MCMC Section 1: Metropolis-Hasting Algorithm

Yuze Zhou

Date

Introduction

- Given $\mathbf{F}(x)$, the CDF of a certain distribution, our goal is to sample an observation from this distribution or to compute some properties, like the mean of it.
- However, in many cases, the underlying pdf or pmf might be very messy, and we do not have access to easy methods that could sample directly from the distribution or calculate certain expectations explicitly
- The idea of MCMC is to take advantage of the asymptotic properties of Markov Chains to get "close" to the wanted distribution.

Introduction:Ising Model

- Ising Model, named after physicists Ernst Ising, is a mathematical model for ferromagnetism in statistical mechanics.
- The model consists of discrete variables representing magnetic dipole moments of atomic "spins" that can be in one of two states (+1 or 1). The spins are usually arranged in a lattice (where the local structure repeats periodically in all directions), allowing each spin to interact with its neighbors.

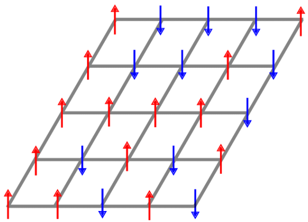

Figure 1: Ising Model on a Lattice

Introduction: Ising Model

- **Definitions:**
  Λ: a set of lattice sites;
  For each lattice site $k \in \Lambda$, there is a discrete variable $\sigma_k \in \{+1, -1\}$;
  A spin configuration $\sigma = \{\sigma_k\}_{k \in \Lambda}$ is an assignment of spin value to each site;

Probability on the Lattice

- **Hamiltonian:** $H(\sigma) = -J \sum_{(i,j)} \sigma_i \sigma_j$

  (i, j) indicates that site i and site j are neighbors on the lattice
  The Hamiltonian is simply an assembling over all interactions between neighbours for a certain spin configuration.
- The probability assigned to a certain spin configuration is:

$$P(\sigma) = \frac{1}{Z_M} exp\{-\beta H(\sigma)\}$$
$$Z_M = \sum_{\sigma} exp\{-\beta H(\sigma)\}$$

- Let M be the number of all different spin configurations, the property we are most concerned with is the **Internal Energy**, which is defined as:

$$U_M = \sum_{\sigma} H(\sigma) \frac{1}{Z_M} exp\{-\beta H(\sigma)\}$$

Introduction: Ising Model

- At first glance, it seems that the Internal Energy could be easily calculated, which is simply the expectation of Hamiltonian over a discrete distribution.

NO!

- Assume our Ising Model is constructed on a 10*10 lattice, then for the total number of spin configurations, $M = 2^{10*10}$, making the seemingly easy computation impossible.

Algorithm 1: Metropolis-Hasting Algorithm for Internal Energy

Initialisation: Choose $\sigma_0 \in \Lambda$;

Set up transition probability $Q(\sigma^{'}|\sigma)$

for *$0 \leq i \leq T-1$* **do**

Generate $\sigma^{'}$ from $Q(\cdot|\sigma_i)$;

Calculate the Hamiltonian $H(\sigma_i)$;

Set $\sigma_{i+1} = \sigma^{'}$ with probability $\alpha(\sigma^{'}|\sigma_i) = \min\{\frac{P(\sigma^{'})}{P(\sigma_i)}\frac{Q(\sigma_i|\sigma^{'})}{Q(\sigma^{'}|\sigma_i)}, 1\}$;

Otherwise set $\sigma_{i+1} = \sigma_i$;

end

Estimate Internal Energy by $\frac{1}{T}\sum_{i=0}^{T-1} H(\sigma_i)$

Review of Markov Chains

For a diecrete Markov Chain $\{X_n\}_{n\in\mathbb{N}_0}$ defined on a countable state space $i \in \mathcal{X}$ with transition probability matrix $P = (P_{ij})$

- **Stationary Distribution:** $\pi = \pi P$, i.e. $\pi_i = \sum_j \pi_j P_{ji}$
- **Detailed Balance Condition:** $\pi_i P_{ij} = \pi_j P_{ji}$
- **Irreducibility:** If all states on the Markov Chain communicates
- **Positive-Recurrent:** $T_j = \inf\{n \geq 1, X_n = j | X_0 = j\}$. If for any state j, $\mathbb{E}T_j < \infty$, then the Chain is called positive recurrent.
- **Aperiodicity:** Let $P^n(i,i) = P(X_n = i|X_0 = i)$, If $gcd(n|P^n(i,i) > 0) = 1$, the state i is called aperiodic. A Markov Chain is called aperiodic if all states are aperiodic.

Review of Markov Chains

- If a Markov Chain is positive recurrent, the stationary distribution is unique with $\pi_i = \frac{1}{\mathbb{E}T_i}$
- If a probability distribution μ defined on the state space of the Markov Chain satisfying the detailed balance condition $\mu_i P_{ij} = \mu_j P_{ji}$, then μ is a stationary distribution for the chain.
- A irreducible Markov Chain defined on a finite state space is positive-recurrent.
- For an irreducible positive-recurrent Markov Chain and a bounded function f:

$$\frac{1}{N}\sum_{n=1}^{N} \mathbf{1}\{X_n = i\} = \pi_i \quad a.s.$$

$$\frac{1}{N}\sum_{n=1}^{N} f(X_n) = \sum_i \pi_i f(i) \quad a.s.$$

Review of Markov Chains

- **Total Variation Distance:** For two probability distributions μ and υ defined on the same discrete space $\mathcal{X}$, the total variation distance is defined as such:

$$d_{TV}(\mu, \upsilon) = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu_(x) - \upsilon(x)|$$

- **Convergence to the Stationary Distribution:** For an irreducible, aperiodic and positive-recurrent Markov Chain with stationary distribution π, and an arbitrary distribution υ on the state space;

$$\lim_{n \to \infty} d_{TV}(\upsilon P^n, \pi) = 0$$
$$\lim_{n \to \infty} \upsilon P^n = \pi$$

Back to Ising Model

Algorithm 2: Metropolis-Hasting Algorithm for Internal Energy

Initialisation: Choose $\sigma_0 \in \Lambda$;

Set up transition probability $Q(\sigma^{'}|\sigma)$

for *0* $\leq i \leq T - 1$ **do**

  Generate $\sigma^{'}$ *from* $Q(\cdot|\sigma_i)$*;*

  Calculate the Hamiltonian $H(\sigma_i)$*;*

  Set $\sigma_{i+1} = \sigma^{'}$ *with probability* $\alpha(\sigma^{'}|\sigma_i) = \min\{\frac{P(\sigma^{'})}{P(\sigma_i)}\frac{Q(\sigma_i|\sigma^{'})}{Q(\sigma^{'}|\sigma_i)}, 1\}$*;*

  Otherwise set $\sigma_{i+1} = \sigma_i$*;*

end

Estimate Internal Energy by $\frac{1}{T}\sum_{i=0}^{T-1} H(\sigma_i)$

Back to the Ising Model

- Assume the transition probability $Q(\sigma^{'}|\sigma)$ guarantees that for a given spin configuration $\sigma^{'}$, it could be accessed from any other spin configuration σ.
- $P(\sigma)\alpha(\sigma^{'}|\sigma) = P(\sigma^{'})\alpha(\sigma|\sigma^{'})$
- Therefore $P(\sigma)$ is the stationary distribution for the Markov Chain with transition probability $\alpha(\sigma^{'}|\sigma)$, and:
$$\frac{1}{T}\sum_{i=0}^{T-1} H(\sigma_i) \to U_M \quad a.s.$$
- Metropolis-Hasting Algorithm provides a valid estimate for the Internal Energy.

In the continuous setting

- $f(x)$ is a certain pdf, the quantity to estimate is
$$\mathbb{E}_f p(x) = \int p(x)f(x)dx$$

Algorithm 3: Metropolis-Hasting for density case

Initialisation: Choose x_0;
Set up transition probability density $f(x^{'}|x)$;
for *0* $\leq i \leq T-1$ **do**
  Generate $x^{'}$ *from* $q(x^{'}|x)$;
  Calculate $p(x_i)$;
  Set $x_{i+1} = x^{'}$ *with probability* $\alpha(x^{'}|x_i) = \min\{\frac{f(x^{'})}{f(x_i)}\frac{q(x_i|x^{'})}{q(x^{'}|x_i)}, 1\}$;
  Otherwise set $x_{i+1} = x_i$;
end
Estimate $\int p(x)f(x)dx$ *by* $\frac{1}{T}\sum_{i=0}^{T-1} p(x_i)$

In the continuous setting

- **Total Variation Distances:** For the continuous setting, for two probability densities $f(x)$ and $g(x)$the total variation distance is defined as:

$$d_{TV}(f,g) = \frac{1}{2}\int |f(x) - g(x)|dx$$

- Let $f_n(x)$ be the density of x_n in the previous algorithm, under some standard conditions:

$$d_{TV}(f_n, f) \to 0 \quad as \quad n \to \infty$$

$$\frac{1}{T}\sum_{i=0}^{T-1} p(x_i) \to \int p(x)f(x)dx \quad a.s.$$

References I