Monte Carlo Method

Meghana Ramakumar[1]

[1] Lane Department of Computer Science and Electrical Engineering
West Virginia University

10 April, 2012

Outline

Outline

Outline

Definition

Approximate Sampling to Approximate Counting

Definition

Approximate Sampling to Approximate Counting

If a solution to a self-reducible combinatorial problem can be sampled, we can construct a randomized algorithm which counts the number of solutions.

Definition

Approximate Sampling to Approximate Counting

If a solution to a self-reducible combinatorial problem can be sampled, we can construct a randomized algorithm which counts the number of solutions.

Definition

Definition

Approximate Sampling to Approximate Counting

If a solution to a self-reducible combinatorial problem can be sampled, we can construct a randomized algorithm which counts the number of solutions.

Definition

Let w be the (random) output of a sampling algorithm for a finite sample space Ω. The sampling algorithm generates an $\varepsilon-$uniform sample of Ω if, for any subset S of Ω,

Definition

Approximate Sampling to Approximate Counting

If a solution to a self-reducible combinatorial problem can be sampled, we can construct a randomized algorithm which counts the number of solutions.

Definition

Let w be the (random) output of a sampling algorithm for a finite sample space Ω. The sampling algorithm generates an $\varepsilon-$uniform sample of Ω if, for any subset S of Ω,

$$\left| P(w \in S) - \frac{|S|}{|\Omega|} \right| \leq \varepsilon.$$

Definition

Approximate Sampling to Approximate Counting

If a solution to a self-reducible combinatorial problem can be sampled, we can construct a randomized algorithm which counts the number of solutions.

Definition

Let w be the (random) output of a sampling algorithm for a finite sample space Ω. The sampling algorithm generates an $\varepsilon-$uniform sample of Ω if, for any subset S of Ω,

$$\left| P(w \in S) - \frac{|S|}{|\Omega|} \right| \leq \varepsilon.$$

A sampling algorithm is a fully polynomial almost uniform sample (*FPAUS*) for a problem if, given an input x and a parameter $\varepsilon > 0$, it generates an ε - uniform sample of $\Omega(x)$ and runs in time that is polynomial in $\ln \varepsilon^{-1}$ and the size of the input x.

Outline

Example

Example

Example

Example

FPAUS for independent sets.

Example

Example

FPAUS for independent sets.

Question

What is an independent set?

Example

Example

FPAUS for independent sets.

Question

What is an independent set?
It is a subset of vertices in a graph, such that no two of which are adjacent.

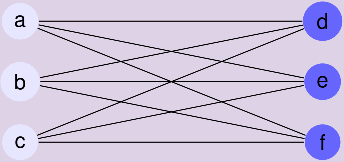
Example

Example

FPAUS for independent sets.

Question

What is an independent set?
It is a subset of vertices in a graph, such that no two of which are adjacent.



Example continued

Example

Example continued

Example

Input : Graph $G = (V, E)$ and parameter ε.

Example continued

Example

Input : Graph $G = (V, E)$ and parameter ε.
Sample space : All independent sets in G.

Example continued

Example

Input : Graph $G = (V, E)$ and parameter ε.
Sample space : All independent sets in G.
Output : ε-uniform sample of the independent sets.

Example continued

Example

Input : Graph $G = (V, E)$ and parameter ε.
Sample space : All independent sets in G.
Output : ε-uniform sample of the independent sets.
Goal : To show that, given an FPAUS for independent sets, we can construct an FPRAS for counting the number of independent sets.

Example

Continued

Assumption:

Example

Continued

Assumption:
Input graph G has m edges.

Example

Continued

Assumption:
Input graph G has m edges.
Let $e_1, e_2, e_3, \ldots, e_m$ be an arbitrary ordering of the edges.

Example

Continued

Assumption:
Input graph G has m edges.
Let $e_1, e_2, e_3, \ldots, e_m$ be an arbitrary ordering of the edges.
E_i be the set of the first i edges in E.

Example

Continued

Assumption:
Input graph G has m edges.
Let $e_1, e_2, e_3, \ldots, e_m$ be an arbitrary ordering of the edges.
E_i be the set of the first i edges in E.
Let $G_i = (V, E_i)$.

Example

Continued

Assumption:
Input graph G has m edges.
Let $e_1, e_2, e_3, \ldots, e_m$ be an arbitrary ordering of the edges.
E_i be the set of the first i edges in E.
Let $G_i = (V, E_i)$.
G_{i-1} is obtained from G_i by removing a single edge.

Example

Continued

Assumption:
Input graph G has m edges.
Let $e_1, e_2, e_3, \ldots, e_m$ be an arbitrary ordering of the edges.
E_i be the set of the first i edges in E.
Let $G_i = (V, E_i)$.
G_{i-1} is obtained from G_i by removing a single edge.

Let $\Omega(G_i)$ denote set of independent sets in G_i.

Example

Continued

Assumption:
Input graph G has m edges.
Let $e_1, e_2, e_3, \ldots, e_m$ be an arbitrary ordering of the edges.
E_i be the set of the first i edges in E.
Let $G_i = (V, E_i)$.
G_{i-1} is obtained from G_i by removing a single edge.

Let $\Omega(G_i)$ denote set of independent sets in G_i.
The number of independent sets in G can be expressed as,

$$|\Omega(G)| = \frac{|\Omega(G_m)|}{|\Omega(G_{m-1})|} \times \frac{|\Omega(G_{m-1})|}{|\Omega(G_{m-2})|} \times \frac{|\Omega(G_{m-2})|}{|\Omega(G_{m-3})|} \times \cdots \times \frac{|\Omega(G_1)|}{|\Omega(G_0)|} \times |\Omega(G_0)|.$$

Example

Continued

Assumption:
Input graph G has m edges.
Let $e_1, e_2, e_3, \ldots, e_m$ be an arbitrary ordering of the edges.
E_i be the set of the first i edges in E.
Let $G_i = (V, E_i)$.
G_{i-1} is obtained from G_i by removing a single edge.

Let $\Omega(G_i)$ denote set of independent sets in G_i.
The number of independent sets in G can be expressed as,

$$|\Omega(G)| = \frac{|\Omega(G_m)|}{|\Omega(G_{m-1})|} \times \frac{|\Omega(G_{m-1})|}{|\Omega(G_{m-2})|} \times \frac{|\Omega(G_{m-2})|}{|\Omega(G_{m-3})|} \times \cdots \times \frac{|\Omega(G_1)|}{|\Omega(G_0)|} \times |\Omega(G_0)|.$$

G_0 has no edges, every subset of V is an independent set,

Example

Continued

Assumption:
Input graph G has m edges.
Let $e_1, e_2, e_3, \ldots, e_m$ be an arbitrary ordering of the edges.
E_i be the set of the first i edges in E.
Let $G_i = (V, E_i)$.
G_{i-1} is obtained from G_i by removing a single edge.

Let $\Omega(G_i)$ denote set of independent sets in G_i.
The number of independent sets in G can be expressed as,

$$|\Omega(G)| = \frac{|\Omega(G_m)|}{|\Omega(G_{m-1})|} \times \frac{|\Omega(G_{m-1})|}{|\Omega(G_{m-2})|} \times \frac{|\Omega(G_{m-2})|}{|\Omega(G_{m-3})|} \times \cdots \times \frac{|\Omega(G_1)|}{|\Omega(G_0)|} \times |\Omega(G_0)|.$$

G_0 has no edges, every subset of V is an independent set,therefore
$\Omega(G_0) =$

Example

Continued

Assumption:
Input graph G has m edges.
Let $e_1, e_2, e_3, \ldots, e_m$ be an arbitrary ordering of the edges.
E_i be the set of the first i edges in E.
Let $G_i = (V, E_i)$.
G_{i-1} is obtained from G_i by removing a single edge.

Let $\Omega(G_i)$ denote set of independent sets in G_i.
The number of independent sets in G can be expressed as,

$$|\Omega(G)| = \frac{|\Omega(G_m)|}{|\Omega(G_{m-1})|} \times \frac{|\Omega(G_{m-1})|}{|\Omega(G_{m-2})|} \times \frac{|\Omega(G_{m-2})|}{|\Omega(G_{m-3})|} \times \cdots \times \frac{|\Omega(G_1)|}{|\Omega(G_0)|} \times |\Omega(G_0)|.$$

G_0 has no edges, every subset of V is an independent set,therefore
$\Omega(G_0) = 2^n$.

Example

Continued

Example

Continued

To estimate $|\Omega(G)|$, we need good estimates for each of the ratios

Example

Continued

To estimate $|\Omega(G)|$, we need good estimates for each of the ratios

$r_i = \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}$, $i = 1, 2, \ldots m$.

Example

Continued

To estimate $|\Omega(G)|$, we need good estimates for each of the ratios

$r_i = \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}$, $i = 1, 2, \ldots m$.

More formally, we will develop estimates $\tilde{r}_i$ for the ratios r_i, then estimates for the number of independent sets in G will be,

Example

Continued

To estimate $|\Omega(G)|$, we need good estimates for each of the ratios

$r_i = \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}$, $i = 1, 2, \ldots m$.

More formally, we will develop estimates $\tilde{r}_i$ for the ratios r_i, then estimates for the number of independent sets in G will be,

$$2^n \prod_{i=1}^{m} \tilde{r}_i$$

Example

Continued

To estimate $|\Omega(G)|$, we need good estimates for each of the ratios

$r_i = \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}$, $i = 1, 2, \ldots m$.

More formally, we will develop estimates $\tilde{r}_i$ for the ratios r_i, then estimates for the number of independent sets in G will be,

$$2^n \prod_{i=1}^{m} \tilde{r}_i$$

while the actual number is

Example

Continued

To estimate $|\Omega(G)|$, we need good estimates for each of the ratios

$r_i = \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}$, $i = 1, 2, \ldots m.$

More formally, we will develop estimates $\tilde{r}_i$ for the ratios r_i, then estimates for the number of independent sets in G will be,

$$2^n \prod_{i=1}^{m} \tilde{r}_i$$

while the actual number is

$$|\Omega(G)| = 2^n \prod_{i=1}^{m} r_i.$$

Example

Continued

To evaluate the error in our estimate, we need to bound the ratio

Example

Continued

To evaluate the error in our estimate, we need to bound the ratio

$$R = \prod_{i=1}^{m} \frac{\tilde{r}_i}{r_i}.$$

Example

Continued

To evaluate the error in our estimate, we need to bound the ratio

$$R = \prod_{i=1}^{m} \frac{\tilde{r}_i}{r_i}.$$

To have an (ε, δ)-approximation, we want $P(|R - 1| \leq \varepsilon) \geq 1 - \delta$.

Lemmas

Lemma

Suppose that for all i, $1 \leq i \leq m$, $\tilde{r}_i$ is an $(\frac{\varepsilon}{2 \cdot m}, \frac{\delta}{m})$-approximation for r_i. Then,

Lemmas

Lemma

Suppose that for all i, $1 \leq i \leq m$, $\tilde{r}_i$ is an $(\frac{\varepsilon}{2 \cdot m}, \frac{\delta}{m})$-approximation for r_i. Then,

$$P(|R - 1| \leq \varepsilon) \geq 1 - \delta.$$

Lemmas

Lemma

Suppose that for all i, $1 \leq i \leq m$, $\tilde{r}_i$ is an $(\frac{\varepsilon}{2 \cdot m}, \frac{\delta}{m})$-approximation for r_i. Then,

$$P(|R - 1| \leq \varepsilon) \geq 1 - \delta.$$

Proof

For each $1 \leq i \leq m$, we have

$$\begin{aligned} P(|\tilde{r}_i - r_i| \leq \frac{\varepsilon}{2 \cdot m} \cdot r_i) &\geq 1 - \frac{\delta}{m} \\ P(|\tilde{r}_i - r_i| > \frac{\varepsilon}{2 \cdot m} \cdot r_i) &< \frac{\delta}{m} \end{aligned}$$

Lemmas

Lemma

Suppose that for all i, $1 \leq i \leq m$, $\tilde{r}_i$ is an $(\frac{\varepsilon}{2 \cdot m}, \frac{\delta}{m})$-approximation for r_i. Then,

$$P(|R - 1| \leq \varepsilon) \geq 1 - \delta.$$

Proof

For each $1 \leq i \leq m$, we have

$$\begin{aligned} P(|\tilde{r}_i - r_i| \leq \frac{\varepsilon}{2 \cdot m} \cdot r_i) &\geq 1 - \frac{\delta}{m} \\ P(|\tilde{r}_i - r_i| > \frac{\varepsilon}{2 \cdot m} \cdot r_i) &< \frac{\delta}{m} \end{aligned}$$

By union bound, the probability that $|\tilde{r}_i - r_i| > (\frac{\varepsilon}{2 \cdot m}) \cdot r_i$ for any i is at most δ;
Therefore $|\tilde{r}_i - r_i| \leq (\frac{\varepsilon}{2 \cdot m} \cdot r_i)$ for all i with probability at least $1 - \delta$. Equivalently,

Lemmas

Lemma

Suppose that for all i, $1 \leq i \leq m$, $\tilde{r}_i$ is an $(\frac{\varepsilon}{2 \cdot m}, \frac{\delta}{m})$-approximation for r_i. Then,

$$P(|R - 1| \leq \varepsilon) \geq 1 - \delta.$$

Proof

For each $1 \leq i \leq m$, we have

$$\begin{aligned} P(|\tilde{r}_i - r_i| \leq \frac{\varepsilon}{2 \cdot m} \cdot r_i) &\geq 1 - \frac{\delta}{m} \\ P(|\tilde{r}_i - r_i| > \frac{\varepsilon}{2 \cdot m} \cdot r_i) &< \frac{\delta}{m} \end{aligned}$$

By union bound, the probability that $|\tilde{r}_i - r_i| > (\frac{\varepsilon}{2 \cdot m}) \cdot r_i$ for any i is at most δ; Therefore $|\tilde{r}_i - r_i| \leq (\frac{\varepsilon}{2 \cdot m} \cdot r_i)$ for all i with probability at least $1 - \delta$. Equivalently,

$$1 - \frac{\varepsilon}{2 \cdot m} \leq \frac{\tilde{r}_i}{r_i} \leq 1 + \frac{\varepsilon}{2 \cdot m}$$

Lemmas

Proof continued

When these bounds hold for all i, we can combine them to obtain,

Lemmas

Proof continued

When these bounds hold for all i, we can combine them to obtain,

$$1-\varepsilon \;\leq\; \left(1-\frac{\varepsilon}{2\cdot m}\right)^m \;\leq\; \prod_{i=1}^{m}\frac{\tilde{r}_i}{r_i} \;\leq\; \left(1+\frac{\varepsilon}{2\cdot m}\right)^m \;\leq\; 1+\varepsilon$$

Lemmas

Proof continued

When these bounds hold for all i, we can combine them to obtain,

$$1-\varepsilon \le (1-\frac{\varepsilon}{2\cdot m})^m \le \prod_{i=1}^{m}\frac{\tilde{r}_i}{r_i} \le (1+\frac{\varepsilon}{2\cdot m})^m \le 1+\varepsilon$$

Notes

We need a method for obtaining an $(\frac{\varepsilon}{2\cdot m}, \frac{\delta}{m})$- approximation for the r_i.
We estimate each of these ratios by a Monte Carlo algorithm that uses *FPAUS* for sampling independent sets.

Lemmas

Proof continued

When these bounds hold for all i, we can combine them to obtain,

$$1 - \varepsilon \;\leq\; (1 - \frac{\varepsilon}{2 \cdot m})^m \;\leq\; \prod_{i=1}^{m} \frac{\tilde{r}_i}{r_i} \;\leq\; (1 + \frac{\varepsilon}{2 \cdot m})^m \;\leq\; 1 + \varepsilon$$

Notes

We need a method for obtaining an $(\frac{\varepsilon}{2 \cdot m}, \frac{\delta}{m})$- approximation for the r_i.
We estimate each of these ratios by a Monte Carlo algorithm that uses *FPAUS* for sampling independent sets.
To estimate r_i, we sample independent sets in G_{i-1} and compute the fraction of these sets that are also independent sets in G_i, as shown in the following algorithm.

Algorithm

Estimating r_i:

Input: Graphs $G_{i-1} = (V, E_{i-1})$ and $G_i = (V, E_i)$.
Output: $\tilde{r}_i$ = an approximation of r_i.

1: $X \leftarrow 0$
2: **repeat**
3: Generate an $(\frac{\varepsilon}{6 \cdot m})-$ uniform sample from $\Omega(G_{i-1})$.
4: If the sample is an independent set in G_i, let $X \leftarrow X + 1$.
5: **until** $M = \lceil 1296 m^2 \varepsilon^{-2} \ln(\frac{2 \cdot m}{\delta}) \rceil$ independent trials
6: **return** $\tilde{r}_i \leftarrow \frac{X}{M}$

Algorithm 2.1: Estimating r_i

Lemmas

Lemma

When $m \geq 1$ and $0 < \varepsilon \leq 1$, the procedure for estimating r_i yields an $(\frac{\varepsilon}{2 \cdot m}, \frac{\delta}{m})$-approximation for r_i.

Lemmas

Lemma

When $m \geq 1$ and $0 < \varepsilon \leq 1$, the procedure for estimating r_i yields an $(\frac{\varepsilon}{2 \cdot m}, \frac{\delta}{m})$-approximation for r_i.

Proof

Read pages 261-262 in the book.

Lemmas

Lemma

When $m \geq 1$ and $0 < \varepsilon \leq 1$, the procedure for estimating r_i yields an $(\frac{\varepsilon}{2 \cdot m}, \frac{\delta}{m})$-approximation for r_i.

Proof

Read pages 261-262 in the book.

Proving involves

We first show that r_i is not too small. $\Omega(G_i) \subseteq \Omega(G_{i-1})$.

Lemmas

Lemma

When $m \geq 1$ and $0 < \varepsilon \leq 1$, the procedure for estimating r_i yields an $(\frac{\varepsilon}{2 \cdot m}, \frac{\delta}{m})$-approximation for r_i.

Proof

Read pages 261-262 in the book.

Proving involves

We first show that r_i is not too small. $\Omega(G_i) \subseteq \Omega(G_{i-1})$.
To bound the size of the set in $\Omega(G_{i-1}) \setminus \Omega(G_i)$,

Lemmas

Lemma

When $m \geq 1$ and $0 < \varepsilon \leq 1$, the procedure for estimating r_i yields an $(\frac{\varepsilon}{2 \cdot m}, \frac{\delta}{m})$-approximation for r_i.

Proof

Read pages 261-262 in the book.

Proving involves

We first show that r_i is not too small. $\Omega(G_i) \subseteq \Omega(G_{i-1})$.
To bound the size of the set in $\Omega(G_{i-1}) \setminus \Omega(G_i)$,
associate each $I \in \Omega(G_{i-1}) \setminus \Omega(G_i)$ with an independent set $I \setminus v \in \Omega(G_i)$.

Lemmas

Lemma

When $m \geq 1$ and $0 < \varepsilon \leq 1$, the procedure for estimating r_i yields an $(\frac{\varepsilon}{2 \cdot m}, \frac{\delta}{m})$-approximation for r_i.

Proof

Read pages 261-262 in the book.

Proving involves

We first show that r_i is not too small. $\Omega(G_i) \subseteq \Omega(G_{i-1})$.
To bound the size of the set in $\Omega(G_{i-1}) \setminus \Omega(G_i)$,
associate each $I \in \Omega(G_{i-1}) \setminus \Omega(G_i)$ with an independent set $I \setminus v \in \Omega(G_i)$.
We have,

$$r_i = \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|} = \frac{|\Omega(G_i)|}{|\Omega(G_i)| + |\Omega(G_{i-1}) \setminus \Omega(G_i)|} \geq \frac{1}{2}.$$

Lemmas

continued

Consider M samples and let X_k be

$$X_k = \begin{cases} 1 & \text{if the } k^{\text{th}} \text{ sample is in } \Omega(G_i) \\ 0 & \text{otherwise} \end{cases}$$

Lemmas

continued

Consider M samples and let X_k be

$$X_k = \begin{cases} 1 & \text{if the } k^{\text{th}} \text{ sample is in } \Omega(G_i) \\ 0 & \text{otherwise} \end{cases}$$

Since our samples are generated by an $(\frac{\varepsilon}{6 \cdot m})$- uniform sampler, by our previous definition each X_i must satisfy,

$$|P(X_k = 1) - \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}| \leq \frac{\varepsilon}{6 \cdot m}.$$

Lemmas

continued

Consider M samples and let X_k be

$$X_k = \begin{cases} 1 & \text{if the } k^{\text{th}} \text{ sample is in } \Omega(G_i) \\ 0 & \text{otherwise} \end{cases}$$

Since our samples are generated by an $(\frac{\varepsilon}{6 \cdot m})$- uniform sampler, by our previous definition each X_i must satisfy,

$$|P(X_k = 1) - \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}| \leq \frac{\varepsilon}{6 \cdot m}.$$

Since X_k are indicator random variables and further by linearity of expectations, we get

$$|\mathbf{E}[\tilde{r}_i] - r_i| \quad = \quad |\mathbf{E}[\frac{\sum_{i=1}^{M} X_k}{M}] - \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}| \leq \frac{\varepsilon}{6 \cdot m}.$$

Lemmas

continued

Consider M samples and let X_k be

$$X_k = \begin{cases} 1 & \text{if the } k^{\text{th}} \text{ sample is in } \Omega(G_i) \\ 0 & \text{otherwise} \end{cases}$$

Since our samples are generated by an $(\frac{\varepsilon}{6 \cdot m})$- uniform sampler, by our previous definition each X_i must satisfy,

$$|P(X_k = 1) - \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}| \leq \frac{\varepsilon}{6 \cdot m}.$$

Since X_k are indicator random variables and further by linearity of expectations, we get

$$|\mathbf{E}[\tilde{r}_i] - r_i| \quad = \quad |\mathbf{E}[\frac{\sum_{i=1}^{M} X_k}{M}] - \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}| \leq \frac{\varepsilon}{6 \cdot m}.$$

Lemma is completed by combining

(a) $\mathbf{E}[\tilde{r}_i]$ is close to r_i and

(b) $\tilde{r}_i$ will be close to $\mathbf{E}[\tilde{r}_i]$ for a sufficiently large number of samples .

Lemmas

continued

Consider M samples and let X_k be

$$X_k = \begin{cases} 1 & \text{if the } k^{\text{th}} \text{ sample is in } \Omega(G_i) \\ 0 & \text{otherwise} \end{cases}$$

Since our samples are generated by an $(\frac{\varepsilon}{6 \cdot m})$- uniform sampler, by our previous definition each X_i must satisfy,

$$|P(X_k = 1) - \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}| \leq \frac{\varepsilon}{6 \cdot m}.$$

Since X_k are indicator random variables and further by linearity of expectations, we get

$$|\mathbf{E}[\tilde{r}_i] - r_i| \quad = \quad |\mathbf{E}[\frac{\sum_{i=1}^{M} X_k}{M}] - \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}| \leq \frac{\varepsilon}{6 \cdot m}.$$

Lemma is completed by combining

(a) $\mathbf{E}[\tilde{r}_i]$ is close to r_i and

(b) $\tilde{r}_i$ will be close to $\mathbf{E}[\tilde{r}_i]$ for a sufficiently large number of samples .

Using the above and $r_i \geq 1/2$ gives the desired $(\frac{\varepsilon}{2 \cdot m}, \frac{\delta}{m})$- approximation.

Theorems

Theorem

Given a fully polynomial almost uniform sampler (*FPAUS*) for independent sets in any graph, we can construct a fully polynomial randomized approximation scheme (*FPRAS*) for the number of independent sets in a graph G.

Theorem

Given a fully polynomial almost uniform sampler (*FPAUS*) for independent sets in any graph with maximum degree at most Δ, we can construct a fully polynomial randomized approximation scheme (*FPRAS*) for the number of independent sets in a graph G with maximum degree at most Δ.

Outline

Introduction

Definition

The Markov chain Monte Carlo (*MCMC*) method provides a very general approach to sampling from a desired probability distribution.

Introduction

Definition

The Markov chain Monte Carlo (*MCMC*) method provides a very general approach to sampling from a desired probability distribution.

Where MCMC is used

(i) Data Mining and Machine Learning
(ii) Bayesian methods
(iii) Biological and generic research

Recall Markov chain

Questions

Recall Markov chain

Questions

(i) Irreducible?

Recall Markov chain

Questions

(i) Irreducible? Markov chain is said to be irreducible if its state space is a single communicating class; i.e. , if it is possible to get to any state from any state.

Recall Markov chain

Questions

(i) Irreducible? Markov chain is said to be irreducible if its state space is a single communicating class; i.e. , if it is possible to get to any state from any state.

(ii) Aperiodic?

Recall Markov chain

Questions

(i) Irreducible? Markov chain is said to be irreducible if its state space is a single communicating class; i.e. , if it is possible to get to any state from any state.

(ii) Aperiodic? A state is periodic if it can only return to itself after a fixed number of transitions greater than 1 (or multiple of a fixed number). A state that is not periodic is aperiodic. A Markov chain is aperiodic if all states of the chain are aperiodic.

Recall Markov chain

Questions

(i) Irreducible? Markov chain is said to be irreducible if its state space is a single communicating class; i.e. , if it is possible to get to any state from any state.

(ii) Aperiodic? A state is periodic if it can only return to itself after a fixed number of transitions greater than 1 (or multiple of a fixed number). A state that is not periodic is aperiodic. A Markov chain is aperiodic if all states of the chain are aperiodic.

(iii) Ergodic Markov chain?

Recall Markov chain

Questions

(i) Irreducible? Markov chain is said to be irreducible if its state space is a single communicating class; i.e. , if it is possible to get to any state from any state.

(ii) Aperiodic? A state is periodic if it can only return to itself after a fixed number of transitions greater than 1 (or multiple of a fixed number). A state that is not periodic is aperiodic. A Markov chain is aperiodic if all states of the chain are aperiodic.

(iii) Ergodic Markov chain? A state i is said to be ergodic if it is aperiodic and positive recurrent. If all states in an irreducible Markov chain are ergodic, then the chain is said to be ergodic.

Recall Markov chain

Questions

(i) Irreducible? Markov chain is said to be irreducible if its state space is a single communicating class; i.e. , if it is possible to get to any state from any state.

(ii) Aperiodic? A state is periodic if it can only return to itself after a fixed number of transitions greater than 1 (or multiple of a fixed number). A state that is not periodic is aperiodic. A Markov chain is aperiodic if all states of the chain are aperiodic.

(iii) Ergodic Markov chain? A state i is said to be ergodic if it is aperiodic and positive recurrent. If all states in an irreducible Markov chain are ergodic, then the chain is said to be ergodic.

(iv) Stationary distribution?

Recall Markov chain

Questions

(i) Irreducible? Markov chain is said to be irreducible if its state space is a single communicating class; i.e. , if it is possible to get to any state from any state.

(ii) Aperiodic? A state is periodic if it can only return to itself after a fixed number of transitions greater than 1 (or multiple of a fixed number). A state that is not periodic is aperiodic. A Markov chain is aperiodic if all states of the chain are aperiodic.

(iii) Ergodic Markov chain? A state i is said to be ergodic if it is aperiodic and positive recurrent. If all states in an irreducible Markov chain are ergodic, then the chain is said to be ergodic.

(iv) Stationary distribution? A stationary distribution of a Markov chain is a probability distribution $\bar{\pi}$ such that $\bar{\pi} = \bar{\pi} \cdot P$

Recall Markov chain continued

Theorem (for stationary distribution)

Consider a finite, irreducible, and ergodic Markov chain with transition matrix P. If there are nonnegative numbers $\bar{\pi} = (\pi_0, ...\pi_n)$ such that $\sum_{i=0}^{n} \pi_i = 1$ and if, for any pair of states i, j,

$$\pi_i \cdot P_{i,j} = \pi_j \cdot P_{j,i},$$

then $\bar{\pi}_i$ is the stationary distribution corresponding to P.

Recall Markov chain continued

Theorem (for stationary distribution)

Consider a finite, irreducible, and ergodic Markov chain with transition matrix P. If there are nonnegative numbers $\bar{\pi} = (\pi_0, ...\pi_n)$ such that $\sum_{i=0}^{n} \pi_i = 1$ and if, for any pair of states i, j,

$$\pi_i \cdot P_{i,j} = \pi_j \cdot P_{j,i},$$

then $\bar{\pi}_i$ is the stationary distribution corresponding to P.

Theorem

A random walk on G converges to a stationary distribution $\bar{\pi}$, where

$$\pi_v = \frac{d(v)}{2 \cdot |E|}.$$

In a stationary distribution of a random walk, the probability of a vertex is proportional to the degree of the vertex.

MCMC Continued

Basic Idea

Define an ergodic Markov chain whose set of states is the sample space and whose stationary distribution is the required sampling distribution.

MCMC Continued

Basic Idea

Define an ergodic Markov chain whose set of states is the sample space and whose stationary distribution is the required sampling distribution.
Let $X_0, X_1, \ldots, X_n$ be a run of the chain.

MCMC Continued

Basic Idea

Define an ergodic Markov chain whose set of states is the sample space and whose stationary distribution is the required sampling distribution.
Let $X_0, X_1, \ldots, X_n$ be a run of the chain.
Markov chain converges to the stationary distribution from any starting state X_0, after a sufficiently large number of steps r, the distribution of the state X_r will be close to stationary distribution and can be used as a sample.

MCMC Continued

Basic Idea

Define an ergodic Markov chain whose set of states is the sample space and whose stationary distribution is the required sampling distribution.
Let $X_0, X_1, \ldots, X_n$ be a run of the chain.
Markov chain converges to the stationary distribution from any starting state X_0, after a sufficiently large number of steps r, the distribution of the state X_r will be close to stationary distribution and can be used as a sample.
Similarly starting from X_r, $X_{2.r}$ can be used as a sample and so on.

MCMC Continued

Basic Idea

Define an ergodic Markov chain whose set of states is the sample space and whose stationary distribution is the required sampling distribution.
Let $X_0, X_1, \ldots, X_n$ be a run of the chain.
Markov chain converges to the stationary distribution from any starting state X_0, after a sufficiently large number of steps r, the distribution of the state X_r will be close to stationary distribution and can be used as a sample.
Similarly starting from X_r, $X_{2.r}$ can be used as a sample and so on.
Therefore the sequence of states $X_r, X_{2.r}, X_{3.r}, \ldots$ can be used as the almost independent samples from the stationary distribution of the Markov chain.

MCMC Continued

Lemma

For a finite state space Ω and neighborhood structure $\{N(X) | x \in \Omega\}$, let $N = \max_{x \in \Omega} |N(x)|$.

MCMC Continued

Lemma

For a finite state space Ω and neighborhood structure $\{N(X)|x \in \Omega\}$, let $N = \max_{x \in \Omega} |N(x)|$.
Let M be any number such that $M \geq N$. Consider a Markov chain where

$$P_{x,y} = \begin{cases} \frac{1}{M}, & \text{if } x \neq y \text{ and } y \in N(x), \\ & \end{cases}$$

MCMC Continued

Lemma

For a finite state space Ω and neighborhood structure $\{N(X)|x \in \Omega\}$, let $N = \max_{x \in \Omega} |N(x)|$.
Let M be any number such that $M \geq N$. Consider a Markov chain where

$$P_{x,y} = \begin{cases} \frac{1}{M}, & \text{if } x \neq y \text{ and } y \in N(x), \\ 0, & \text{if } x \neq y \text{ and } y \notin N(x), \end{cases}$$

MCMC Continued

Lemma

For a finite state space Ω and neighborhood structure $\{N(X)|x \in \Omega\}$, let $N = \max_{x\in\Omega} |N(x)|$.
Let M be any number such that $M \geq N$. Consider a Markov chain where

$$P_{x,y} = \begin{cases} \frac{1}{M}, & \text{if } x \neq y \text{ and } y \in N(x), \\ 0, & \text{if } x \neq y \text{ and } y \notin N(x), \\ 1 - \frac{N(x)}{M}, & \text{if } x = y. \end{cases}$$

MCMC Continued

Lemma

For a finite state space Ω and neighborhood structure $\{N(X)|x \in \Omega\}$, let $N = \max_{x \in \Omega} |N(x)|$.
Let M be any number such that $M \geq N$. Consider a Markov chain where

$$P_{x,y} = \begin{cases} \frac{1}{M}, & \text{if } x \neq y \text{ and } y \in N(x), \\ 0, & \text{if } x \neq y \text{ and } y \notin N(x), \\ 1 - \frac{N(x)}{M}, & \text{if } x = y. \end{cases}$$

If this chain is irreducible and aperiodic, then the stationary distribution is the uniform distribution.

MCMC Continued

Proof

We show that chain is time reversible and apply Theorem (for stationary distribution).

MCMC Continued

Proof

We show that chain is time reversible and apply Theorem (for stationary distribution).
For any $x \neq y$, if $\pi_x = \pi_y$ then,

MCMC Continued

Proof

We show that chain is time reversible and apply Theorem (for stationary distribution).
For any $x \neq y$, if $\pi_x = \pi_y$ then,

$$\pi_x \cdot P_{x,y} \quad = \quad \pi_y \cdot P_{y,x},$$

MCMC Continued

Proof

We show that chain is time reversible and apply Theorem (for stationary distribution).
For any $x \neq y$, if $\pi_x = \pi_y$ then,

$$\begin{aligned} \pi_x \cdot P_{x,y} &= \pi_y \cdot P_{y,x}, \\ \text{since } P_{x,y} &= P_{y,x} = \frac{1}{M}. \end{aligned}$$

MCMC Continued

Proof

We show that chain is time reversible and apply Theorem (for stationary distribution).
For any $x \neq y$, if $\pi_x = \pi_y$ then,

$$\begin{aligned} \pi_x \cdot P_{x,y} &= \pi_y \cdot P_{y,x}, \\ \text{since } P_{x,y} &= P_{y,x} = \frac{1}{M}. \\ \implies \pi_x &= \frac{1}{|\Omega|} \text{ is the stationary distribution.} \end{aligned}$$

Example

Example

Consider the following simple Markov chain, whose states are independent sets in a graph $G = (V, E)$.

Example

Example

Consider the following simple Markov chain, whose states are independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

Example

Example

Consider the following simple Markov chain, whose states are independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

(2) To compute X_{i+1} :

Example

Example

Consider the following simple Markov chain, whose states are independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.
(2) To compute X_{i+1} :
   (a) choose a vertex v uniformly at random from V,

Example

Example

Consider the following simple Markov chain, whose states are independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

(2) To compute X_{i+1} :

  (a) choose a vertex v uniformly at random from V,

  (b) if $v \in X_i$ then $X_{i+1} = X_i \backslash \{v\}$,

Example

Example

Consider the following simple Markov chain, whose states are independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

(2) To compute X_{i+1} :

 (a) choose a vertex v uniformly at random from V,
 (b) if $v \in X_i$ then $X_{i+1} = X_i \backslash \{v\}$,
 (c) if $v \notin X_i$ and if adding v to X_i still gives an independent set, then $X_{i+1} = X_i \cup \{v\}$,

Example

Example

Consider the following simple Markov chain, whose states are independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

(2) To compute X_{i+1} :

  (a) choose a vertex v uniformly at random from V,
  (b) if $v \in X_i$ then $X_{i+1} = X_i \setminus \{v\}$,
  (c) if $v \notin X_i$ and if adding v to X_i still gives an independent set, then $X_{i+1} = X_i \cup \{v\}$,
  (d) otherwise, $X_{i+1} = X_i$.

Example

Example

Consider the following simple Markov chain, whose states are independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

(2) To compute X_{i+1} :

  (a) choose a vertex v uniformly at random from V,
  (b) if $v \in X_i$ then $X_{i+1} = X_i \backslash \{v\}$,
  (c) if $v \notin X_i$ and if adding v to X_i still gives an independent set, then $X_{i+1} = X_i \cup \{v\}$,
  (d) otherwise, $X_{i+1} = X_i$.

contd...

Chain is :

Example

Example

Consider the following simple Markov chain, whose states are independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

(2) To compute X_{i+1} :
   - (a) choose a vertex v uniformly at random from V,
   - (b) if $v \in X_i$ then $X_{i+1} = X_i \backslash \{v\}$,
   - (c) if $v \notin X_i$ and if adding v to X_i still gives an independent set, then $X_{i+1} = X_i \cup \{v\}$,
   - (d) otherwise, $X_{i+1} = X_i$.

contd...

Chain is :

(i) Irreducible

Example

Example

Consider the following simple Markov chain, whose states are independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

(2) To compute X_{i+1} :
   - (a) choose a vertex v uniformly at random from V,
   - (b) if $v \in X_i$ then $X_{i+1} = X_i \backslash \{v\}$,
   - (c) if $v \notin X_i$ and if adding v to X_i still gives an independent set, then $X_{i+1} = X_i \cup \{v\}$,
   - (d) otherwise, $X_{i+1} = X_i$.

contd...

Chain is :

(i) Irreducible

(ii) Aperiodic

Example

Example

Consider the following simple Markov chain, whose states are independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

(2) To compute X_{i+1} :

   (a) choose a vertex v uniformly at random from V,
   (b) if $v \in X_i$ then $X_{i+1} = X_i \backslash \{v\}$,
   (c) if $v \notin X_i$ and if adding v to X_i still gives an independent set, then $X_{i+1} = X_i \cup \{v\}$,
   (d) otherwise, $X_{i+1} = X_i$.

contd...

Chain is :

(i) Irreducible

(ii) Aperiodic

(iii) When $y \neq x$, it follows that $P_{x,y} = \frac{1}{|V|}$ or 0.

Example

Example

Consider the following simple Markov chain, whose states are independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

(2) To compute X_{i+1} :

- (a) choose a vertex v uniformly at random from V,
- (b) if $v \in X_i$ then $X_{i+1} = X_i \backslash \{v\}$,
- (c) if $v \notin X_i$ and if adding v to X_i still gives an independent set, then $X_{i+1} = X_i \cup \{v\}$,
- (d) otherwise, $X_{i+1} = X_i$.

contd...

Chain is :

(i) Irreducible

(ii) Aperiodic

(iii) When $y \neq x$, it follows that $P_{x,y} = \frac{1}{|V|}$ or 0.

$\Longrightarrow$ Stationary distribution is the uniform distribution

Outline

Metropolis Algorithm

Metropolis Algorithm

Metropolis Algorithm

Metropolis Algorithm

It refers to a general construction algorithm that transforms any irreducible Markov chain on a state space Ω to a time-reversible Markov chain with a required stationary distribution.

Metropolis Algorithm

Metropolis Algorithm

It refers to a general construction algorithm that transforms any irreducible Markov chain on a state space Ω to a time-reversible Markov chain with a required stationary distribution.

Notes

Assume we have designed an irreducible state space for a Markov chain.

Metropolis Algorithm

Metropolis Algorithm

It refers to a general construction algorithm that transforms any irreducible Markov chain on a state space Ω to a time-reversible Markov chain with a required stationary distribution.

Notes

Assume we have designed an irreducible state space for a Markov chain.
We want to construct Markov chain on this state space with a stationary distribution, $\pi_x = \frac{b(x)}{B}$, $\forall x \in \Omega$ we have $b(x) > 0$ such that $B = \sum_{x \in \Omega} b(x)$ is finite.

Metropolis Algorithm

Lemma

For a finite state space Ω and neighborhood structure $\{N(X)|x \in \Omega\}$, let $N = \max_{x\in\Omega} |N(x)|$.

Metropolis Algorithm

Lemma

For a finite state space Ω and neighborhood structure $\{N(X)|x \in \Omega\}$, let $N = \max_{x\in\Omega} |N(x)|$. Let M be any number such that $M \geq N$. For all $x \in \Omega$, let $\pi_x > 0$ be the desired probability of state x in the stationary distribution.

Metropolis Algorithm

Lemma

For a finite state space Ω and neighborhood structure $\{N(X)|x \in \Omega\}$, let $N = \max_{x \in \Omega} |N(x)|$. Let M be any number such that $M \geq N$. For all $x \in \Omega$, let $\pi_x > 0$ be the desired probability of state x in the stationary distribution.
Consider a Markov chain where

$$P_{x,y} = \begin{cases} \frac{1}{M}\min(1, \frac{\pi_y}{\pi_x}) & \text{if } x \neq y \text{ and } y \in N(x), \\ & \end{cases}$$

Metropolis Algorithm

Lemma

For a finite state space Ω and neighborhood structure $\{N(X) | x \in \Omega\}$, let $N = \max_{x \in \Omega} |N(x)|$. Let M be any number such that $M \geq N$. For all $x \in \Omega$, let $\pi_x > 0$ be the desired probability of state x in the stationary distribution.
Consider a Markov chain where

$$P_{x,y} = \begin{cases} \frac{1}{M} \min(1, \frac{\pi_y}{\pi_x}) & \text{if } x \neq y \text{ and } y \in N(x), \\ 0 & \text{if } x \neq y \text{ and } y \notin N(x), \end{cases}$$

Metropolis Algorithm

Lemma

For a finite state space Ω and neighborhood structure $\{N(X)|x \in \Omega\}$, let $N = \max_{x\in\Omega} |N(x)|$. Let M be any number such that $M \geq N$. For all $x \in \Omega$, let $\pi_x > 0$ be the desired probability of state x in the stationary distribution.
Consider a Markov chain where

$$P_{x,y} = \begin{cases} \frac{1}{M}\min(1, \frac{\pi_y}{\pi_x}) & \text{if } x \neq y \text{ and } y \in N(x), \\ 0 & \text{if } x \neq y \text{ and } y \notin N(x), \\ 1 - \sum_{y \neq x} P_{x,y} & \text{if } x = y. \end{cases}$$

Metropolis Algorithm

Lemma

For a finite state space Ω and neighborhood structure $\{N(X)|x \in \Omega\}$, let $N = \max_{x\in\Omega} |N(x)|$. Let M be any number such that $M \geq N$. For all $x \in \Omega$, let $\pi_x > 0$ be the desired probability of state x in the stationary distribution.
Consider a Markov chain where

$$P_{x,y} = \begin{cases} \frac{1}{M}\min(1, \frac{\pi_y}{\pi_x}) & \text{if } x \neq y \text{ and } y \in N(x), \\ 0 & \text{if } x \neq y \text{ and } y \notin N(x), \\ 1 - \sum_{y \neq x} P_{x,y} & \text{if } x = y. \end{cases}$$

Then, if this chain is irreducible and aperiodic, the stationary distribution is given by the probabilities π_x.

Lemma

Proof

We show that chain is time reversible and apply Theorem (for stationary distribution).

Lemma

Proof

We show that chain is time reversible and apply Theorem (for stationary distribution).
For any $x \neq y$ if $\pi_x \leq \pi_y$ then $P_{x,y} = 1$ and $P_{y,x} = \frac{\pi_x}{\pi_y}$.

Lemma

Proof

We show that chain is time reversible and apply Theorem (for stationary distribution).
For any $x \neq y$ if $\pi_x \leq \pi_y$ then $P_{x,y} = 1$ and $P_{y,x} = \frac{\pi_x}{\pi_y}$.
$\implies \pi_x \cdot P_{x,y} = \pi_y \cdot P_{y,x}$.

Lemma

Proof

We show that chain is time reversible and apply Theorem (for stationary distribution).
For any $x \neq y$ if $\pi_x \leq \pi_y$ then $P_{x,y} = 1$ and $P_{y,x} = \frac{\pi_x}{\pi_y}$.
$\implies \pi_x \cdot P_{x,y} = \pi_y \cdot P_{y,x}$.
Similarly, if $\pi_x > \pi_y$, then $P_{x,y} = \frac{\pi_y}{\pi_x}$ and $P_{y,x} = 1$.

Lemma

Proof

We show that chain is time reversible and apply Theorem (for stationary distribution).
For any $x \neq y$ if $\pi_x \leq \pi_y$ then $P_{x,y} = 1$ and $P_{y,x} = \frac{\pi_x}{\pi_y}$.
$\implies \pi_x \cdot P_{x,y} = \pi_y \cdot P_{y,x}$.
Similarly, if $\pi_x > \pi_y$, then $P_{x,y} = \frac{\pi_y}{\pi_x}$ and $P_{y,x} = 1$.
$\implies \pi_x \cdot P_{x,y} = \pi_y \cdot P_{y,x}$.

Lemma

Proof

We show that chain is time reversible and apply Theorem (for stationary distribution).
For any $x \neq y$ if $\pi_x \leq \pi_y$ then $P_{x,y} = 1$ and $P_{y,x} = \frac{\pi_x}{\pi_y}$.
$\implies \pi_x \cdot P_{x,y} = \pi_y \cdot P_{y,x}$.
Similarly, if $\pi_x > \pi_y$, then $P_{x,y} = \frac{\pi_y}{\pi_x}$ and $P_{y,x} = 1$.
$\implies \pi_x \cdot P_{x,y} = \pi_y \cdot P_{y,x}$.
By Theorem (for stationary distribution), the stationary distribution is given by π_x.

Lemma

Proof

We show that chain is time reversible and apply Theorem (for stationary distribution).
For any $x \neq y$ if $\pi_x \leq \pi_y$ then $P_{x,y} = 1$ and $P_{y,x} = \frac{\pi_x}{\pi_y}$.
$\implies \pi_x \cdot P_{x,y} = \pi_y \cdot P_{y,x}$.
Similarly, if $\pi_x > \pi_y$, then $P_{x,y} = \frac{\pi_y}{\pi_x}$ and $P_{y,x} = 1$.
$\implies \pi_x \cdot P_{x,y} = \pi_y \cdot P_{y,x}$.
By Theorem (for stationary distribution), the stationary distribution is given by π_x.

Independent sets

Suppose we want to create Markov chain where, in the stationary distribution, each independent set I has probability proportional to $\lambda^{|I|}$ for some constant parameter $\lambda > 0$.

Lemma

Proof

We show that chain is time reversible and apply Theorem (for stationary distribution).
For any $x \neq y$ if $\pi_x \leq \pi_y$ then $P_{x,y} = 1$ and $P_{y,x} = \frac{\pi_x}{\pi_y}$.
$\implies \pi_x \cdot P_{x,y} = \pi_y \cdot P_{y,x}$.
Similarly, if $\pi_x > \pi_y$, then $P_{x,y} = \frac{\pi_y}{\pi_x}$ and $P_{y,x} = 1$.
$\implies \pi_x \cdot P_{x,y} = \pi_y \cdot P_{y,x}$.
By Theorem (for stationary distribution), the stationary distribution is given by π_x.

Independent sets

Suppose we want to create Markov chain where, in the stationary distribution, each independent set I has probability proportional to $\lambda^{|I|}$ for some constant parameter $\lambda > 0$.
i.e. $\pi_x = \frac{\lambda^{|I_x|}}{B}$,

Lemma

Proof

We show that chain is time reversible and apply Theorem (for stationary distribution).
For any $x \neq y$ if $\pi_x \leq \pi_y$ then $P_{x,y} = 1$ and $P_{y,x} = \frac{\pi_x}{\pi_y}$.
$\implies \pi_x \cdot P_{x,y} = \pi_y \cdot P_{y,x}$.
Similarly, if $\pi_x > \pi_y$, then $P_{x,y} = \frac{\pi_y}{\pi_x}$ and $P_{y,x} = 1$.
$\implies \pi_x \cdot P_{x,y} = \pi_y \cdot P_{y,x}$.
By Theorem (for stationary distribution), the stationary distribution is given by π_x.

Independent sets

Suppose we want to create Markov chain where, in the stationary distribution, each independent set I has probability proportional to $\lambda^{|I|}$ for some constant parameter $\lambda > 0$.
i.e. $\pi_x = \frac{\lambda^{|I_x|}}{B}$,
where I_x is independent set corresponding to state x and $B = \sum_x \lambda^{|I_x|}$.

Lemma

Proof

We show that chain is time reversible and apply Theorem (for stationary distribution).
For any $x \neq y$ if $\pi_x \leq \pi_y$ then $P_{x,y} = 1$ and $P_{y,x} = \frac{\pi_x}{\pi_y}$.
$\implies \pi_x \cdot P_{x,y} = \pi_y \cdot P_{y,x}$.
Similarly, if $\pi_x > \pi_y$, then $P_{x,y} = \frac{\pi_y}{\pi_x}$ and $P_{y,x} = 1$.
$\implies \pi_x \cdot P_{x,y} = \pi_y \cdot P_{y,x}$.
By Theorem (for stationary distribution), the stationary distribution is given by π_x.

Independent sets

Suppose we want to create Markov chain where, in the stationary distribution, each independent set I has probability proportional to $\lambda^{|I|}$ for some constant parameter $\lambda > 0$.
i.e. $\pi_x = \frac{\lambda^{|I_x|}}{B}$,
where I_x is independent set corresponding to state x and $B = \sum_x \lambda^{|I_x|}$.
When $\lambda = 1$, this is the uniform distribution,

Lemma

Proof

We show that chain is time reversible and apply Theorem (for stationary distribution).
For any $x \neq y$ if $\pi_x \leq \pi_y$ then $P_{x,y} = 1$ and $P_{y,x} = \frac{\pi_x}{\pi_y}$.
$\implies \pi_x \cdot P_{x,y} = \pi_y \cdot P_{y,x}$.
Similarly, if $\pi_x > \pi_y$, then $P_{x,y} = \frac{\pi_y}{\pi_x}$ and $P_{y,x} = 1$.
$\implies \pi_x \cdot P_{x,y} = \pi_y \cdot P_{y,x}$.
By Theorem (for stationary distribution), the stationary distribution is given by π_x.

Independent sets

Suppose we want to create Markov chain where, in the stationary distribution, each independent set I has probability proportional to $\lambda^{|I|}$ for some constant parameter $\lambda > 0$.
i.e. $\pi_x = \frac{\lambda^{|I_x|}}{B}$,
where I_x is independent set corresponding to state x and $B = \sum_x \lambda^{|I_x|}$.
When $\lambda = 1$, this is the uniform distribution,
when $\lambda > 1$, larger independent sets have a larger probability than smaller independent sets,

Lemma

Proof

We show that chain is time reversible and apply Theorem (for stationary distribution).
For any $x \neq y$ if $\pi_x \leq \pi_y$ then $P_{x,y} = 1$ and $P_{y,x} = \frac{\pi_x}{\pi_y}$.
$\implies \pi_x \cdot P_{x,y} = \pi_y \cdot P_{y,x}$.
Similarly, if $\pi_x > \pi_y$, then $P_{x,y} = \frac{\pi_y}{\pi_x}$ and $P_{y,x} = 1$.
$\implies \pi_x \cdot P_{x,y} = \pi_y \cdot P_{y,x}$.
By Theorem (for stationary distribution), the stationary distribution is given by π_x.

Independent sets

Suppose we want to create Markov chain where, in the stationary distribution, each independent set I has probability proportional to $\lambda^{|I|}$ for some constant parameter $\lambda > 0$.
i.e. $\pi_x = \frac{\lambda^{|I_x|}}{B}$,
where I_x is independent set corresponding to state x and $B = \sum_x \lambda^{|I_x|}$.
When $\lambda = 1$, this is the uniform distribution,
when $\lambda > 1$, larger independent sets have a larger probability than smaller independent sets,
when $\lambda < 1$, larger independent sets have smaller probability than smaller independent sets.

Example

Example

Consider the following variation on previous Markov chain for independent sets in a graph $G = (V, E)$.

Example

Example

Consider the following variation on previous Markov chain for independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

Example

Example

Consider the following variation on previous Markov chain for independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

(2) To compute X_{i+1} :

Example

Example

Consider the following variation on previous Markov chain for independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

(2) To compute X_{i+1} :

  (a) choose a vertex v uniformly at random from V,

Example

Example

Consider the following variation on previous Markov chain for independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

(2) To compute X_{i+1} :
   (a) choose a vertex v uniformly at random from V,
   (b) if $v \in X_i$, set $X_{i+1} = X_i \backslash \{v\}$ with probability $\min(1, \frac{1}{\lambda})$,

Example

Example

Consider the following variation on previous Markov chain for independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

(2) To compute X_{i+1} :

  (a) choose a vertex v uniformly at random from V,
  (b) if $v \in X_i$, set $X_{i+1} = X_i \backslash \{v\}$ with probability $\min(1, \frac{1}{\lambda})$,
  (c) if $v \notin X_i$ and if adding v to X_i still gives an independent set, then put $X_{i+1} = X_i \cup \{v\}$ with probability $\min(1, \lambda)$,

Example

Example

Consider the following variation on previous Markov chain for independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

(2) To compute X_{i+1} :

  (a) choose a vertex v uniformly at random from V,
  (b) if $v \in X_i$, set $X_{i+1} = X_i \backslash \{v\}$ with probability $\min(1, \frac{1}{\lambda})$,
  (c) if $v \notin X_i$ and if adding v to X_i still gives an independent set, then put $X_{i+1} = X_i \cup \{v\}$ with probability $\min(1, \lambda)$,
  (d) otherwise, set $X_{i+1} = X_i$.

Example

Example

Consider the following variation on previous Markov chain for independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

(2) To compute X_{i+1} :

  (a) choose a vertex v uniformly at random from V,
  (b) if $v \in X_i$, set $X_{i+1} = X_i \backslash \{v\}$ with probability $\min(1, \frac{1}{\lambda})$,
  (c) if $v \notin X_i$ and if adding v to X_i still gives an independent set, then put $X_{i+1} = X_i \cup \{v\}$ with probability $\min(1, \lambda)$,
  (d) otherwise, set $X_{i+1} = X_i$.

Two-step approach

Choose a vertex v to add or delete with probability $\frac{1}{M}$; here $M = |V|$.

Example

Example

Consider the following variation on previous Markov chain for independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

(2) To compute X_{i+1} :

  (a) choose a vertex v uniformly at random from V,
  (b) if $v \in X_i$, set $X_{i+1} = X_i \backslash \{v\}$ with probability $\min(1, \frac{1}{\lambda})$,
  (c) if $v \notin X_i$ and if adding v to X_i still gives an independent set, then put $X_{i+1} = X_i \cup \{v\}$ with probability $\min(1, \lambda)$,
  (d) otherwise, set $X_{i+1} = X_i$.

Two-step approach

Choose a vertex v to add or delete with probability $\frac{1}{M}$; here $M = |V|$.
This proposal is accepted with probability $min(1, \frac{\pi_y}{\pi_x})$, where x is current state and y is proposed state.

Example

Example

Consider the following variation on previous Markov chain for independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

(2) To compute X_{i+1} :

- (a) choose a vertex v uniformly at random from V,
- (b) if $v \in X_i$, set $X_{i+1} = X_i \backslash \{v\}$ with probability $\min(1, \frac{1}{\lambda})$,
- (c) if $v \notin X_i$ and if adding v to X_i still gives an independent set, then put $X_{i+1} = X_i \cup \{v\}$ with probability $\min(1, \lambda)$,
- (d) otherwise, set $X_{i+1} = X_i$.

Two-step approach

Choose a vertex v to add or delete with probability $\frac{1}{M}$; here $M = |V|$.
This proposal is accepted with probability $min(1, \frac{\pi_y}{\pi_x})$, where x is current state and y is proposed state.

(i) If chain adds a vertex, $\frac{\pi_y}{\pi_x}$ is λ

Example

Example

Consider the following variation on previous Markov chain for independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

(2) To compute X_{i+1} :

  (a) choose a vertex v uniformly at random from V,
  (b) if $v \in X_i$, set $X_{i+1} = X_i \backslash \{v\}$ with probability $\min(1, \frac{1}{\lambda})$,
  (c) if $v \notin X_i$ and if adding v to X_i still gives an independent set, then put $X_{i+1} = X_i \cup \{v\}$ with probability $\min(1, \lambda)$,
  (d) otherwise, set $X_{i+1} = X_i$.

Two-step approach

Choose a vertex v to add or delete with probability $\frac{1}{M}$; here $M = |V|$.
This proposal is accepted with probability $min(1, \frac{\pi_y}{\pi_x})$, where x is current state and y is proposed state.

(i) If chain adds a vertex, $\frac{\pi_y}{\pi_x}$ is λ

(ii) If chain deletes a vertex $\frac{\pi_y}{\pi_x}$ is $\frac{1}{\lambda}$

Example

Example

Consider the following variation on previous Markov chain for independent sets in a graph $G = (V, E)$.

(1) X_0 is an arbitrary independent set in G.

(2) To compute X_{i+1} :

(a) choose a vertex v uniformly at random from V,

(b) if $v \in X_i$, set $X_{i+1} = X_i \backslash \{v\}$ with probability $\min(1, \frac{1}{\lambda})$,

(c) if $v \notin X_i$ and if adding v to X_i still gives an independent set, then put $X_{i+1} = X_i \cup \{v\}$ with probability $\min(1, \lambda)$,

(d) otherwise, set $X_{i+1} = X_i$.

Two-step approach

Choose a vertex v to add or delete with probability $\frac{1}{M}$; here $M = |V|$.
This proposal is accepted with probability $min(1, \frac{\pi_y}{\pi_x})$, where x is current state and y is proposed state.

(i) If chain adds a vertex, $\frac{\pi_y}{\pi_x}$ is λ

(ii) If chain deletes a vertex $\frac{\pi_y}{\pi_x}$ is $\frac{1}{\lambda}$

The transition probability $P_{x,y} = \frac{1}{M} \cdot min(1, \frac{\pi_y}{\pi_x})$, so Lemma applies.

Metropolis Algorithm

Notes

We dint need to know $B = \sum_x \lambda^{|I_x|}$.
Graphs with n vertices can have exponentially many independent sets calculating whose sum would be expensive task.
Markov chain here gives the stationary distribution by using only the ratios $\frac{\pi_y}{\pi_x}$.