Discrete MALA – A New Method for Sampling Discrete Distribution

Yuze Zhou; Advisor: Zhiqiang Tan

Apr.2021

Ising Model

Definitions:

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

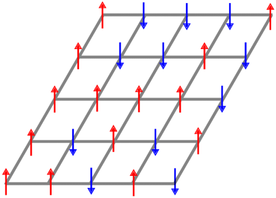


Figure 1: Ising Model

Ising Model

Distribution:

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

  (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.

- Given parameter β (also called inverse temperature as $\beta \propto 1/T$), the probability assigned to a certain spin configuration σ is:

$$P_\beta(x) = \frac{1}{Z_\beta} exp\{-\beta H(x)\}$$
$$Z_\beta = \sum_x exp\{-\beta H(x)\}$$

Ising Model

Properties:

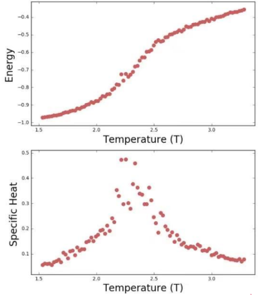
- Two properties of the Ising Model are of special interest
- **Internal Energy:**

$$U(\beta) = \mathbb{E}_\beta \frac{1}{|\Lambda|} H(x)$$

- **Specific Heat:**

$$C(\beta) = \beta^2 Var_\beta[\frac{1}{|\Lambda|} H(x)]$$

- **Phase Transition**: If we choose $J = 1$, $\beta = 1/T$, a spike is observed in the specific heat:



Review of Methods

- For simplicity, denote the distribution we want to sample from the Ising Model is $P(x)$
- **Metropolis-Hasting:**

Algorithm 1: Metropolis-Hasting

Initialisation: Choose initial configuration x_0;

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

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

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

    Set new configuration $x_{i+1} = x^{'}$ with probability

    $\alpha(x^{'}|x_i) = \min\{\frac{P(x^{'})}{P(x_i)}\frac{Q(x_i|x^{'})}{Q(x^{'}|x_i)}, 1\}$;

    Otherwise set $x_{i+1} = x_i$;

end

Review of Methods

- Randomized Gibbs

Algorithm 2: Randomized Gibbs

Initialisation: Choose initial configuration x_0;

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

Randomly choose site $k \in \Lambda$

From the current configuration x_i, *sample* $P(\sigma_k|\sigma_{-k})$ *to get new configuration* x_{i+1}

end

Discrete MALA

- **Global Discrete MALA:**
- For a given configuration x, denote $\mathcal{N}(x)$ as the set of neighboring configurations of x, which is achieved by flipping a single site from x,

---

Algorithm 3: Global Discrete MALA

Initialisation: Choose initial configuration x_0;

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

  sample $\tilde{x} \in \mathcal{N}(x)$ *with probability* $\frac{1}{|\mathcal{N}(x)|}$

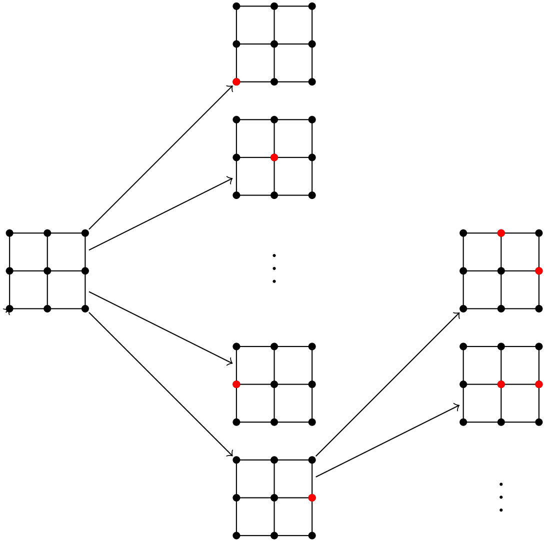
  sample $x^{'} \in \mathcal{N}(\tilde{x})$ *with probability* $\frac{p(x^{'})}{\sum\limits_{y \in \mathcal{N}(\tilde{x})} p(y)}$

  Accept $x^{'}$ *as the new configuration* x_{i+1}

end

- It can be shown that Global Discrete MALA would satisfy Detailed Balance Condition.

Discrete MALA



Discrete MALA

- **Local Discrete MALA:**
- In global discrete MALA, we have to browse through all $|\mathcal{N}(x)|$ neighbours of x, sometimes the memory and computational cost will be too high.
- We select an anchor point $k \in \Lambda$, and denote C_k as the set of sites that are in a certain small neighbourhood of k
- For a configuration x, instead of considering all candidates in $\mathcal{N}(x)$, we only consider neighboring configurations that differ from x by flipping a site in C_k, which we denotes as $C_k(x)$.

Discrete MALA

- Using the same notations as above, the local discrete MALA algorithm goes as followed:

Algorithm 4: Local Discrete MALA

Initialisation: Choose initial configuration x_0;

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

Randomly select $k \in \Lambda$ *as the anchor point*

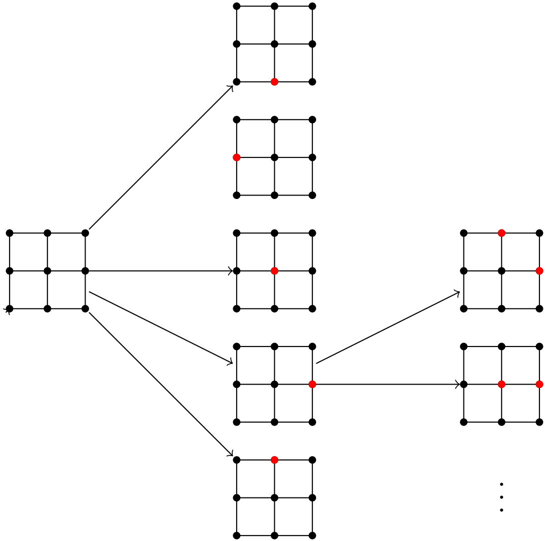
sample $\tilde{x} \in C_k(x)$ *with probability* $\frac{1}{|C_k(x)|}$

sample $x^{'} \in C_k(\tilde{x})$ *with probability* $\frac{p(x^{'})}{\sum\limits_{y \in C_k(\tilde{x}))} p(y)}$

Accept $x^{'}$ *as the new configuration* x_{i+1}

end

Discrete MALA



Results

- First we run Metroplis, Randomized Gibbs, Local Discrete MALA with a C_k of size 5 and global discrete mala with 2000 burn-in samples and a sample size of 2000.

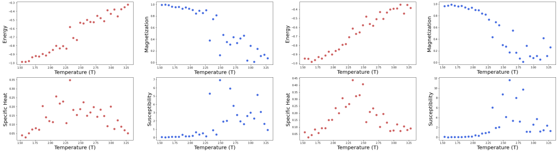


Figure 3: Global Discrete MALA

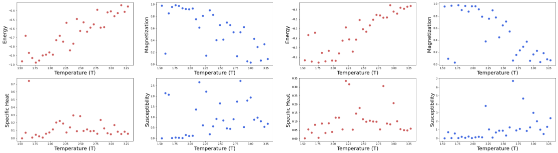


Figure 4: Local Discrete MALA

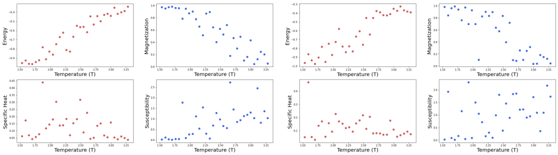


Figure 5: Metropolis

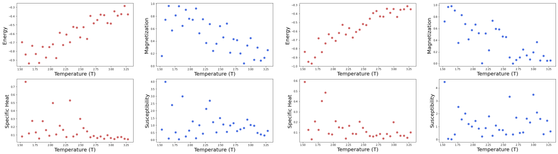
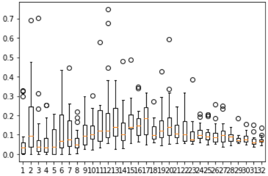


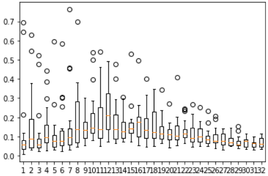
Figure 6: Randomized Gibbs

Results

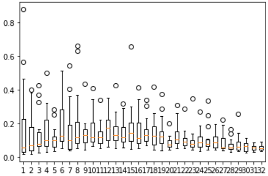
- Although Global Discrete MALA shows very good convergence properties, the computational and memory cost is rather high and undesirable. Thus, we would like to only compare the results of Local Discrete MALA, Metropolis and Randomized Gibbs.
- We first compare results from 20 independent runs with 2000 samples and 2000 burn-in samples and give the boxplots for specific heat.



(a) Local Discrete MALA



(b) Metropolis

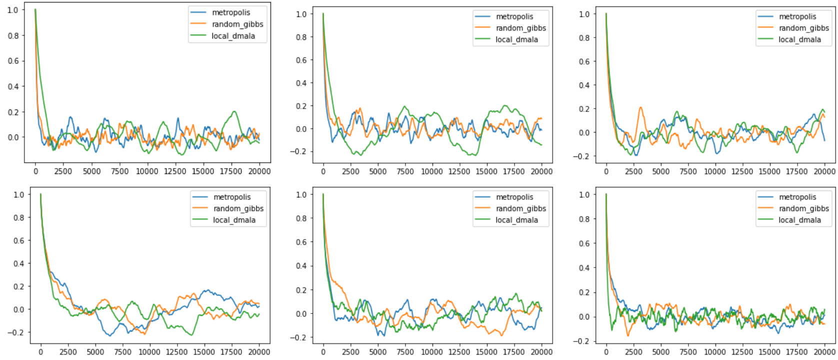


(c) Randomized Gibbs

Figure 7: Boxplot

Results

- We also give the acf plots of the Internal Energies of the samples at different temperature from different methods:



Discussion

- **Trade-off:** Increasing the size of neighbourhood C_k will lead to faster convergence within the samples but the computational and memory cost would also increase.
- **Auto-correlation:** Though Local Discrete-MALA might outperform both Metropolis and Randomized Gibbs when estimating both Internal Energies and Specific Heats, the acf plots display worse patterns compared to the other two when the Temperature is lower (the acceptance rate is also lower at the same time).