Theory of Stochastic Processes
8. Markov chain Monte Carlo

Tomonari Sei
sei@mist.i.u-tokyo.ac.jp

Department of Mathematical Informatics, University of Tokyo

June 8, 2017

http://www.stat.t.u-tokyo.ac.jp/~sei/lec.html

Handouts & Announcements

Handouts:

- Slides (this one)
- Solutions of the midterm exam
- Copy of §6.5 and §6.14 of PRP

Note:

- Your answer sheets of the midterm exam will be returned today unless you took an additional exam (for conference attendee or sick leave).

Schedule

Apr 6	Overview
Apr 13	Simple random walk
Apr 20	Generating functions
Apr 27	Markov chain
May 11	Continuous-time Markov chain
May 18	Review
May 25	(midterm exam)
June 8	Markov chain Monte Carlo
June 15	Stationary processes
June 22	Martingales
June 29	Queues
July 6	Diffusion processes
July 13	Review
July 20	(final exam) $\leftarrow$ the date has been decided.

Outline today

1. Remarks about the midterm exam
2. Markov chain Monte Carlo
   - Detailed balance equation
   - Gibbs sampler
   - Metropolis-Hastings algorithm
3. Recommended problems

How to score

- I sincerely apologize that the problems Q1-(c) and Q4-(b) were very difficult.
- So I decided that the full marks is 70 marks.
- Therefore the final score will be based on

$$40\% \times \frac{10}{7} \times (\text{midterm exam}) + 60\% \times (\text{final exam})$$

- The score of Q5 of the midterm exam is either α, β, or 0, where α and β are functions of the total score of Q1 to Q4, and satisfy $0 \leq \beta \leq \alpha \leq 10$. I have not decided the specific function yet...
- The final exam will consist of easier questions...

Q5 of the midterm exam

Research interest

- statistics, numerical algorithm, optimization, intelligent systems, music data, computational linguistics, causal discovery, mathematical finance, computer graphics, random graph, economics, data structure, hybrid system, brain consciousness.

Comments on the lecture (selected)

- loud
- explanation could be more detailed
- FAQ
- easier exercises
- more proofs

Review: Stationary distribution

Let p_{ij} denote the transition probability (from i to j) of a Markov chain.

Definition (reminder)

A mass function $\boldsymbol{\pi} = (\pi_i)_{i \in S}$ is called the stationary distribution of the Markov chain if

$$\sum_{i \in S} \pi_i p_{ij} = \pi_j \quad \text{for all} \quad j \in S.$$

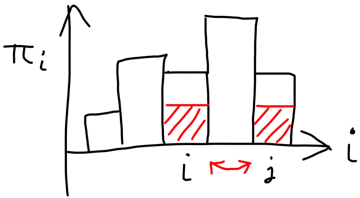
Detailed balance equation

Definition

A Markov chain is called reversible if it has the stationary distribution $\boldsymbol{\pi} = (\pi_i)_{i \in S}$ satisfying

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \text{for all} \quad i, j \in S.$$

This equation is called the detailed balance equation.



$\pi_i p_{ij} = \pi_j p_{ji}$

Detailed balance implies stationarity

Lemma

If a mass function π_i satisfies the detailed balance equation, then π_i is a stationary distribution.

Proof:

$$\begin{aligned}\sum_i \pi_i p_{ij} &= \sum_i \pi_j p_{ji} \quad \text{(detailed balance)} \\ &= \pi_j \sum_i p_{ji} \\ &= \pi_j \quad \text{(property of transition matrix)}\end{aligned}$$

$\square$

Exercises

Exercise 1

Let $\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 1/6 & 1/2 & 1/3 \\ 0 & 2/3 & 1/3 \end{pmatrix}$ and $\boldsymbol{\pi} = (1/10, 6/10, 3/10)$.
Show that the detailed balance equation is satisfied.

Exercise 2

Let $\mathbf{P} = \begin{pmatrix} a & b & c \\ c & a & b \\ b & c & a \end{pmatrix}$ and $\boldsymbol{\pi} = (1/3, 1/3, 1/3)$, where $a, b, c \geq 0$ and $a + b + c = 1$.

1. Show that $\boldsymbol{\pi}$ is a stationary distribution.
2. Determine the condition that the detailed balance equation holds.

Note: a matrix $\mathbf{P}$ is called doubly stochastic if all the row sums and column sums are 1.

Markov Chain Monte Carlo (MCMC)

We begin with an example.

A toy example

Let Θ be the set of all strings of length 20 that consist of four letters 'A', 'G', 'C', 'T'. For example,

$$\text{“AGATACACATTTAACGGCAT”} \in \Theta.$$

Define a probability mass function on Θ by

$$\pi(\theta) = \frac{2^{c(\theta)}}{Z}, \quad Z = \sum_{\theta \in \Theta} 2^{c(\theta)},$$

where $c(\theta)$ is the number of “CAT” in θ. Find the expected number

$$\sum_{\theta \in \Theta} c(\theta)\pi(\theta).$$

MCMC

- Our target is $\sum_{\theta \in \Theta} c(\theta)\pi(\theta)$.
- Since $|\Theta| = 4^{20} \simeq 10^{12}$, computation of the sum is not so easy.
- Instead, construct a Markov chain in the following manner.

Gibbs sampler for the above problem

1. Determine an initial state $\theta^{(0)}$. For example, $\theta^{(0)} =$ "AAAAAAAAAAAAAAAAAAAA".
2. Choose a site from the 20 at random, and replace it with a letter in such a way that the detailed balance equation is satisfied. Repeat this procedure N times. Let $\theta^{(n)}$ for $n = 1, \ldots, N$ be the obtained sequence.
3. An estimate is $(1/N) \sum_{n=1}^{N} c(\theta^{(n)})$.

MCMC

For example, if

$$\theta = \text{“AGATACA}\underline{\text{CAT}}\text{TTAACGG}\underline{\text{CAT}}\text{”}$$

and the 2nd letter is chosen (with probability $1/20$), then move from θ to ϕ in the following probability:

ϕ	$2^{c(\phi)}$	probability
“AAATACA<u>CAT</u>TTAACGG<u>CAT</u>”	4	$1/5$
“AGATACA<u>CAT</u>TTAACGG<u>CAT</u>”	4	$1/5$
“A<u>CAT</u>ACA<u>CAT</u>TTAACGG<u>CAT</u>”	8	$2/5$
“ATATACA<u>CAT</u>TTAACGG<u>CAT</u>”	4	$1/5$

The detailed balance equation is satisfied. It will be checked later in more general form.

A result

Sorry, I had no time to implement it.

Ising model

Here is another example.

Example (Ising model)

Let $\Theta = \{-1, 1\}^{n \times n}$. The Ising model is defined by

$$\pi(\theta) = \frac{\pi^*(\theta)}{Z}, \quad \pi^*(\theta) = \exp\left(\beta \sum_{(v,w):v\sim w} \theta_v \theta_w\right), \quad \theta \in \Theta.$$

Here $v \sim w$ means that v and w are "adjacent" in the $n \times n$ lattice, Z is defined by $Z = \sum_{\theta \in \Theta} \pi^*(\theta)$, and $\beta \in \mathbb{R}$ is a given parameter.

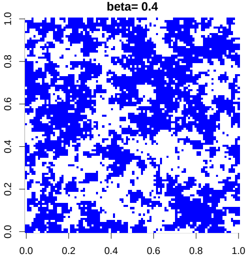
How to generate a random sample from $\pi(\theta)$?

It is hopeless to calculate Z since $|\Theta| = 2^{n^2}$.

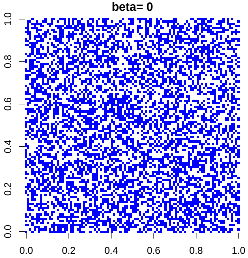
$\rightarrow$ MCMC avoids such computation.

Output of MCMC

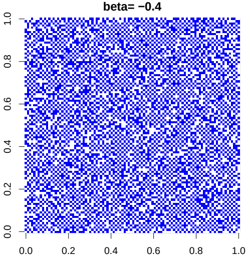
Random samples from the Ising model are shown.
The size of the lattice is 100×100.



$\beta = 0.4$



$\beta = 0$
(uniform distribution)



$\beta = -0.4$

Idea of MCMC

- Purpose: generate a sample from $\pi_i = \pi_i^*/Z$, where $Z = \sum_{i \in S} \pi_i^*$.
- Assumption: π_i^* is easily calculated, but Z is not.
- Solution: construct a Markov chain with the stationary distribution π.

Example (cont.)

For the Ising model, we used the following Markov chain:

- Choose $v \in V = \{1, \ldots, n\}^2$ randomly.
- Choose $\tilde{\theta}_v \in \{-1, 1\}$ with the probability

$$P(\tilde{\theta}_v = \pm 1) = \frac{e^{\pm \beta \sum_{w \sim v} \theta_w}}{e^{\beta \sum_{w \sim v} \theta_w} + e^{-\beta \sum_{w \sim v} \theta_w}}.$$

  and $\tilde{\theta}_w = \theta_w$ for all $w \neq v$.
- Move from θ to $\tilde{\theta}$.

This is an example of the Gibbs sampler. $\rightarrow$ next slide

Gibbs sampler

- The state space is $\Theta = S^V$, where S is "local state space" and V is a finite set.
- Let $\Theta_{i,v} = \{j \in \Theta \mid j_w = i_w \text{ for } w \neq v\}$ for $i \in \Theta$ and $v \in V$.
- Assumption: the quantity

$$h_{ij} = \frac{\pi_j}{\sum_{k \in \Theta_{i,v}} \pi_k}, \quad j \in \Theta_{i,v},$$

  is easy to compute.
- Markov chain: choose $v \in V$ randomly, and choose $j \in \Theta_{i,v}$ randomly according to $\{h_{ij}\}$. Then move from i to j.

Example (cont.)

For the Ising model, $S = \{-1, 1\}$, $V = \{1, \ldots, n\}^2$, $\pi_i \propto e^{\beta \sum_{(v,w):v \sim w} i_v i_w}$, and

$$h_{ij} = \frac{e^{\beta \sum_{w \sim v} j_v i_w}}{e^{\beta \sum_{w \sim v} i_w} + e^{-\beta \sum_{w \sim v} i_w}}, \quad j \in \Theta_{i,v}.$$

Validity of the Gibbs sampler

Lemma

The Markov chain constructed above is reversible.

Proof: The transition probability is

$$\begin{aligned} p_{ij} &= \frac{1}{|V|} \sum_{v \in V} 1_{\{j \in \Theta_{i,v}\}} h_{ij} \\ &= \frac{1}{|V|} \sum_{v \in V} 1_{\{j \in \Theta_{i,v}\}} \frac{\pi_j}{\sum_{k \in \Theta_{i,v}} \pi_k} \end{aligned}$$

Then

$$\pi_i p_{ij} = \frac{1}{|V|} \sum_{v \in V} 1_{\{j \in \Theta_{i,v}\}} \frac{\pi_i \pi_j}{\sum_{k \in \Theta_{i,v}} \pi_k}.$$

This is symmetric because $j \in \Theta_{i,v} \Leftrightarrow i \in \Theta_{j,v}$. $\square$

Exercise

Exercise

Let $\Theta = \{1,2\}^2 = \{(1,1),(1,2),(2,1),(2,2)\}$ and

$$\pi_i = \frac{1}{(i_1 + i_2)Z}, \quad i \in \Theta,$$

where $Z = \sum_{i \in \Theta} \frac{1}{i_1 + i_2}$.

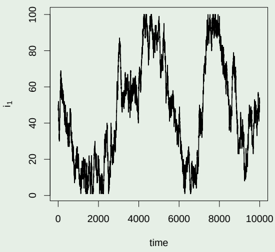
1. Write down the transition matrix of the Gibbs sampler.
2. Confirm that the obtained chain is reversible.

A drawback

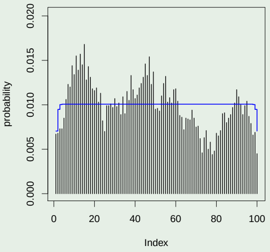
- In general, the Gibbs sampler is quite powerful.
- However, if the "correlation" of the target distribution is high, the convergence to the stationary distribution may be slow.

Example

Let $\Theta = \{1, \ldots, 100\}^2$ and $\pi_i \propto \exp(-|i_1 - i_2|^2/2)$ ($\leftarrow$ much correlated).



A sample path of i_1



Histogram of the states
(blue curve shows exact values)

Metropolis-Hastings algorithm

Given $X_n = i$, generate X_{n+1} as follows.

- Pick $Y \in \Theta$ according to

$$P(Y = j \mid X_n = i) = h_{ij},$$

  where $\mathbf{H} = (h_{ij})$ is an irreducible transition matrix called the proposal.
- Given that $Y = j$, set

$$X_{n+1} = \begin{cases} j & \text{with probability } a_{ij}, \\ X_n & \text{with probability } 1 - a_{ij}, \end{cases}$$

  where the acceptance probability $a_{ij} \in [0, 1]$ is determined to satisfy the detailed balance equation

$$\pi_i h_{ij} a_{ij} = \pi_j h_{ji} a_{ji}.$$

  For example,

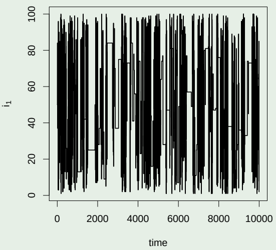
$$a_{ij} = \min\left(1, \frac{\pi_j h_{ji}}{\pi_i h_{ij}}\right), \quad a_{ij} = \frac{\pi_j h_{ji}}{\pi_i h_{ij} + \pi_j h_{ji}} \quad \text{etc.}$$

Example (cont.)

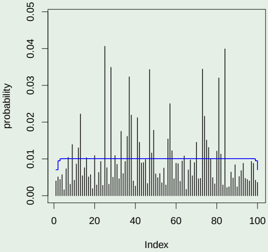
Let $\Theta = \{1, \ldots, 100\}^2$ and $\pi_i \propto \exp(-|i_1 - i_2|^2/2)$. For example, determine a proposal matrix as follows:

$$h_{ij} = \frac{1}{100}\left(\frac{1}{2} \cdot 1_{\{j_1 = j_2\}} + \frac{1}{198} \cdot 1_{\{j_1 \neq j_2\}}\right), \quad i, j \in \Theta.$$

Note that h_{ij} does not depend on i. We use $a_{ij} = \min(1, \pi_j h_{ji}/(\pi_i h_{ij}))$.

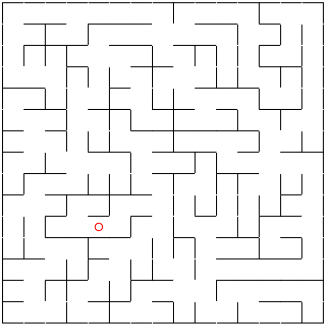


Trajectory of i_1

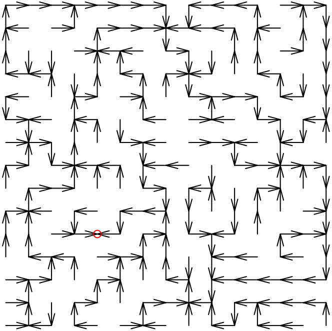


Histogram of i_1

Application: random sampling of mazes



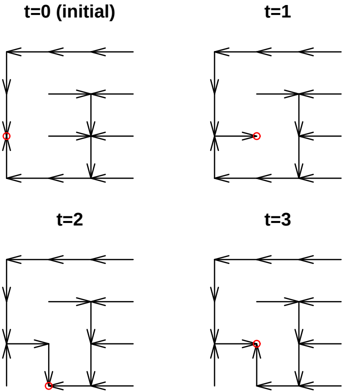
Maze



Rooted tree
(The root is indicated by ◯)

Application: random sampling of mazes

- Proposal: Select a neighbour point of the root with probability $1/4$. Then modify the tree.
- For example,



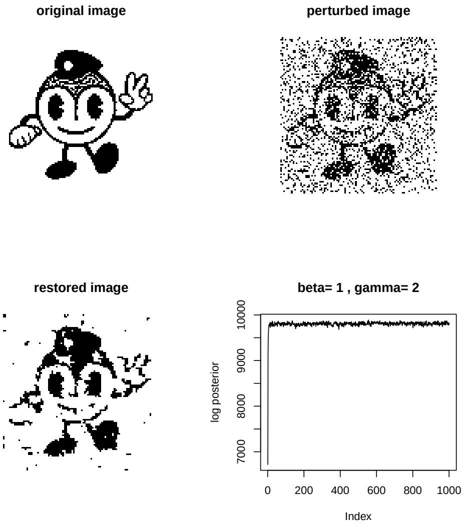
This chain is not reversible, but has the uniform stationary distribution.

Application: Bayesian inference

- θ is a parameter (unobservable)
- x is a data (observable)
- $\pi(\theta)$: prior
- $f(x|\theta)$: likelihood
- After the data x is observed, the posterior is given by

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\sum_\theta f(x|\theta)\pi(\theta)}.$$

Image restoration



http://www.city.yokohama.lg.jp/koutuu/kids/profile.html

Recommended problems

Recommended problems:

- §6.5, Problem 1, 8*, 9.
- §6.14, Problem 1.
- §6.15, Problem 2,

The asterisk (*) shows difficulty.