Markov Chain Monte Carlo and Simulated Annealing with Applications and Implementations

Hongda Li

December 2, 2022

Abstract

In this report, we prove the fundamentals for the convergence of the Metropolis Hasting Chain under the discrete case; then introduce some ideas from the continuous case. We discuss the Simulated Annealing algorithm as a particular case of the Metropolis Hasting and use both algorithms to construct several numerical experiments in Julia. The first experiment is sampling from complicated distribution functions on 2D, the second is applying Simulated Annealing for the knapsack problem, and in the third experiment, we test simulated Annealing on the Rastrigin function using different base chains. We collect data and illustrate the behaviors of these algorithms.

1 Introduction

Metropolis Hasting is a type of Markov Chain Monte Carlo (MCMC) method. It uses the convergence of a Markov chain to sample from a distribution. For motivations, it's easy to sample from a 1D distribution function if we have the inverse of the CDF, however, in general, it's very hard to sample from a high-dimension distribution even if we have the PDF function.

The Metropolis Hasting Chain (MHC) is a type of Markov chain whose stationary distribution equals the targeted distribution function. In this report, we are interested in the theoretical foundations for MHC and its applications. Additionally, we are also interested in understanding Simulated Annealing (SA) using stochastic processes.

Definition 1 (Stationary distributions)**.** Let $p(x, y)$ be a transition kernel for a Markov chain with state space that is countable, then π is said to be a stationary distribution for p if it satisfies:

$$\pi(y) = \sum_{x \in S} p(x, y)\pi(x) \quad \forall y \in S.$$

Definition 2 (Detailed balance)**.** Let $p(x, y)$ be the transition kernel for a Markov chain with state space S which is countable, then the distribution π satisfies detailed balance if:

$$\pi(y)p(y, x) = \pi(x)p(x, y) \forall x, y \in S.$$

Remark 1.0.1. Take note that we did in class that if a Markov chain has a distribution such that it satisfies detailed balance, then the distribution is going to be the stationary distribution for the Markov chain.

2 Preliminaries

Theorem 1 (Convergence to stationary distributions)**.** Let $(X_n)_{n \geq 0}$ be a discrete Markov chain with countable/finite state space S. Assuming that it's irreducible, aperiodic, and has a stationary distribution π, then as $n \to \infty$, we have $p^n(x, y) = \pi(y)$.

Proof. The theorem is listed as theorem 1.19 in Rick's book [1]. $\qquad\qquad\square$

Remark 2.0.1. This theorem plays a central role in understanding the regularity conditions for the convergence properties of MHC. Moreover, we skip the discussion regarding the case when the Markov chain has a state space that is uncountable in this report. For more details about the analysis of the ergodic theorem, see chapter 6 of the book by Robert, Casella [2].

2.1 Metropolis Hasting chain and its convergence

In this subsection, we present the proof and theoretical foundations for the convergence of the MHC when the underlying state space is countable. We seek the answer following questions:

1. What is the MHC?

2. What are the regularity conditions that can assert that the stationary distribution is equal to the targeted distribution?

3. What we the regularity conditions imply about the choices of a base chain when it comes to actual applications?

The quantities in algorithm 1 has the following quantities and their expectations:

1. $q(x|y)$ is the base chain, and it should be defined on S.

2. $f(x) : S \mapsto \mathbb{R}_+$ is a probability mass function on the state space S.

3. ρ is the acceptance function, given $X^{(t)}$, it decides whether to accept $Y^{(t+1)}$ from q.

Algorithm 1 Metropolis Chain

1: **Input:** $X^{(t)}$
2: $Y^{(t)} \sim q(\cdot|x^{(t)})$
3: $X^{(t+1)} := \begin{cases} Y^{(t)} & \text{w.p}: \rho(X^{(t)}, Y^{(t)}) \\ X^{(t)} & \text{else} \end{cases}$
4: $\rho(x, y) := \min \left\{ \frac{f(y)}{f(x)} \frac{q(x|y)}{q(y|x)}, 1 \right\}$

Theorem 2 (Staiontary distribution for the MHC)**.** The support set of f, denoted with $\operatorname{supp}(f)$ satisfies detailed balance if the base chain $q(x|y)$ is non-negative for all $x, y \in S$, then the MHC has the support sets of f as the stationary distributions.

References

[1] R. Durrett. *Essentials of Stochastic Processes.* Springer Texts in Statistics. Springer New York, 2012.

[2] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods.* Springer, 2005.