Markov Chain to MCMC - Take Gibbs Sampling as an example

Min-Yi Chen

October 17, 2022

Abstract

This is the note about how Markov chain works in MCMC, Gibbs sampling. Most of the contents are taken from [1] and [2] and under Dr. Ting-Li Chen's instruction.

1 Introdiction

MCMC is a popular framework that provides a general approach for generating samples from the posterior distribution. However, have you ever been confused about why MCMC works? In this note, I write down what I think about this question.

2 Invariant Equilibrium and Stationary

When we talk about Markov chain, the words invariant equilibrium and stationary appear frequently. Some people may think those terms are used to mean the same, however, they are slightly different.

Theorem 2.1. *Invariant [2]*
We say λ is invariant if

$$\lambda P = \lambda$$

Theorem 2.2. *Stationary [2]*
Let $(X_n)_{n\geq 0}$ be $Markov(\lambda, P)$ and suppose that λ is invariant for P. Then $(X_{m+n})_{n\geq 0}$ is also $Markov(\lambda, P)$

Invariant and Stationary are similar. It just a terms used by different people.In [1], the author mentioned that

Definition 2.3. *A distribution $\pi(X)$ is a stationary distribution for Markov Chain $\mathcal{T}$ if it satisfies:*

$$\pi(X=j) = \textstyle\sum_{i\in I} \pi(X=i)\mathcal{T}(i \to j), \text{ where } \mathcal{T} \text{ is the transition probability[1]}$$

A stationary distribution is also called an invariant distribution

However, most of the time when we use stationary to enhance that after n state, $n+m$ state doesn't change. (Theorem2.2)

Theorem 2.4. *Equilibrium [2]*
Let I be finite. Suppose for some $i \in I$ that

$$\mathcal{T}^{(n)}(i \to j) \to \pi_j \text{ as } n \to \infty \ \forall j \in I$$

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

Equilibrium is to describe the state $X_t \approx X_{t+1}$, that is, equilibrium is used at convergence. The equation below can show equilibrium.

$$\mathcal{T}(i \to j) = \lim_{n\to\infty} \mathcal{T}^{(n)}(i \to j) = \pi_j$$

3 Gibbs sampling

Recall Naive Bayes, we have

$$P(\theta|\mathcal{X}) = \frac{P(\mathcal{X}|\theta)P(\theta)}{P(\mathcal{X})}$$

Our goal in Gibbs sampling is to approach $P(\theta|\mathcal{X}) \approx \pi(\theta)$ by iterative process. Let $\theta = (\theta^{(1)}, \ldots, \theta^{(d)})$, $d \in \mathbb{N}$. Gibbs sampling update θ one dimension per time during iteration. That is
before update: $\theta = (\theta^{(1)}, \ldots, \theta^{(d)})$, after update: $\theta' = (\theta^{(1)}, \ldots, \theta'^{(i)}, \ldots, \theta^{(d)})$
We define the process as $\mathcal{T}_i(\theta \to \theta')$ and the probability to θ' is

$$\mathcal{T}_i(\theta \to \theta') = \frac{\pi(\theta^{(1)}, \ldots, \theta^{(i-1)}, \theta'^{(i)}, \ldots, \theta^{(d)})}{\sum_{\mathcal{Z}} \pi(\theta^{(1)}, \ldots, \theta^{(i-1)}, \mathcal{Z}, \ldots, \theta^{(d)})}$$

A practical example is exhibited below.

Example 3.1. *Let $\theta = (\theta^{(1)}, \theta^{(2)}, \theta^{(3)})$, $\theta^{(i)} \in \{1, 2, 3, 4, 5\}$,$\theta = (2, 3, 5)$*

$$\mathcal{T}_2(\theta \to \theta') = \frac{\pi(2, \theta'^{(2)}, 5)}{\pi(2,1,5) + \ldots + \pi(2,5,5)}$$

Now, we know how Gibbs sampling work. However, there're two questions that emerges.

1. The kernel $\mathcal{T}_i$ and $\mathcal{T}_{i+1}$ is different. This means we have to consider if this multiple kernel chain is stationary.
2. $\mathcal{T}_i(\theta^{(i)} \to \theta'^{(i)})$ does not depend on $x_i \in X_i$, is $P(\theta|\mathcal{X})$ a stationary distribution ?

To show the posterior distribution is stationary with Gibbs sampling, we discuss the transition kernel first.

Let $(\theta^{(-i)}, \theta^{(i)})$ be i state and $(\theta^{(-i)}, \theta'^{(i)})$ be i' state after Gibbs sampling.

$$\mathcal{T}_i((\theta^{(-i)}, \theta^{(i)}) \to (\theta^{(-i)}, \theta'^{(i)})) = P(\theta'^{(i)}|\theta^{(-i)}) = \pi_x(\theta)$$

We can describe the transition kernel as a function

$$\mathcal{T}_i : (a,b) \to (x,d), \text{ if } a \neq c,\ P(\mathcal{T}_i(a,b) = (c,d)) = 0$$

Thus, set $x = (x_{-i}, x_i)$ and $y = (y_{-i}, y_i)$ if $x_i \neq y_i$,$P(\mathcal{T}_i(x_{-i}, x_i) = (y_{-i}, y_i)) = 0$

With the statement above, we write

$$\pi(x) = \sum_y \mathcal{T}_i(y \to x)\pi(y)$$

where $\mathcal{T}_i(y \to x) = \mathcal{T}_i((x_{-i}, y_i) \to (x_{-i}, x_i)) = \frac{\pi((x_{-i},x_i))}{\sum_Z \pi((x_{-i},Z))}$

To write it clearly, we have

$\pi(x) = \sum_{(y_{-i}=x_{-i},y_i)} \mathcal{T}_i(y \to x)\pi(y) = \sum_{(y_{-i}=x_{-i},y_i)} \frac{\pi(x_{-i},x_i)}{\sum_Z \pi(x_{-i},Z)} \pi(x_{-i}, y_i)$

$= \frac{\pi(x_{-i},x_i)}{\sum_Z \pi(x_{-i},Z)} \sum_{(y_{-i}=x_{-i},y_i)} \pi(x_{-i}, y_i)$

Since $\sum_{(y_{-i}=x_{-i},y_i)} \pi(x_{-i}, y_i) = \sum_Z \pi(x_{-i}, Z)$,we have

$\pi(x) = \pi(x_{-i}, x_i)$,thus, we show that the posterior distribution is stationary.

4 Unique stationary distribution

We have shown Gibbs sampling is stationary, however, do we really sure the posterior is exactly the right distribution? To support it, we have to prove MCMC method based on Markov chain have a unique stationary distribution.

According to [1] and [2], I find out some proofs and arrange it by myself.

In [1], the author explains how Markov Chain has a unique stationary distribution with a theorem related to regular Markov Chain.

Definition 4.1. *Regular Markov Chain[1]*
A Markov Chain is said to be regular if there exist some number k such that, for every $i, j \in I$, the probability of getting from i to j in exactly k steps is > 0.

The theorem is described below:

Theorem 4.2. *If a finite state Markov Chain is regular, then it has a unique stationary distribution.[1]*

I found some theorem to prove theorem 4.2 in [2].

Theorem 4.2 holds because $\mathcal{T}^{(k)}(i \to j) > 0$ in regular Markov Chain. Then, by theorem4.3.

Theorem 4.3. *Let $\mathcal{T}$ be irreducible. Then the following are equivalent:*

1. *every state is positive recurrent;*
2. *some state i is positive recurrent;*

3. $\mathcal{T}$ has an invariant distribution, π say.[2]

We can imply the regular Markov chain is irreducible. Furthermore, we can use the ergodic theorem to show the stationary distribution is the unique stationary distribution.

Theorem 4.4. *Ergodic theorem [2]*
Let P be irreducible and let λ be any distribution. If $(X_n)_{n\geq 0}$ is Markov(λ, P) then

$$\mathbb{P}(\tfrac{V_i(n)}{n} \to \tfrac{1}{m_i} \text{ as } n \to \infty) = 1$$

where $m_i = \mathbb{E}_i(T_i)$ is the expected return time to state i and $V_i(n)$ is the number of visits to state i before n. Moreover, in the positive recurrent case, for any bounded function $f : I \to \mathbb{R}$ we have

$$\mathbb{P}(\tfrac{1}{n}\textstyle\sum_{k=0}^{n-1} f(X_k) \to \bar{f} \text{ as } nto\infty) = 1$$

where

$$\bar{f} = \textstyle\sum_{i\in I} \pi_i f_i$$

and where $(\pi_i : i \in I)$ is the unique invariant distribution.

5 Detailed Balalnce

In MCMC methods, we often use detailed balance to verify the Markov chain has the desired stationary distribution.

Theorem 5.1. *Reversible Markov Chain [1]*
A finite-state Markov chain $\mathcal{T}$ is reversible if there exists a unique distribution π such that, for all $i, j \in I$:

$$\pi(i)\mathcal{T}(i \to j) = \pi(j)(j \to i)$$

This equation is called the detailed balance.

By theorem 5.1, reversibility means that exists unique stationary distribution π, however, reversibility does not guarantee the chain will converge i.e. example 5.2.

Example 5.2. $\mathcal{T} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$*, Then $\mathcal{T}^{(2)} = I$, so $\mathcal{T}^{(2n)} = I$ and $\mathcal{T}^{(2n+1)} = \mathcal{T}$ $\forall n$. Thus, there is no convergence to equilibrium.*

Additionally, the condition of detailed balance is stronger than invariant.

1. invariant : $\pi(x) = \sum_y \pi(y)\mathcal{T}(y \to x)$
2. detailed balance : $\pi(x)\mathcal{T}(x \to y) = \pi(y)\mathcal{T}(y, x)$

If detailed balance holds, $\pi(x) = \sum_y \pi(y)\mathcal{T}(y \to x) = \pi(x)\sum_y \mathcal{T}(x \to y)$. We have $\sum_y \mathcal{T}(x \to y) = 1$ and $\pi(x) = \pi(x)$.

References

[1] Koller, D., Friedman, N. (2009). Probabilistic Graphical Models: Principles and Techniques. MIT Press. ISBN: 9780262013192

[2] J. Norris, "Markov Chains," Cambridge University Press, Cambridge, 1998.