approx \sum_{n \geq n_0} \frac{1}{\sqrt{\pi n}} (4p(1-p))^n (1 + o(1)). \end{aligned}$$

which is ∞ if $p = 1/2$ and finite if $p \neq 1/2$.

More on Irreducible and Recurrent Chain

Theorem

In an irreducible and recurrent chain, $f_{ij} = 1$ for all i, j

Why is it true? If $f_{ij} < 1$, there's a non-zero chance of the chain starting from j, getting to i, and never come back to j. However, j is recurrent!

Example: Birth-and-Death Chain

Consider a DTMC on state space $\mathbb{N}$ where

- $p_{i,i+1} = a_i$, $p_{i,i-1} = b_i$, $p_{ii} = c_i$
- $a_i + b_i + c_i = 1, \ \forall i \in \mathbb{N}$, and implicitly $b_0 = 0$
- $a_i, b_i > 0$ for all i, except for b_0

Question

When is this chain transient/recurrent?

To answer this question, we need some results about computing hitting probabilities

Hitting Times and Hitting Probabilities

- Let $\mathbf{P}$ be a DTMC on I. Let $A \subseteq I$.
- The *hitting time* H^A

$$H^A := \min\{n \geq 0 \; : \; X_n \in A\}.$$

- The probability of hitting A starting from i

$$h_i^A := \mathsf{Pr}_i[H^A < \infty].$$

- If A is a closed class, the h_i^A are called the *absorption probabilities*
- The *mean hitting time* μ_i^A is defined by

$$\mu_i^A := \mathsf{E}_i[H^A] = \sum_{n<\infty} n\mathsf{Pr}[H^A = n] + \infty\mathsf{Pr}[H^A = \infty]$$

Computing Hitting Probabilities

Theorem

The vector $(h_i^A : i \in I)$ *is the minimal non-negative solution to the following system*

$$\begin{cases} h_i^A = 1 & \text{for } i \in A \\ h_i^A = \sum_{j \in I} p_{ij} h_j^A & \text{for } i \notin A \end{cases}$$

Back to the Birth-and-Death Chain

- Note that $f_{00} = c_0 + a_0 h_1^{\{0\}} = 1 - a_0(1 - h_1^{\{0\}})$

$$\text{The system: } \begin{cases} h_0^{\{0\}} = 1 \\ h_i^{\{0\}} = a_i h_{i+1}^{\{0\}} + c_i h_i^{\{0\}} + b_i h_{i-1}^{\{0\}} & \text{for } i \geq 1 \end{cases}$$

- Define $d_n := \frac{b_1 \dots b_n}{a_1 \dots a_n}$, $n \geq 1$, and $d_0 = 1$
- When $\sum_{n=0}^{\infty} d_n = \infty$, $h_i^{\{0\}} = 1, \forall i$ is the solution
- When $\sum_{n=0}^{\infty} d_n < \infty$, we have the following solution

$$\begin{cases} h_0^{\{0\}} = 1 \\ h_i^{\{0\}} = \frac{\sum_{j=i}^{\infty} d_j}{\sum_{j=0}^{\infty} d_j} < 1 & \text{for } i \geq 1 \end{cases}$$

Thus,

- the DTMC is recurrent $(f_{00} = 1)$ when $\sum_{j=0}^{\infty} d_j = \infty$
- the DTMC is transient $(f_{00} < 1)$ when $\sum_{j=0}^{\infty} d_j < \infty$

Brief Summary of Recurrent and Transient Properties

- We often only need to look at closed classes (that's where the chain will eventually end up).
- We can then consider irreducible chains instead.

Let $\mathbf{P}$ be an irreducible chain.

- If $\mathbf{P}$ is finite, then $\mathbf{P}$ is recurrent.
- If $\mathbf{P}$ is infinite, then $\mathbf{P}$ could be either transient or recurrent.

Invariant Distribution

- A distribution λ is a *stationary* (also *equilibrium* or *invariant*) distribution if $\lambda^T \mathbf{P} = \lambda$

Theorem

(i) *Let* $(X_n)_{n\geq 0} = \text{Markov}(\mathbf{P}, \lambda)$*, where* λ *is stationary, then* $(X_{n+m})_{n\geq 0} = \text{Markov}(\mathbf{P}, \lambda)$ *for any fixed* m*.*

(ii) *In a* **finite** *DTMC, suppose for some* $i \in I$ *we have*

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j, \forall j \in I,$$

then $\pi = (\pi_j : j \in I)$ *is an invariant distribution.*

Infinite State Space Could be Strange

- In an infinite DTMC, it is possible that $\lim_{n\to\infty} p_{ij}^{(n)}$ exists for all i, j, producing a vector π for each i, yet π is **not** a distribution.
- Consider the DTMC with state space $\mathbb{Z}$ and

$$p_{i,i+1} = p = 1 - q = 1 - p_{i,i-1}, \quad \forall i \in \mathbb{Z}.$$

(This is a random walk on $\mathbb{Z}$ we have considered.)

$$\lim_{n\to\infty} p_{ij}^{(n)} = 0, \forall i, j.$$

Positive and Null Recurrent States

Define

$$\mu_{ij} = \sum_{n=1}^{\infty} n f_{ij}^{(n)} = \mathsf{E}_i[T_j]$$

Definition

A recurrent state i is positive recurrent if $\mu_{ii} < \infty$

Definition

A recurrent state i is null recurrent if $\mu_{ii} = \infty$

Example of Positive Recurrent States

$$\mathbf{P} = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}, \text{for } 0 < p < 1.$$

Let the states be 0 and 1, then

$$\begin{aligned} f_{00}^{(1)} &= f_{11}^{(1)} = p \\ f_{00}^{(n)} &= f_{11}^{(n)} = (1-p)^2 p^{n-2}, \quad n \geq 2 \end{aligned}$$

Both states are recurrent. Moreover,

$$\mu_{00} = \mu_{11} = p + \sum_{n=2}^{\infty} n(1-p)^2 p^{n-2} = 2.$$

Hence, both states are positive recurrent states.

Example of Null Recurrent States

Consider a Markov chain with $I = \mathbb{N}$ where $p_{01} = 1$, and

$$\begin{aligned} P_{m,m+1} &= \frac{m}{m+1}, \quad \forall m \geq 1 \\ P_{m,0} &= \frac{1}{m+1}, \quad \forall m \geq 1. \end{aligned}$$

Then,

$$\begin{aligned} f_{00}^{(1)} &= 0 \\ f_{00}^{(n)} &= \frac{1}{n(n-1)} \\ f_{00} = \sum_{n=1}^{\infty} f_{00}^{(n)} &= 1 \\ \mu_{00} = \sum_{n=1}^{\infty} n f_{00}^{(n)} &= \sum_{n=1}^{\infty} \frac{1}{n} = \infty. \end{aligned}$$

Consequently, 0 is a null recurrent state.

Positive and Null Recurrence are Class Properties

Theorem

Positive and null recurrence are class properties, i.e. in a recurrent communication class either all states are positive recurrent or all states are null recurrent.

Example: Birth-and-Death Chain

Consider a DTMC on state space $\mathbb{N}$ where

- $p_{i,i+1} = a_i$, $p_{i,i-1} = b_i$, $p_{ii} = c_i$
- $a_i + b_i + c_i = 1, \ \forall i \in \mathbb{N}$, and implicitly $b_0 = 0$
- $a_i, b_i > 0$ for all i, except for b_0

Question

When is this chain positive/null recurrent?

To answer this question, we need a result on computing mean hitting times

Mean Hitting Times

- Let $\mathbf{P}$ be a DTMC on I. Let $A \subseteq I$.
- The *hitting time* H^A

$$H^A := \min\{n \geq 0 \;:\; X_n \in A\}.$$

- The probability of hitting A starting from i

$$h_i^A := \mathsf{Pr}_i[H^A < \infty].$$

- If A is a closed class, the h_i^A are called the *absorption probabilities*
- The *mean hitting time* μ_i^A is defined by

$$\mu_i^A := \mathsf{E}_i[H^A] = \sum_{n<\infty} n\mathsf{Pr}[H^A = n] + \infty\mathsf{Pr}[H^A = \infty]$$

Computing Mean Hitting Times

Theorem

The vector $(\mu_i^A : i \in I)$ *is the minimal non-negative solution to the following system*

$$\begin{cases} \mu_i^A = 0 & \text{for } i \in A \\ \mu_i^A = 1 + \sum_{j \notin A} p_{ij} \mu_j^A & \text{for } i \notin A \end{cases}$$

Back to the Birth-and-Death Chain

- Note that $\mu_{00} = c_0 + a_0(1 + \mu_1^{\{0\}}) = 1 + a_0\mu_1^{\{0\}}$

$$\text{The system:} \quad \begin{cases} \mu_0^{\{0\}} = 0 \\ \mu_1^{\{0\}} = 1 + a_1\mu_2^{\{0\}} + c_1\mu_1^{\{0\}} \\ \mu_i^{\{0\}} = 1 + a_i\mu_{i+1}^{\{0\}} + c_i\mu_i^{\{0\}} + b_i\mu_{i-1}^{\{0\}} & \text{for } i \geq 2 \end{cases}$$

- Define $e_n := \frac{a_0 \dots a_{n-1}}{b_1 \dots b_n}$, $n \geq 1$.
- When $\sum_{n=1}^{\infty} e_n = \infty$, $\mu_i^{\{0\}} = \infty, \forall i \geq 1$ is the solution
- When $\sum_{n=1}^{\infty} e_n < \infty$, we have the following solution

$$\begin{cases} \mu_0^{\{0\}} = 0 \\ \mu_1^{\{0\}} = \frac{1}{a_0}\sum_{n=1}^{\infty} e_n \\ \mu_i^{\{0\}} = \frac{d_i}{a_0}(\sum_{j=i}^{\infty} e_j) & \text{for } i \geq 2 \end{cases}$$

Thus, (conditioned on the chain being recurrent)

- the DTMC is positive recurrent when $\sum_{j=1}^{\infty} e_j < \infty$
- the DTMC is null recurrent when $\sum_{j=1}^{\infty} e_j = \infty$

Existence of a Stationary Distribution

Theorem

An irreducible DTMC $\mathbf{P}$ has a stationary distribution if and only if one of its states is positive recurrent.
Moreover, if $\mathbf{P}$ has a stationary distribution π, then

$$\pi_i = 1/\mu_{ii}$$

.

Line of proof

- Every irreducible and recurrent $\mathbf{P}$ basically has a unique invariant measure (unique up to rescaling)
- Due to positive recurrence, the measure can be normalized to be come an invariant distribution

Proof of the Existence of a Stationary Distribution

Define the expected time spent in i between visits to k

$$\gamma_i^{(k)} = \mathsf{E}_k \left[\sum_{n=0}^{T_k - 1} 1_{X_n = i} \right]$$

Lemma

If $\mathbf{P}$ is irreducible and recurrent, then

(i) $\gamma_k^{(k)} = 1$

(ii) *the vector $\gamma^{(k)} = (\gamma_i^{(k)} \mid i \in I)$ is an invariant measure, namely*

$$\gamma^{(k)} \mathbf{P} = \gamma^{(k)}$$

(iii) $0 < \gamma_i^{(k)} < \infty$ *for all $i \in I$*

Conversely, if $\mathbf{P}$ is irreducible and λ is an invariant measure with $\lambda_k = 1$, then $\lambda \geq \gamma^{(k)}$. Moreover, if $\mathbf{P}$ is also recurrent, then $\lambda = \gamma^{(k)}$

Periodicity

- For a state $i \in I$, let $d_i = \gcd\{n \geq 1 : p_{ii}^{(n)} > 0\}$.
- When $d_i \geq 2$, state i is periodic with period d_i.
- When $d_i = 1$, state i is *aperiodic*.
- A DTMC is periodic if it has a periodic state. Otherwise, the chain is violetaperiodic.

Theorem

If $i \leftrightarrow j$, then $d_i = d_j$.
If i is aperiodic, then $\exists n_0 : p_{ii}^{(n)} > 0$, $\forall n \geq n_0$.

Corollary

If $\mathbf{P}$ is irreducible and has an aperiodic state i, then $\mathbf{P}^n$ has all strictly positive entries for sufficiently large n.

Ergodicity

- An *ergodic* state is an aperiodic and positive recurrent state.
- An *ergodic Markov chain* is a Markov chain in which all states are ergodic.
  (Basically, a "well-behaved" chain.)

Convergence to equilibrium

Theorem

Suppose $\mathbf{P}$ is irreducible and ergodic. Then, it has an invariant distribution π. Moreover,

$$\frac{1}{\mu_{jj}} = \pi_j = \lim_{n\to\infty} p_{ij}^{(n)}, \quad \forall j \in I.$$

Thus, π is the unique invariant distribution of $\mathbf{P}$.

Note: there is a generalized version of this theorem for irreducible chains with period $d \geq 2$. (And the chain is not even required to be positive recurrent.)

Ergodic Theorem

Let

$$V_i(n) = \sum_{k=0}^{n-1} 1_{\{X_k = i\}}.$$

Theorem (Ergodic Theorem)

Let $\mathbf{P}$ be an irreducible DTMC. Then

$$\Pr\left[\lim_{n\to\infty} \frac{V_i(n)}{n} = \frac{1}{\mu_{ii}}\right] = 1$$

Moreover, if $\mathbf{P}$ is positive recurrent with (unique) invariant distribution π, then for any bounded function $f : I \to \mathbb{R}$

$$\Pr\left[\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \sum_{i \in I} \pi_i f_i\right] = 1,$$