Markov chains and applications

Dr. András Pongrácz

Week 11, University of Debrecen

Mixing time

Total variation distance

A good way to measure mixing in ergodic chains is based on the total variation distance.

Total variation distance

A good way to measure mixing in ergodic chains is based on the total variation distance.

Given two distribution μ, ν on the state space Ω, their total variation distance is defined as

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

Total variation distance

A good way to measure mixing in ergodic chains is based on the total variation distance.

Given two distribution μ, ν on the state space Ω, their total variation distance is defined as

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

We can measure that, given any initial distribution represented as a row vector $\underline{\mu}^*$, how far are we from the stationary distribution after k steps:

Total variation distance

A good way to measure mixing in ergodic chains is based on the total variation distance.

Given two distribution μ, ν on the state space Ω, their total variation distance is defined as

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

We can measure that, given any initial distribution represented as a row vector $\underline{\mu}^*$, how far are we from the stationary distribution after k steps :

$$|\underline{\mu}^* P^k - \underline{w}^*|_{TV}$$

Let P be an ergodic Markov chain with stationary distribution $\underline{w}^*$.

Mixing time

Let P be an ergodic Markov chain with stationary distribution $\underline{w}^*$.

The mixing time $t_{mix}(\varepsilon)$ is the smallest t such that starting the walk from any initial state, the distribution of states after k moves is within ε local variation distance from $\underline{w}^*$.

Mixing time

Let P be an ergodic Markov chain with stationary distribution $\underline{w}^*$.

The mixing time $t_{mix}(\varepsilon)$ is the smallest t such that starting the walk from any initial state, the distribution of states after k moves is within ε local variation distance from $\underline{w}^*$.

rapid mixing : polynomial in $\log \Omega$
slow mixing : exponential in $\log \Omega$

Mixing time

Let P be an ergodic Markov chain with stationary distribution $\underline{w}^*$.

The mixing time $t_{mix}(\varepsilon)$ is the smallest t such that starting the walk from any initial state, the distribution of states after k moves is within ε local variation distance from $\underline{w}^*$.

rapid mixing : polynomial in $\log \Omega$
slow mixing : exponential in $\log \Omega$
phase transition

Remarks

The parameter ε is not so important, often $\varepsilon = 1/4$.

Remarks

The parameter ε is not so important, often $\varepsilon = 1/4$. This makes little difference, as $t_{mix}(\varepsilon) \leq \lceil \log_2 \varepsilon^{-1} \rceil t_{mix}(1/4)$.

Remarks

The parameter ε is not so important, often $\varepsilon = 1/4$. This makes little difference, as $t_{mix}(\varepsilon) \leq \lceil \log_2 \varepsilon^{-1} \rceil t_{mix}(1/4)$. Sometimes $t_{mix}(1/4)$ is denoted by t_{mix}, and it is called the mixing time.

Remarks

The parameter ε is not so important, often $\varepsilon = 1/4$. This makes little difference, as $t_{mix}(\varepsilon) \leq \lceil \log_2 \varepsilon^{-1} \rceil t_{mix}(1/4)$. Sometimes $t_{mix}(1/4)$ is denoted by t_{mix}, and it is called the mixing time.

Furthermore, in the definition, we start from a Dirac distribution (concentrated on one state). We could start from any initial distribution, it would not affect the notion. See the exercises.

Shuffling cards

There are 52! orders of a deck of cards.

Shuffling cards

There are 52! orders of a deck of cards. These are the sates of an ergodic Markov chain.

Shuffling cards

There are 52! orders of a deck of cards. These are the sates of an ergodic Markov chain.

From each state, we move to one that can be reached by riffle shuffle uniformly at random.

Shuffling cards

There are 52! orders of a deck of cards. These are the sates of an ergodic Markov chain.

From each state, we move to one that can be reached by riffle shuffle uniformly at random. There are $2^{52} - 52$ such states ; see the exercises.

Shuffling cards

There are 52! orders of a deck of cards. These are the sates of an ergodic Markov chain.

From each state, we move to one that can be reached by riffle shuffle uniformly at random. There are $2^{52} - 52$ such states; see the exercises. How many shuffles are needed for the deck to be mixed sufficiently arbitrarily? If certain orders of cards occur with significantly larger probability than others, a professional gambler can turn it against us. ("ace tracking")

Shuffling cards

There are 52! orders of a deck of cards. These are the sates of an ergodic Markov chain.

From each state, we move to one that can be reached by riffle shuffle uniformly at random. There are $2^{52} - 52$ such states; see the exercises. How many shuffles are needed for the deck to be mixed sufficiently arbitrarily? If certain orders of cards occur with significantly larger probability than others, a professional gambler can turn it against us. ("ace tracking")

According to a famous result, $t_{mix} = t_{mix}(1/4) = 7$.

Shuffling cards

There are 52! orders of a deck of cards. These are the sates of an ergodic Markov chain.

From each state, we move to one that can be reached by riffle shuffle uniformly at random. There are $2^{52} - 52$ such states ; see the exercises. How many shuffles are needed for the deck to be mixed sufficiently arbitrarily ? If certain orders of cards occur with significantly larger probability than others, a professional gambler can turn it against us. ("ace tracking")

According to a famous result, $t_{mix} = t_{mix}(1/4) = 7$. If the 25% error rate is not good enough, ask for 12 shuffles.

Walks on graphs

Given a connected graph G with m edges.

Given a connected graph G with m edges. In the random walk on G, we choose the next vertex uniformly at random from the neighbors: if the current vertex v has degree $d(v)$, then each neighbor follows by probability $1/d(v)$.

Given a connected graph G with m edges. In the random walk on G, we choose the next vertex uniformly at random from the neighbors : if the current vertex v has degree $d(v)$, then each neighbor follows by probability $1/d(v)$.

This is an irreducible Markov chain.

Given a connected graph G with m edges. In the random walk on G, we choose the next vertex uniformly at random from the neighbors: if the current vertex v has degree $d(v)$, then each neighbor follows by probability $1/d(v)$.

This is an irreducible Markov chain.

Unless the graph is bipartite, the chain is aperiodic.

Given a connected graph G with m edges. In the random walk on G, we choose the next vertex uniformly at random from the neighbors: if the current vertex v has degree $d(v)$, then each neighbor follows by probability $1/d(v)$.

This is an irreducible Markov chain.

Unless the graph is bipartite, the chain is aperiodic. If G is bipartite, then the period is 2.

Reversibility

The distribution assigning $d(v)/(2m)$ to each vertex v satisfies the detailed balance equations.

Reversibility

The distribution assigning $d(v)/(2m)$ to each vertex v satisfies the detailed balance equations. Indeed, if uv is an edge, then $\frac{d(u)}{2m} \cdot \frac{1}{d(u)} = \frac{1}{2m} = \frac{d(v)}{2m} \cdot \frac{1}{d(v)}$,

Reversibility

The distribution assigning $d(v)/(2m)$ to each vertex v satisfies the detailed balance equations. Indeed, if uv is an edge, then $\frac{d(u)}{2m} \cdot \frac{1}{d(u)} = \frac{1}{2m} = \frac{d(v)}{2m} \cdot \frac{1}{d(v)}$, and if uv is a non-edge, then both sides are zero.

Reversibility

The distribution assigning $d(v)/(2m)$ to each vertex v satisfies the detailed balance equations. Indeed, if uv is an edge, then $\frac{d(u)}{2m} \cdot \frac{1}{d(u)} = \frac{1}{2m} = \frac{d(v)}{2m} \cdot \frac{1}{d(v)}$, and if uv is a non-edge, then both sides are zero.

So by Kelly's lemma, the chain is reversible with (unique) stationary distribution assigning $d(v)/(2m)$ to each vertex v.

Reversibility

The distribution assigning $d(v)/(2m)$ to each vertex v satisfies the detailed balance equations. Indeed, if uv is an edge, then $\frac{d(u)}{2m} \cdot \frac{1}{d(u)} = \frac{1}{2m} = \frac{d(v)}{2m} \cdot \frac{1}{d(v)}$, and if uv is a non-edge, then both sides are zero.

So by Kelly's lemma, the chain is reversible with (unique) stationary distribution assigning $d(v)/(2m)$ to each vertex v.

The frequency of being at a given vertex v is proportionate to the degree $d(v)$ of v.

Reversibility

The distribution assigning $d(v)/(2m)$ to each vertex v satisfies the detailed balance equations. Indeed, if uv is an edge, then $\frac{d(u)}{2m} \cdot \frac{1}{d(u)} = \frac{1}{2m} = \frac{d(v)}{2m} \cdot \frac{1}{d(v)}$, and if uv is a non-edge, then both sides are zero.

So by Kelly's lemma, the chain is reversible with (unique) stationary distribution assigning $d(v)/(2m)$ to each vertex v.

The frequency of being at a given vertex v is proportionate to the degree $d(v)$ of v. The mean recurrence time to v is $2m/d(v)$.

If the graph is not bipartite, then there is a number $0 < c < 1$ such that the convergence of P^n to W has exponential speed with rate at most μ.

If the graph is not bipartite, then there is a number $0 < c < 1$ such that the convergence of P^n to W has exponential speed with rate at most μ. Small mixing time (at most $O(\log n)$).

Mixing time, cover time

If the graph is not bipartite, then there is a number $0 < c < 1$ such that the convergence of P^n to W has exponential speed with rate at most μ. Small mixing time (at most $O(\log n)$).

Cover time : how long do we have to wait before all nodes are visited ?

Mixing time, cover time

If the graph is not bipartite, then there is a number $0 < c < 1$ such that the convergence of P^n to W has exponential speed with rate at most μ. Small mixing time (at most $O(\log n)$).

Cover time : how long do we have to wait before all nodes are visited ?

Theorem (Feige)

The cover time from any starting node in a graph with n nodes is at least $(1 + o(1))n \log n$ and at most $(4/27 + o(1))n^3$. The cover time of a regular graph on n nodes is at most $2n^2$.

Expanders

Given $\varepsilon > 0, k \in \mathbb{N}$.

Given $\varepsilon > 0, k \in \mathbb{N}$. A finite k-regular graph G with spectrum of the associated random walk $1 = \lambda_1 > \lambda_2 > \cdots > \lambda_n$ is a one-sided expander if $\lambda_2 < 1 - \varepsilon$, Given $\varepsilon > 0, k \in \mathbb{N}$. A finite k-regular graph G with spectrum of the associated random walk $1 = \lambda_1 > \lambda_2 > \cdots > \lambda_n$ is a one-sided expander if $\lambda_2 < 1 - \varepsilon$, and two sided if furthermore $\lambda_n \geq -(1 - \varepsilon)$.

Given $\varepsilon > 0, k \in \mathbb{N}$. A finite k-regular graph G with spectrum of the associated random walk $1 = \lambda_1 > \lambda_2 > \cdots > \lambda_n$ is a one-sided expander if $\lambda_2 < 1 - \varepsilon$, and two sided if furthermore $\lambda_n \geq -(1 - \varepsilon)$.

A sequence $G_i = (V_i, E_i)$ of finite k-regular graphs is a one-sided (resp. two-sided) expander family if there is an $\varepsilon > 0$ such that G_i is a one-sided (resp. two-sided) ε-expander for all sufficiently large i.

According to the Perron-Frobenius theorem, every connected graph is a one-sided expander with some ε.

According to the Perron-Frobenius theorem, every connected graph is a one-sided expander with some ε.

Furthermore, if the graph is connected and not bipartite, then it is two-sided. (See the Perron-Frobenius and Hoffman theorems.)

According to the Perron-Frobenius theorem, every connected graph is a one-sided expander with some ε.

Furthermore, if the graph is connected and not bipartite, then it is two-sided. (See the Perron-Frobenius and Hoffman theorems.)

These are used in computer science for generating random numbers, de-randomizing non-deterministic algorithms, and constructing good error-correcting codes.

Exercises

Exercises

1. Show that the total variation distance from any distribution $\underline{w}^*$ attains its maximum on a Dirac measure.

Exercises

1. Show that the total variation distance from any distribution $\underline{w}^*$ attains its maximum on a Dirac measure.
2. Prove that the number of (p, q) riffle shuffles (when the deck of N cards is cut into piles of size p and q) is $\binom{p+q}{q}$.

Exercises

1. Show that the total variation distance from any distribution $\underline{w}^*$ attains its maximum on a Dirac measure.
2. Prove that the number of (p, q) riffle shuffles (when the deck of N cards is cut into piles of size p and q) is $\binom{p+q}{q}$.
3. Using the result of Problem 2, show that the total number of riffle shuffles is $2^N - N$. (Watch out for the identity permutation!)

Exercises

1. Show that the total variation distance from any distribution $\underline{w}^*$ attains its maximum on a Dirac measure.
2. Prove that the number of (p, q) riffle shuffles (when the deck of N cards is cut into piles of size p and q) is $\binom{p+q}{q}$.
3. Using the result of Problem 2, show that the total number of riffle shuffles is $2^N - N$. (Watch out for the identity permutation!)
4. Prove that the transition matrix of a random walk on a simple, connected undirected graph is symmetrical iff the graph is regular.

1. Show that the total variation distance from any distribution $\underline{w}^*$ attains its maximum on a Dirac measure.
2. Prove that the number of (p, q) riffle shuffles (when the deck of N cards is cut into piles of size p and q) is $\binom{p+q}{q}$.
3. Using the result of Problem 2, show that the total number of riffle shuffles is $2^N - N$. (Watch out for the identity permutation!)
4. Prove that the transition matrix of a random walk on a simple, connected undirected graph is symmetrical iff the graph is regular.
5. Prove that the stationary distribution of the transition matrix of a random walk on a simple, connected undirected graph is $\underline{1}^*$ iff the graph is regular.

Exercises

1. Show that the total variation distance from any distribution $\underline{w}^*$ attains its maximum on a Dirac measure.
2. Prove that the number of (p, q) riffle shuffles (when the deck of N cards is cut into piles of size p and q) is $\binom{p+q}{q}$.
3. Using the result of Problem 2, show that the total number of riffle shuffles is $2^N - N$. (Watch out for the identity permutation!)
4. Prove that the transition matrix of a random walk on a simple, connected undirected graph is symmetrical iff the graph is regular.
5. Prove that the stationary distribution of the transition matrix of a random walk on a simple, connected undirected graph is $\underline{1}^*$ iff the graph is regular.