APMA1690: Homework # 9 (Due by 11pm on November 30)

"*He was too simple to wonder when he had attained humility. But he knew he had attained it, and he knew it was not disgraceful and it carried no loss of true pride.*"

— sentences from *The Old Man and the Sea*, by Ernest Hemingway

1 Review

I would suggest you go through the review section before going to the problem set.

1.1 Some Concepts in Graph Theory

We list as follows the concepts needed for graphical models

- A **graph** is an ordered pair $G = (V, E)$ comprising:
  - V is a set of **vertices**; the elements of V are usually indexed by positive integers.
  - E is a subset of $\{(i, j) \mid i, j \in V \text{ and } i \neq j\}$, i.e., pairs of vertices; if we view the pairs as unordered, i.e., $(i, j) = (j, i)$ for all i and j, this graph is called an **undirected graph**, otherwise, it is called a **directed graph**. (We focus on undirected graphs when talking about graphical models, i.e., we will keep assuming $(i, j) = (j, i)$.)
- For a given graph $G = (V, E)$, two vertices i and j are called **adjacent** if there is an edge between them, i.e., $(i, j) = (j, i) \in E$.
- For a given graph $G = (V, E)$ and a vertex $i \in V$, the **neighborhood** of i is defined as the collection of vertices adjacent to i, i.e., $\mathcal{N}(i) := \{j \in V \mid (i, j) \in E\}$. Since we do not allow 'self connecting edges' (i.e., vertices i and j should be different if then form an edge $(i, j) \in E$), we have $i \notin \mathcal{N}(i)$.
- For a given graph $G = (V, E)$, a set of vertices (i.e., a subset $c \subset V$) is called a **clique** if every pair of vertices in this set are adjacent. Furthermore, we denote $\mathcal{C}(G) := \{\text{all cliques of graph } G\}$. In addition, we adopt the convention that each vertex itself is a clique, i.e., $\{i\} \in \mathcal{C}(G)$ for all $i \in V$.

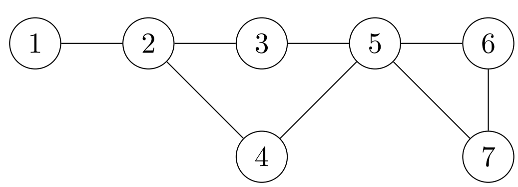


Figure 1: A graph $G = (V, E)$ with $V = \{1, 2, \ldots, 7\}$
and $E = \{(1, 2), (2, 3), (2, 4), (3, 5), (4, 5), (5, 6), (5, 7), (6, 7)\}$.

Figure 1 provides an example of a graph. In this graph, for example, vertices 1 and 2 are adjacent, vertices 2 and 3 are adjacent. Additionally, the neighborhood of vertex 5 is $\mathcal{N}(5) = \{3, 4, 6, 7\}$. We list all the elemnts of $\mathcal{C}(G)$ as follows

$$\begin{aligned} &\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7\}, \\ &\{1,2\}, \{2,3\}, \{2,4\}, \{3,5\}, \{4,5\}, \{5,6\}, \{5,7\}, \{6,7\}, \\ &\{5,6,7\}. \end{aligned} \tag{1.1}$$

1.2 Notations

We also adopt the following notations

- Let $G = (V, E)$ be a graph with $V = \{1, 2, \ldots, d\}$ and $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)^\intercal$ a vector in the product space $\mathcal{X} = \mathcal{X}_1 \times \cdots \mathcal{X}_d$.
- For any subset $c = \{i_1, \ldots, i_k\} \subseteq V$ with $i_1 < i_2 < \cdots < i_k$, we denote $x_c := (x_{i_1}, x_{i_2}, \ldots, x_{i_k})^\intercal$ and $\mathcal{X}_c := \mathcal{X}_{i_1} \times \cdots \times \mathcal{X}_{i_k}$. For example, let $V = \{1, \ldots, 7\}$ and $c = \{1, 3, 5\}$, then $x_c = (x_1, x_3, x_5)^\intercal$ and $\mathcal{X}_c = \mathcal{X}_1 \times \mathcal{X}_3 \times \mathcal{X}_5$.

1.3 Application of Gibbs Sampling to the 2-Dimensional Ising Model

This subsection helps discover a classical result from **?**[1] in a numerical way.

1.3.1 Periodic Lattices

`(This subsection on periodic lattices is only for HW 9 and will not be required for the final exam.)`

We use periodic lattices, following the convention in the standard literature on the Ising model. The periodic structure will make the coding part easier. Without the periodic structure, we would have to deal with the boundaries of nonperiodic lattices separately.

A 3-by-3 periodic lattice Λ_3 is presented in Figure 2. An N-by-N periodic lattice Λ_N with a generic size N is defined in a similar way. Specifically, for the vertex in the i-th row and j-th column of an N-by-N periodic lattice, its neighbors are the following four vertices[2]

- the vertex in the k-th row and l-th column, where $k \in \{1, \ldots, N\}$ with $k-1 \equiv i-1 \mod (N)$ and $l \in \{1, \ldots, N\}$ with $l - 1 \equiv j - 2 \mod (N)$.
- the vertex in the k-th row and l-th column, where $k \in \{1, \ldots, N\}$ with $k-1 \equiv i-1 \mod (N)$ and $l \in \{1, \ldots, N\}$ with $l - 1 \equiv j \mod (N)$.
- the vertex in the k-th row and l-th column, where $k \in \{1, \ldots, N\}$ with $k - 1 \equiv i \mod (N)$ and $l \in \{1, \ldots, N\}$ with $l - 1 \equiv j - 1 \mod (N)$.

[1] Professor Onsager taught statistical mechanics at Brown University and did the research work important enough to gain him the unshared Nobel Prize in Chemistry in 1968. However, "the Great Depression limited Brown's ability to support a faculty member who was only useful as a researcher and not a teacher; he was let go by Brown, being hired by Yale University" (see Wikipedia).

[2] For integers a and b, the notation '$a \equiv b \mod (N)$' denotes the following: N is a divisor of $a - b$, i.e., there exists an integer k (not necessarily positive) such that $a - b = kN$.

- the vertex in the k-th row and l-th column, where $k \in \{1, \dots, N\}$ with $k-1 \equiv i-2 \mod (N)$ and $l \in \{1, \dots, N\}$ with $l-1 \equiv j-1 \mod (N)$.

Neighborhoods of 5 and 3 of the 3-by-3 periodic lattice Λ_3 (see Figure 2) are presented in Figures 3 and 4, respectively. The two figures visually interpret the complicated modular arithmetic notation 'mod (N)' above.

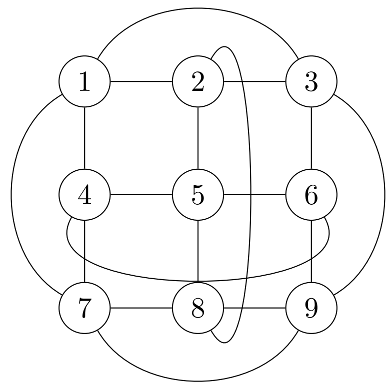


Figure 2: A 3-by-3 periodic lattice Λ_3, which is a graph.

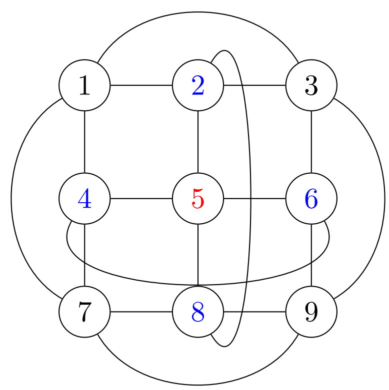


Figure 3: The **neighborhood** of vertex 5 (red) is $\mathcal{N}(5) = \{2, 4, 6, 8\}$, presented in blue.

1.3.2 Ising Model PMF

Suppose Λ_N is an N-by-N periodic lattice. Each vertex (also called a 'site,' i.e., Section 18.3 of **?**) is labeled by a positive integer, e.g., vertices of a 3-by-3 periodic lattice (see Figure 2) are labeled by $\{1, 2, \dots, 9\}$.

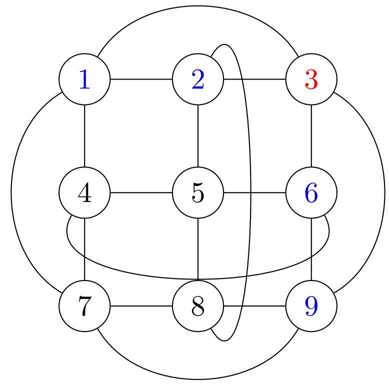


Figure 4: The **neighborhood** of vertex 3 (red) is $\mathcal{N}(3) = \{2, 1, 6, 9\}$, presented in blue.

At each vertex $k \in \Lambda_N$, there exists an atom with spin $s_k \in \{-1, 1\}$. A **spin configuration** $\boldsymbol{s}$ is a vector of spin values assigned to the N^2 vertices, i.e.,

$$\boldsymbol{s} = (s_k)_{k\in\Lambda} = (s_1, s_2, \ldots, s_{N^2}) .$$

Let $\mathcal{X}$ denote the collection of all possible spin configurations $\boldsymbol{s}$, i.e.,

$$\mathcal{X} = \underbrace{\{-1, 1\} \times \{-1, 1\} \times \cdots \times \{-1, 1\}}_{\text{Cartesian product of } N^2 \text{ sets}} . \tag{1.2}$$

It is straightforward that $\#\mathcal{X} = 2^{N^2}$.

Because of the perturbation from heat, the spin values are not deterministic, i.e, $\boldsymbol{s}$ is a random vector. The Ising model gives PMFs describing the randomness of $\boldsymbol{s}$. A simplified Ising model-based PMF is the following one characterizing a "ferromagnetic zero-field" model[3]

$$\begin{aligned} \pi_{N,\beta}(\boldsymbol{s}) &= \frac{1}{Z_\beta} \cdot \exp\left\{ \beta \cdot \sum_{\langle i,j \rangle} s_i s_j \right\} \\ &= \frac{1}{Z_\beta} \cdot \exp\left\{ \beta \cdot \left(\sum_{\text{all edges } (i,j) \text{ on the } N\text{-by-}N \text{ periodic lattice}} s_i s_j \right) \right\}, \end{aligned} \tag{1.3}$$

where $\langle i, j \rangle$ denotes that i and j are neighbors. The PMF formula in Eq. (1.3) is called the 2-dimensional Ising model[4]. The notations above are explained as follows:

- $\sum_{\langle i,j \rangle}$ denotes the sum over all pairs (i, j) such that i and j are neighbors. It is the standard notation in the Ising model literature, e.g., the Wikipedia page on the Ising Model.

[3]The PMF $\pi_{N,\beta}(\boldsymbol{s}) = \frac{1}{Z_\beta} \cdot \exp\left\{ -\beta \cdot \sum_{\langle i,j \rangle} s_i s_j \right\}$ with an extra negative sign represents an "antiferromagnetic zero-field" model, which is not of interest.

[4]German pronunciation is like the English word "easing" (see the remark in Example 18.16 of **?**).

- β is a model parameter. Specifically, $\beta = 1/T$, where T indicates the temperature of the magnetic system.
- $Z_\beta = \sum_{\boldsymbol{s}\in\mathcal{X}} \exp\left\{\beta \cdot \sum_{\langle i,j\rangle} s_i s_j\right\}$ is the normalizing constant. As a function of β, the quantity Z_β is the partition function of the model. The letter Z stands for the German word *Zustandssumme*, "sum over states," and $\sum_{\boldsymbol{s}\in\mathcal{X}}$ is the sum over all possible spin configurations.
- $H_N(\boldsymbol{s}) = -\sum_{\langle i,j\rangle} s_i s_j$ presents the energy of the magnetic system, and $\pi_{N,\beta}(\boldsymbol{s}) = \frac{1}{Z_\beta} \cdot \exp\{-\beta \cdot H_N(\boldsymbol{s})\}$ is referred to as the Boltzmann distribution of the magnetic system in the statistical mechanics literature. Notably, the negative sign in $H_N(\boldsymbol{s})$ and the negative sign in the Boltzmann distribution are canceled out. Forgetting to cancel out the negative sign will make a ferromagnetic model become an antiferromagnetic model.

1.3.3 Curie Temperature

The Curie temperature is the one above which certain materials lose their permanent magnetic properties. The Curie temperature is named after Pierre Curie. In this section, we describe the Curie temperature using the Ising model. The discussion in this section provides a nontrivial question that can be solved by the Gibbs sampling algorithm.

Suppose the distribution $\pi_{N,\beta}(\boldsymbol{s})$ of spin configurations $\boldsymbol{s}$ is given by the ferromagnetic zero-field model in Eq. (1.3). Macroscopically, the individual spins cannot be observed, but the following average magnetization can

$$m_N(\beta) := \sum_{\boldsymbol{s}\in\mathcal{X}} \pi_{N,\beta}(\boldsymbol{s}) \cdot \left|\frac{\sum_{i\in\Lambda_N} s_i}{N^2}\right| = \mathbb{E}_{\pi_{N,\beta}} \left|\frac{\sum_{i\in\Lambda_N} s_i}{N^2}\right|, \tag{1.4}$$

where "$\sum_{\boldsymbol{s}\in\mathcal{X}}$" means the sum over all possible spin configurations. If we consider a very large system, then we are close to the so-called thermodynamic limit

$$m(\beta) := \lim_{N\to\infty} m_N(\beta). \tag{1.5}$$

The interpretation of $m(\beta)$ is roughly presented as follows: when $m(\beta) > 0$, the material of interest is magnetic; when $m(\beta) = 0$, it is not magnetic. The magic of Curie Temperature comes from this limiting procedure in Eq. (1.5). **?** essentially showed that there exists a critical value $\beta_C = \frac{1}{T_C}$, where T_C is the Curie temperature of interest, such that

$$m(\beta) \begin{cases} > 0, & \text{if } \beta > \beta_C, \\ = 0, & \text{if } \beta < \beta_C. \end{cases} \tag{1.6}$$

That is, we have the Curie temperature T_C is represented as follows

$$T_C = \frac{1}{\beta_C}. \tag{1.7}$$

Below the Curie temperature, the material of interest is magnetic; above the Curie temperature, it is not. **?** provided the exactly value of β_C as follows

$$\boxed{\beta_C = \frac{\log(1+\sqrt{2})}{2} \approx 0.44.} \tag{1.8}$$

1.3.4 Estimation of β_C

Mathematically deriving the value β_C in Eq. (1.8) is nearly a Nobel prize-level question. Using the Gibbs sampling, we can estimate this value numerically. To estimate β_C, we need to estimate the $m(\beta)$ defined in Eq. (1.5). When the size N of the periodic lattice Λ_N is large, we have $m(\beta) \approx m_N(\beta)$.

Suppose we have a homogeneous Markov chain (HMC)

$$\left\{ \boldsymbol{X}^{(n)} = \left(X_1^{(n)}, X_2^{(n)}, \ldots, X_{N^2}^{(n)} \right) \right\}_{n=0}^{\infty} \tag{1.9}$$

whose state space is the $\mathcal{X}$ defined in Eq. (1.2); furthermore, this HMC is irreducible and apperiodic, and its stationary distribution is the $\pi_{N,\beta}(\boldsymbol{s})$ defined in Eq. (1.3). Visualizations of $\boldsymbol{X}^{(10000)}$ with lattice size $N = 100$ for different β values are presented in Figure 5.

Then, the ergodic theorem implies the following with probability one

$$\lim_{n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\sum_{k\in\Lambda_N} X_k^{(i)}}{N^2} \right| \right\} = m_N(\beta).$$

Therefore, when both the lattice size N and sample size n are large, we have

$$m(\beta) \approx m_N(\beta) \approx \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\sum_{k\in\Lambda_N} X_k^{(i)}}{N^2} \right| =: \widehat{v}_{N,n}(\beta). \tag{1.10}$$

Then, we can apply Eq. (1.10) (i.e., the estimator $\widehat{v}_{N,n}(\beta)$) to estimate $m(\beta)$, which help us estimate β_C. Now, it remains to get such an HMC in Eq. (1.9), which can be done through the Gibbs sampling.

1.3.5 Application of the Gibbs Sampling to Ising Model

For each vertex i, we define the neighborhood $\mathcal{N}(i)$ of i by the following

$$\mathcal{N}(i) := \{\text{the neighbors of site } i\}.$$

Suppose we are interested in the spin s_{i^*} at the vertex i^*. The Boltzmann distribution $\pi_{N,\beta}(\boldsymbol{s})$ in Eq. (1.3) can be represented as follows

$$\pi_{N,\beta}(\boldsymbol{s}) = \frac{1}{Z_\beta} \cdot \exp\left\{ \beta \left[\sum_{j\in\mathcal{N}(i^*)} s_{i^*} s_j \right] + \beta \left[\sum_{\langle i',j'\rangle \text{ and } i',j'\neq i^*} s_{i'} s_{j'} \right] \right\}. \tag{1.11}$$

1.4 Conditional Distributions

A key quantity in the Gibbs sampling (see Algorithm 1) is the conditional distribution of s_{i^*} given the values of other coordinates, i.e.,

$$\pi_{N,\beta}(s_{i^*} | \boldsymbol{s}_{-i^*}) = \pi_{N,\beta}(s_{i^*} | s_1, \ldots, s_{i^*-1},\, s_{i^*+1}, \ldots, s_{N^2}).$$

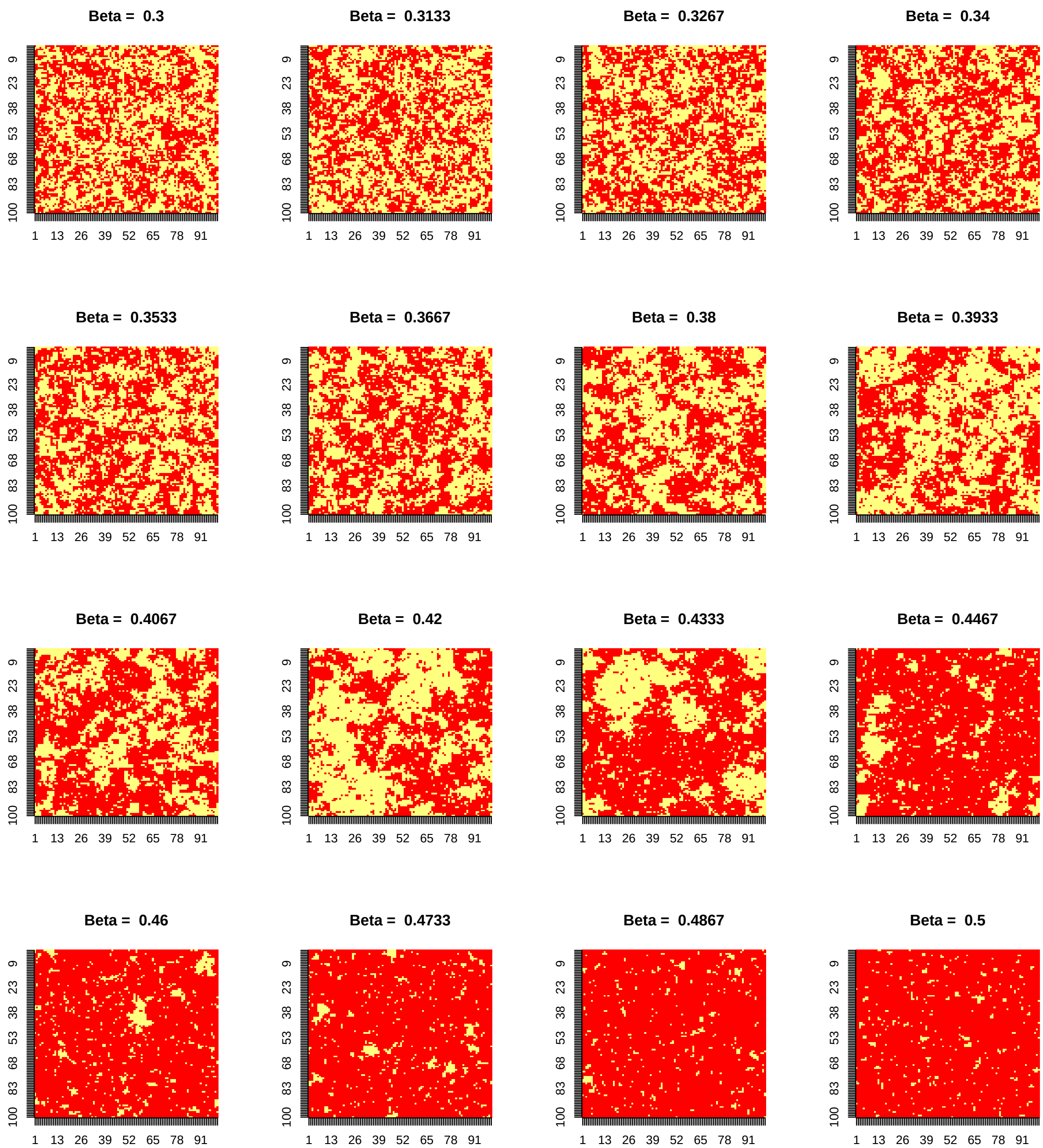


Figure 5: Visualizations of $\boldsymbol{X}^{(10000)}$ with lattice size $N = 100$ for different β values. Recall the critical value $\beta_C = \frac{\log(1+\sqrt{2})}{2} \approx 0.44$.

Using the representation in Eq. (1.11), the conditional distribution $\pi_{N,\beta}(s_{i^*}|\boldsymbol{s}_{-i^*})$ can be represented as follows

(1.12)
$$\begin{aligned}\pi_{N,\beta}(s_{i^*}|\boldsymbol{s}_{-i^*}) &= \frac{\pi_{N,\beta}(\boldsymbol{s})}{\sum_{s_{i^*}\in\{-1,1\}}\pi_{N,\beta}(\boldsymbol{s})}\\ &= \frac{\frac{1}{Z_\beta}\cdot\exp\left\{\beta\left[\sum_{\langle i',j'\rangle \text{ and } i',j'\neq i^*} s_{i'}s_{j'}\right]\right\}\cdot\exp\left\{\beta\left[\sum_{j\in\mathcal{N}(i^*)} s_{i^*}s_j\right]\right\}}{\frac{1}{Z_\beta}\cdot\exp\left\{\beta\left[\sum_{\langle i',j'\rangle \text{ and } i',j'\neq i^*} s_{i'}s_{j'}\right]\right\}\cdot\sum_{s_{i^*}\in\{-1,1\}}\exp\left\{\beta\left[\sum_{j\in\mathcal{N}(i^*)} s_{i^*}s_j\right]\right\}}\\ &= \frac{\exp\left\{\beta\left[\sum_{j\in\mathcal{N}(i^*)} s_{i^*}s_j\right]\right\}}{\sum_{s_{i^*}\in\{-1,1\}}\exp\left\{\beta\left[\sum_{j\in\mathcal{N}(i^*)} s_{i^*}s_j\right]\right\}}\\ &= \frac{\exp\left\{\beta\left[\sum_{j\in\mathcal{N}(i^*)} s_{i^*}s_j\right]\right\}}{\exp\left\{-\beta\left[\sum_{j\in\mathcal{N}(i^*)} s_j\right]\right\}+\exp\left\{\beta\left[\sum_{j\in\mathcal{N}(i^*)} s_j\right]\right\}},\end{aligned}$$

More precisely, we have the following

$$\begin{aligned}\pi_{N,\beta}(\,1\,|\boldsymbol{s}_{-i^*}) &= \frac{\exp\left\{\beta\left[\sum_{j\in\mathcal{N}(i^*)} s_j\right]\right\}}{\exp\left\{-\beta\left[\sum_{j\in\mathcal{N}(i^*)} s_j\right]\right\}+\exp\left\{\beta\left[\sum_{j\in\mathcal{N}(i^*)} s_j\right]\right\}},\\ \pi_{N,\beta}(\,-1\,|\boldsymbol{s}_{-i^*}) &= \frac{\exp\left\{-\beta\left[\sum_{j\in\mathcal{N}(i^*)} s_j\right]\right\}}{\exp\left\{-\beta\left[\sum_{j\in\mathcal{N}(i^*)} s_j\right]\right\}+\exp\left\{\beta\left[\sum_{j\in\mathcal{N}(i^*)} s_j\right]\right\}}.\end{aligned} \tag{1.13}$$

The factor $\frac{1}{Z_\beta}\cdot\exp\left\{\beta\left[\sum_{\langle i',j'\rangle \text{ and } i',j'\neq i^*} s_{i'}s_{j'}\right]\right\}$ in Eq. (1.12) is canceled out. This cancelation is extremely important as it reduces the redundant computation.

1.5 Sampling from Conditional Distributions

To sample a random variable $X_{i^*}^{(n)}$ from the conditional distribution $\pi_{N,\beta}(\,\cdot\,|\boldsymbol{s}_{-i^*})$ defined in Eq. (1.13), we can adopt the following procedure:

- Compute $a \leftarrow \exp\left\{\beta\left[\sum_{j\in\mathcal{N}(i^*)} s_j\right]\right\}$.
- Compute $b \leftarrow \exp\left\{-\beta\left[\sum_{j\in\mathcal{N}(i^*)} s_j\right]\right\}$.
- Compute $p \leftarrow \frac{a}{a+b}$.
- Generate $Z \sim \text{Bernoulli}(p)$, i.e., $\mathbb{P}(Z=1)=p$ and $\mathbb{P}(Z=0)=1-p$.
- $X_{i^*}^{(n)} \leftarrow 2\cdot Z-1$, i.e., $\mathbb{P}(X_{i^*}^{(n)}=1)=\frac{a}{a+b}$ and $\mathbb{P}(X_{i^*}^{(n)}=-1)=\frac{b}{a+b}$.

With the procedure above, we can sample the HMC in Eq. (1.9) through Gibbs sampling (Algorithm 1).

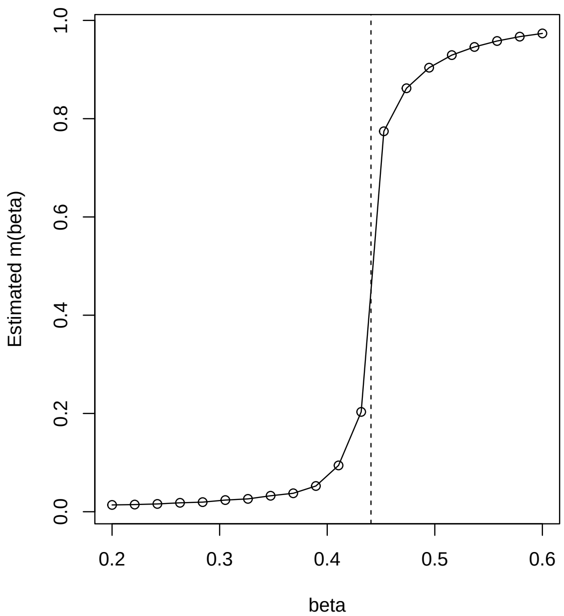


Figure 6: The '$\widehat{v}_{N,n}(\beta)$ vs. β' plot. The vertical dashed line indicates the critical value $\beta_C \approx 0.44$.

1.5.1 Numerical Experiment

As an ending section, we present a numerical experiment of the estimation $m(\beta) \approx \widehat{v}_{N,n}(\beta)$ presented in Eq. (1.10). In this experiment, we choose the lattice size to be $N = 100$ and the sample size to be $n = 10000$. The '$\widehat{v}_{N,n}(\beta)$ vs. β' plot is presented in Figure 6, where the vertical dashed line indicates the critical value $\beta_C \approx 0.44$. Figure 6 is compatible with Eq. (1.6).

1.5.2 Appendix: Gibbs Sampling

Algorithm 1 : Gibbs Sampling

Input: (i) the given distribution $\pi(\xi_1, \ldots, \xi_d)$; (ii) a starting point $\boldsymbol{x}^{(0)} = (\xi_1^{(0)}, \ldots, \xi_d^{(0)})^\intercal$.

Output: A homogeneous Markov chain $\{\boldsymbol{X}^{(n)} = (\xi_1^{(n)}, \ldots, \xi_d^{(n)})^\intercal\}_{n=0}^\infty$ with π as its stationary distribution.

Set $\boldsymbol{X}^{(0)} \leftarrow \boldsymbol{x}^{(0)}$.
for all $n = 1, 2, \ldots,$ **do**
  Sample $\xi_1^{(n)} \sim \pi_{1|-1}(\xi_1 | \xi_2^{(n-1)}, \ldots, \xi_d^{(n-1)})$
  for all $j = 2, \ldots, d$ **do**
    Sample $\xi_j^{(n)} \sim \pi_{j|-j}(\xi_j | \xi_1^{(n)}, \ldots, \xi_{j-1}^{(n)}, \xi_{j+1}^{(n-1)}, \ldots, \xi_d^{(n-1)})$
  end for $\boldsymbol{X}^{(n)} \leftarrow (\xi_1^{(n)}, \xi_2^{(n)}, \ldots, \xi_d^{(n)})^\intercal$
end for

2 Problem Set

1. (3 points) Let $G = (V, E)$ be a graph. Prove the following identity

$$\mathcal{N}(i) \bigcup \{i\} = \left(\bigcup_{c \in \mathcal{C}(G) \text{ and } i \in c} c \right), \tag{2.1}$$

where the union sign $\bigcup$ on the right hand side of Eq. (2.1) denotes the union over all cliques c such that $i \in c$. For the details of the union notation, please refer to the Wikipedia page on 'Union (set theory),' especially the "arbitrary unions" section therein.

Hint: see the definition of cliques.

Remark: The set identity in Eq. (2.1) partially implies the Hammersley–Clifford theorem.

$$\begin{aligned} \mathcal{N}(i) \cup \{i\} &= \{j \in V : (i, j) \in E\} \cup \{i\} \\ &= \{i, j_1, \ldots,\ j_k :\ (i, j_1), \ldots, (i, j_k) \in E\} \\ &= \{c \in \mathcal{C}(G) : i \in c\} \\ &= \bigcup_{\substack{c \in \mathcal{C}(G) \\ i \in c}} c \quad \blacksquare \end{aligned}$$

2. (7 points) In this question, we consider the Ising model on a 100-by-100 **periodic** lattice (i.e., the lattice size $N = 100$). We focus on the following 20 values of the parameter β

$$\beta \in \{0.2 + 0.02k \,|\, k = 1, 2, \ldots, 20\}\,. \tag{2.2}$$

Please do the following tasks using Gibbs sampling:

- For each of the β values in Eq. (2.2), generate the sequence

$$\left\{\boldsymbol{X}^{(i)} = \left(X_1^{(i)}, X_2^{(i)}, \ldots, X_{N^2}^{(i)}\right)\right\}_{i=0}^{1000}, \tag{2.3}$$

  which is the first 1000 (or 1001) components of an irreducible and aperiodic HMC with the $\pi_{N,\beta}$ in Eq. (1.3) as its stationary distribution.
- Plot $\boldsymbol{X}^{(1000)}$ for $\beta \in \{0.32, 0.4, 0.44, 0.6\}$. Each of your four plots should be similar to one of the panels in Figure 5.
- For each of the twenty β values in Eq. (2.2), compute

$$\widehat{v}(\beta) := \frac{1}{1000}\sum_{i=1}^{1000}\left|\frac{\sum_{k\in\Lambda_N} X_k^{(i)}}{N^2}\right|.$$

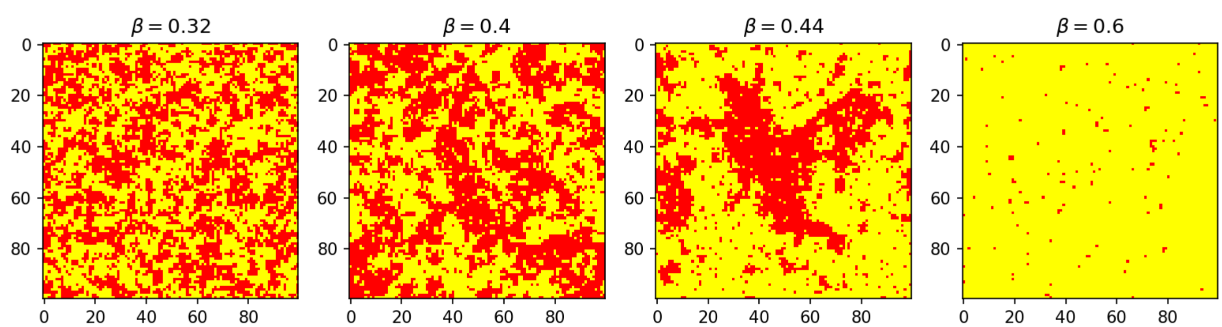
- List the following twenty values of $\widehat{v}(\beta)$

$$\left\{\widehat{v}(\beta) \,\middle|\, \beta = 0.2 + 0.02k \text{ for } k = 1, 2, \ldots, 20\right\}. \tag{2.4}$$

- Plot the twenty values in Eq. (2.4) in a '$\widehat{v}(\beta)$ vs. β' fashion, which should be similar to Figure 6.

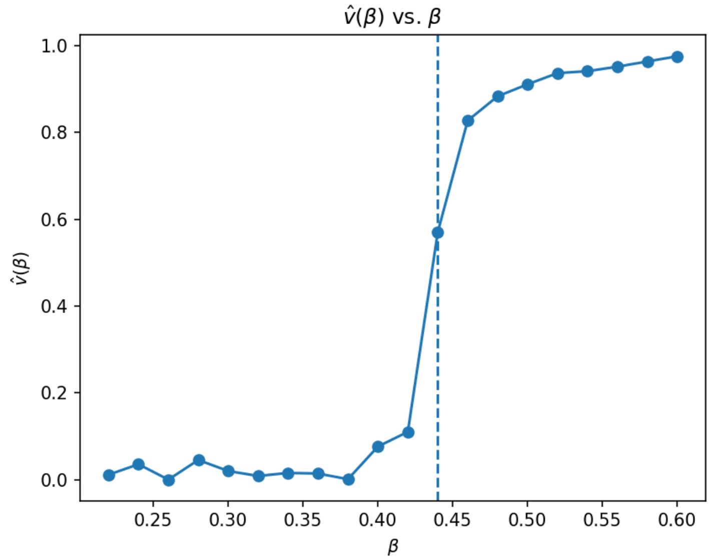
Please provide your code for completing each of the tasks.

Here are the plots for $\beta \in \{0.32, 0.4, 0.44, 0.6\}$:



The full list of values for $\widehat{v}(\beta)$ along with the plot of $\widehat{v}(\beta)$ vs. β are below:

β	$\widehat{v}(\beta)$
0.22	0.0116
0.24	0.0358
0.26	0.0004
0.28	0.0456
0.30	0.0202
0.32	0.0088
0.34	0.0157
0.36	0.0144
0.38	0.0014
0.40	0.0706
0.42	0.1104
0.44	0.5702
0.46	0.8272
0.48	0.8832
0.50	0.9108
0.52	0.9366
0.54	0.9410
0.56	0.9514
0.58	0.9634
0.60	0.9754



Finally, here is the full Python implementation of the model, plots, and estimator:

```python
import numpy as np
import matplotlib.pyplot as plt
from numba import njit

@njit
def bernoulli(p: float) -> int:
    if np.random.uniform() < p:
        return 1
    else:
        return 0

@njit
def modulo(coord : int) -> int:
    return coord % LATTICE_SIZE

@njit
def indexify(s : int) -> int:
    #Converts {-1, 1} to {0, 1} for Python indexing
    return (s+1)//2

@njit
def init_probability_matrix(beta : float) -> list[list[float]]:
    p = np.zeros((2, 2, 2, 2))
    for s1 in range(2):
        for s2 in range(2):
            for s3 in range(2):
                for s4 in range(2):
                    a = np.exp(beta * (2*(s1+s2+s3+s4)-4))
                    b = np.exp(-beta * (2*(s1+s2+s3+s4)-4))

                    p[s1, s2, s3, s4] = a/(a+b)
    return p
```

```python

@njit
def ising(n : int, beta : float, x0 : list[list[int]]) -> list[list[list[int]]]:
    #TRANSLATED AND MODIFIED FROM PROVIDED R CODE
    #Returns a list of n 2D arrays of size N x N whose entries are samples from $\pi_{N, \beta}$

    p = init_probability_matrix(beta)

    X = [x0]
    for k in range(n):
        Xi = X[k]
        for j in range(LATTICE_SIZE):
            jp1 = modulo(j+1)
            jm1 = modulo(j-1)

            for i in range(LATTICE_SIZE):
                ip1 = modulo(i+1)
                im1 = modulo(i-1)

                pij = p[indexify(Xi[ip1,j]),
                        indexify(Xi[im1,j]),
                        indexify(Xi[i,jp1]),
                        indexify(Xi[i,jm1])]

                Z = bernoulli(pij)
                Xi[i, j] = 2*Z - 1

        X.append(Xi)
    return X
```

```
@njit
def estimator(beta: float, x0 : list[list[int]]) -> float:
    print("Estimating with beta =", beta)
    X = ising(CHAIN_LENGTH, beta, x0)

    total = 0
    for i in range(CHAIN_LENGTH):
        total += abs(np.sum(X[i]) / (LATTICE_SIZE**2))

    return total/CHAIN_LENGTH

def plot_X1000(x0 : list[list[int]]) -> None:
    ax1 = plt.subplot(1, 4, 1)
    ax1.imshow(ising(CHAIN_LENGTH, 0.32, x0)[-1], cmap='autumn')
    ax1.set_title(r'$\beta = 0.32$')
    print("Plotting beta = 0.32")

    ax2 = plt.subplot(1, 4, 2)
    ax2.imshow(ising(CHAIN_LENGTH, 0.4, x0)[-1], cmap='autumn')
    ax2.set_title(r'$\beta = 0.4$')
    print("Plotting beta = 0.45")

    ax3 = plt.subplot(1, 4, 3)
    ax3.imshow(ising(CHAIN_LENGTH, 0.44, x0)[-1], cmap='autumn')
    ax3.set_title(r'$\beta = 0.44$')
    print("Plotting beta = 0.44")

    ax4 = plt.subplot(1, 4, 4)
    ax4.imshow(ising(CHAIN_LENGTH, 0.6, x0)[-1], cmap='autumn')
    ax4.set_title(r'$\beta = 0.6$')
    print("Plotting beta = 0.6")
```

```python

def plot_estimates(x0):
    betas = np.linspace(0.22, 0.6, 20)
    estimates = [estimator(beta, x0) for beta in betas]

    print("\nEstimates:")
    for beta, estimate in zip(betas, estimates):
        print(f"Beta: {round(beta, 4)}, Estimator: {round(estimate, 4)}")

    plt.figure(2)
    plt.plot(betas, estimates, marker='o')
    plt.axvline(0.44, linestyle='--')
    plt.title(r'$\hat{v}(\beta)$' + " vs. " r'$\beta$')
    plt.xlabel(r'$\beta$')
    plt.ylabel(r'$\hat{v}(\beta)$')

LATTICE_SIZE = 100
CHAIN_LENGTH = 1000

x0 = np.full((LATTICE_SIZE, LATTICE_SIZE), -1)

plot_X1000(x0)
plot_estimates(x0)

plt.show()
```

References