An Emperical Study of Phase Transitions in the Ising Model

Leon Ma [1]

Abstract

Ising models have traditionally been solved with Markov chain Monte Carlo methods, the most popular of which is the Gibbs' sampler. However, MCMC suffers from correlated samples and long mixing times. In recent years, a new class of algorithms based on variational inference has come to light. These algorithms promise faster convergence abit at the cost of convergence. Belief propagation is an inference algorithm that uses message passing in order to approximate the probability distribution over a factor graph. In this paper, we apply belief propagation to solve the Ising model and show that during belief propagation, the 2D Ising model exhibits a phase transition similar to simulations using MCMC. We compare metrics of phase transitions to MCMC approaches and known theoretical results and conclude that belief propagation can solve the Ising model to a high degree of accuracy and retain many interesting properties that the Ising model is known for.

1. Introduction

The Ising model is a mathematical model of ferromagnestim in statistical mechanics. Invented first by Wilhelm Lenz in 1920 and the equations of the one-dimensional Ising model were solved in by Lenz's student Ernst Ising in 1924. In 1944, Lars Onsager solved the equations for the two-dimensional case. As of today, no analytic solution has been found in dimensions three or higher. Over the last 100 years, many other problems such as percolation, min cut max flow, error correction, neural networks, and neurodegenerative diseases were shown to be very closely mathematically related to the Ising model. The Ising model's relatively simple formulation yet rich and general mathematical properties makes a *drosophila* of physics, mathematics, statistics, and computer science.

Since Onsager's solution, it is known that the Ising model exhibits a phase transition just like magnets in the real world. Above the a special temperature, the critical temperature (or more generally the critical point), the system under goes a rapid change, changing from an ordered magnetic state to a disordered non-magnetic state. This rapid change of order is fundementally related to the relationship between energy, entropy and free energy. When the system is at the critical point, it is said to be at criticality and many interesting properties such as scale invariance and long range correlations are present during the phase transition.

However, due to the difficulty of solving the Ising models in higher dimensions, many approxiomation methods have been devolved over the years. Statistical physicists developed mean field theory and the cavity method to find approximate analytic solutions. In more recent times, the rise of computational power allowed computer algorithms to approximate solutions to the Ising model. The most well known class of algorithms are Markov Chain Monte Carlo (MCMC) algorithms such as Metropolis-Hasting algorithm and Gibbs' sampling. Although these algorithms are powerful and still widely used, they suffer long convergence times and autocorrelated samples.

In this paper, we explore an alternative algorithm to approximately solve the Ising model called belief propagation. Belief propagation works by sending messages from each node of a factor graph to iteratively find the probability distrubition. We show that this method can calculate solutions comparable to Metropolis-Hastings and Gibbs sampling and show that solutions given by belief propegation exhibit a phase translation like predicted theoretically.

2. Background

2.1. Ising Model

The Ising model models ferromagnetic materials as a two-dimensional lattice where each lattice site represents an electron that is either spin up $\sigma_i = +1$ or spin down $\sigma_i = -1$. Each lattice point only interacts with it's adjacent neighbors. In this paper, we impose periodic boundary conditions. This is equivalent to the lattice wrapped around the surface of a torus. This simulates an infinitely large lattice and midigates boundary effects.

Thus, the total energy (also known as the Hamiltonian H in

[1]Department of Statistics, UC Berkeley. Correspondence to: Leon Ma <lma00@berkeley.edu>.

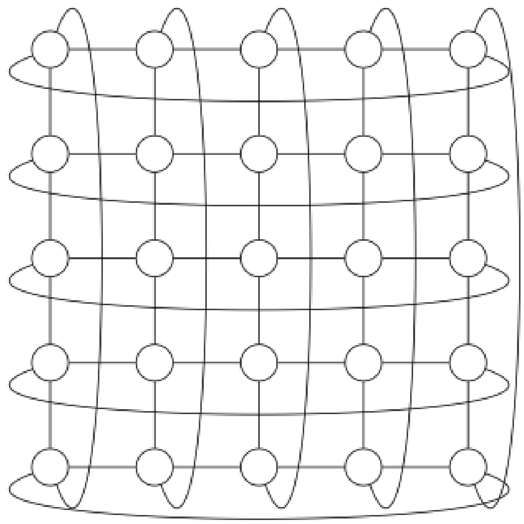

Figure 1. Ising Model with Periodic boundary conditions

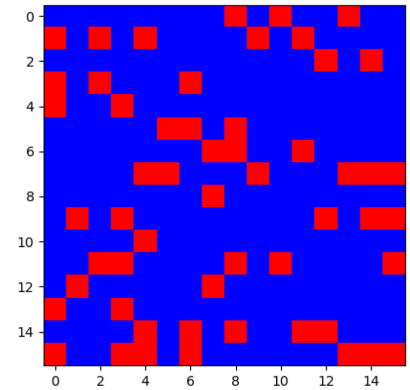

Figure 2. 16x16 Lattice where an electron is spin up with probability 0.75. Blue is spin up and red is spin down.

physics literature) is

$$E(\sigma) = -J \sum_{(i,j)\in\mathcal{N}} \sigma_i \sigma_j - B \sum_j \sigma_j \tag{1}$$

where $(i, j) \in \mathcal{N}$ is understood to be the sum over all pairs of (i, j) that are adjacent neighbors, J is the physical constant knows as the coupling strength and B is the strength of the external magnetic field. We assume $J = 1$ for simplicity and no external magnetic field $B = 0$. Then our total energy is

$$E(\sigma) = -\sum_{(i,j)\in\mathcal{N}} \sigma_i \sigma_j \tag{2}$$

We notice that the energy is lower when σ_i and σ_j are the same sign and greater when σ_i and σ_j are different signs. The energy is minimized when the spins are alligned in the same direction, corresponding to a magnetized state. However, the entropy is maximized when the spins are misalligned corresponding to a demagnetized state.

We can then introduce the partition function,

$$Z = Z(\beta) = \sum e^{-\beta E(\sigma)} \tag{3}$$

The PMF of the system is given by the Boltzmann distribution

$$P(\sigma) = \frac{e^{-\beta E(\sigma)}}{Z} \tag{4}$$

where the inverse temperature $\beta = 1/(kT)$, k is Boltzmann's constant, and the partition function Z is the normalization constant. Given an initial square lattice of σ and inverse temperature β the goal is to find the equalibrium distrubition over σ.

2.2. Phase Transition

In physics a phase transition refers to a change of a system from one state to another. A classical example is ice melting into water. However, more generally it refers to the rapid change of a systems order parameter as a function of one or more of the systems parameters. This more general defintion can be applied to other fields such as rate distortion theory. The Ising model predicts a phase transition of a ferromagnetic system like an iron magnet from a demagnetized state to a magnetized state. It is known emperically that heating a magnet above a certain temperature known as the Curie temperature (770 °C for iron) causes rapid demagnetization. The characteristic feature of the phase transition is that the magnetization strength changes very rapidly over a short range of temperature similar to how ice melts into water over a narrow temperature of range of 0°C. This transition from order to disorder is due to the changing strength of energetic and entropic forces.

In nature, systems tend to minimize a thermodynamic quanity known as free energy. The free energy is given by

$$F = E - TS \tag{5}$$

where S is the thermodynamic entropy. At high temperatures the entropy term is dominant and at low temperature the energetic term is dominant. This explains why the spins are misalligned at high temperature. The thermal energy from the envirnoment is sufficent to flip the lattice site spins. This thermal function acts like a corrupting force constantly distrupting the spins of the system. It is also convient to define the free energy as

$$F = kT \ln Z \tag{6}$$

These two formulas are equivalent

$$\begin{align}
S &= -k \sum_i p_i \ln p_i \tag{7}\\
&= -k \sum_i \frac{e^{-\beta E_i}}{Z} \ln\left(\frac{e^{-\beta E_i}}{Z}\right) \tag{8}\\
&= -k \sum_i \frac{e^{-\beta E_i}}{Z}(-\beta E_i - \ln Z) \tag{9}\\
&= k\beta \sum_i \frac{e^{-\beta E_i}}{Z} E_i + k \sum_i \frac{e^{-\beta E_i}}{Z} \ln Z \tag{10}\\
&= k\beta \langle E \rangle + k \ln Z \tag{11}\\
&= k\beta E + k \ln Z \tag{12}
\end{align}$$

since the energy E is simply the energy averaged all over microstates $\langle E \rangle$. Plugging this into (5) yields

$$\begin{align}
F &= E - T(k\beta E + k \ln Z) \tag{13}\\
&= E - T(k\frac{1}{kT}E + k \ln Z) \tag{14}\\
&= E - E + kT \ln Z \tag{15}\\
&= kT \ln Z \tag{16}
\end{align}$$

The energy can also be defined in terms of the partition function as

$$\begin{align}
E &= -\frac{\partial \ln Z}{\partial \beta} \tag{17}\\
&= -\frac{1}{Z}\frac{Z}{\beta} \tag{18}\\
&= -\frac{1}{Z}\frac{\partial}{\partial \beta}\sum_i e^{-\beta E_i} \tag{19}\\
&= -\frac{1}{Z}\sum_i -E_i e^{-\beta E_i} \tag{20}\\
&= \sum_i E_i \frac{e^{-\beta E_i}}{Z} \tag{21}\\
&= \langle E \rangle \tag{22}\\
&= E \tag{23}
\end{align}$$

These equations will be relevant when we discuss the connection between belief propagation and minimization of the Bethe free energy.

By analyzing at what temperature the energy and entropy terms balance, Onsager solved the 2D Ising model analytitically in the case of no external magnetic field and showed that the critical temperature is

$$T_c = \frac{2J}{k \ln\left(1+\sqrt{2}\right)} \tag{24}$$

or in terms of $\beta_c = \frac{\ln\left(1+\sqrt{2}\right)}{2J}$. This derivation is very technical and will not be covered here.

Systems at the critical point exhibits many interesting properties. The first is the rapid change of the order parameter. In the case of the Ising model, the order parameter is the magnetization given the average value of the spins.

$$\langle m \rangle = \frac{1}{N}\sum_{i=1}^{N} \sigma_i \tag{25}$$

In order to not give preference to major spin down or spin up configurations, it is convient to define the mangetization squared $\langle m^2 \rangle = \frac{1}{N}\sum_{i=1^N} \sigma_i^2$.

At the critical temperature, the system reaches a critical state where order and disorder are perfectly matched. The spins exhibit coordination and thus in order to demagnetized the system, lots of energy must be exerted to break the coordination of the spins. We can then define the specific heat as a measure of how difficult it is to break the coordination of the spins.

$$\begin{align}
C_v &= \frac{\partial E}{\partial T} \tag{26}\\
&= -\frac{\partial E}{\partial \beta}\frac{\partial \beta}{\partial T} \tag{27}\\
&= \frac{\partial^2 \ln Z}{\partial \beta^2}\left(-\frac{1}{k_B T^2}\right) \tag{28}\\
&= k_B \beta^2 (\langle E^2 \rangle - \langle E \rangle^2) \tag{29}
\end{align}$$

where $\langle E^2 \rangle - \langle E \rangle^2 = \mathrm{Var}(E)$. This is analgous to how when ice melts into water, adding more energy to melting ice does not increae it's temperature as all the energy is used to break the molecular bounds between the water molecules in the ice crystal to form liquid water.

This breaking of structure at the critical point is closely related to another phenomenon, long range correlations. The Ising model only takes into account interactions between neigboring lattice sites. Therefore, we expect the correlation between lattice sites to be very short and lattice sites long distances apart to act inpendendently. This is in general true except when the system is at criticality. When order and disorder are perfectly balanced, local interactions have a ripple effect, meaning that local interactions lead to long range communication. The correlation function is given by

$$C_{i,j} = \langle (\sigma_i - \langle \sigma_i \rangle)(\sigma_j - \langle \sigma_j \rangle) \rangle \tag{30}$$

where $\langle \cdot \rangle$ denotes the time average. Then we can define the correlation length function,

$$C(d) = \mathbb{E}\left[C_{i,j} \mid D(i,j) = d\right] \tag{31}$$

which is the correlation strength as a function of the distance between lattice sites.

Another interesting phenomona observed at criticality is scale invariance. Suppose for each 3 by 3 block of spins, we replace that block with the spin corresponding to the majority spin value out of those 9 smaller spins. Apply this procedure to the whole lattice and we have a new lattice with a third of the distance scale. This procedure is called a course graining. We can continue this procedure indefinitely on an infinitely large lattice. At the critical point, all these lattices of varying length scales look the same, in other words there is no characteristic scale. To show this idea graphically, we plot the distribution of the cluster sizes where a cluster is a group of lattice points with the same spin. This gives us a scaling equation

$$\frac{f(kx)}{f(x)} = g(k) \tag{32}$$

where $f(x)$ is the probability of observing a cluster size with size x and k is a scaling factor. Suppose $k = 3$. Then this tells us that the probability of observing a cluster three times larger then x is a function $g(k)$. It turns out that we can show the only function g that satifies this property is $g(k) = k^{-\gamma}$ for some constant γ which means

$$f(x) = Ax^{-\gamma} \tag{33}$$

This is a power law. We can visualize this as a straight line on a log-log plot $\log f(x) = -\gamma \log x + \log A$. The slope of such line on a log-log plot is $-\gamma$.

3. Methods

3.1. Metropolis-Hasting

In the previous sections we have described the mathematical background behind the Ising model and theory of phase transitions. However, solving the Ising model analytitically is extremely difficult and in practice computational methods are used. The most common of these methods are Monte Carlo Markov Chain methods. The simpliest example of MCMC is the Metropolis-Hasting algorithm. Metropolis-Hasting works by building an implict Markov Chain model of the system and sampling from the stationary distrubtion of the chain to generate samples of the system's true probability distrubtion. This works because when the Ising model is at thermodynamic equalibrium, it's Markov chain satifies the detailed balance equations

$$\pi_\mu P(\mu \to \nu) = \pi_\nu P(\nu \to \mu) \tag{34}$$

where μ and ν are states, p_μ is the probability that the system is in state μ and $P(\mu \to \nu)$ is the probability that the system will transition from state μ to state ν. Then we have that

$$\begin{align}
\frac{P(\mu \to \nu)}{P(\nu \to \mu)} &= \frac{\pi_\nu}{\pi_\mu} \tag{35}\\
&= \frac{e^{-\beta E_\nu}/Z}{e^{-\beta E_\mu}/Z} \tag{36}\\
&= e^{-\beta(E_\nu - E\mu)} \tag{37}\\
&= e^{-\beta \Delta E} \tag{38}
\end{align}$$

where the change in energy $\Delta E = E_\nu - E\mu$. Suppose that we start in state μ. At each step of the algorithm, we flip a random spin $\sigma_j \mid j \sim U\{1, 2, \ldots, n\}$ to create a new state ν. If $E_\nu < E_\mu$ then $\Delta E < 0$ and the system would enter the lower energy state so $P(\nu \to \mu) = 1$. This gives us $P(\mu \to \nu) = e^{-\beta \Delta E}$. Note that the probability that the system moves to a higher energy state is non zero due to random fluctuations due to the thermal energy. Furthermore as T increases, β decreases causing $e^{-\beta \Delta E}$ to increase meaning that spin flips are more likely at higher temperature. If $E_\mu < E_\nu$ we can similarly define $P(\mu \to \nu) = 1$ and $P(\nu \to \mu) = e^{-\beta(-\Delta E)}$. This gives us the acceptance probability for $\mu \to \nu$ to be

$$\alpha = \min(1, e^{-\beta \Delta E}) \tag{39}$$

Thus, we only need to calculate the difference in energy $\Delta E = E_\mu - E_\nu$ when determining wether to accept or reject a canidate state. However, as we'll see, this naive method of randomly flipping a spin can lead to pathelogical a Markov chain.

3.2. Gibbs Sampler

The Gibbs sampler begins with $\mathbf{X} = (x_1, \ldots, x_2)$ with joint distribution $p(x_1, \ldots, x_n)$. Denote the ith sample $\mathbf{X}^{(i)} = (x_1^{(i)}, \ldots, x_n^{(i)})$. Then the Gibbs sampler samples

$$x_j^{(i+1)} \sim P(x_j^{(i+1)} \mid x_1^{(i+1)}, \ldots, x_{j-1}^{(i+1)}, x_{j+1}^{(i+1)}, \ldots, x_n^{(i+1)}) \tag{40}$$

In our specific case we choose the index $j \sim U\{1, 2, \ldots, n\}$. $x_j^{(i+1)}$ is given by 40 and $x_{\setminus j}^{(i+1)} = x_{\setminus j}^{(i)}$. The advantage of Gibbs sampling is that in some cases is it simpler ot sample from the conditional distrubtion then it is to marginalize the joint distrubtion.

In the case of our Ising model, at each time step we sample a lattice point j uniformly. Then according to our update rule, the probability of σ_j being spin up is given by

$$P(\sigma_j = +1) = \frac{\exp(-\beta E(\ldots, \sigma_j = +1, \ldots))}{\exp(-\beta E(\ldots, \sigma_j = +1, \ldots)) + \exp(-\beta E(\ldots, \sigma_j = -1, \ldots))} \tag{41}$$

where $\ldots$ denotes the spins of the other lattice points at the previous time step. $P(\sigma_j = -1)$ is given similarly. Notice that this update rule does not depend on the value of σ_j at the previous time step. This update rule is referred to as the heat bath update rule because the orientation of each spin depends only on the energy transferred from the thermal energy of the heat bath.

Now lets try to reformulate this update rule to better match our Metropolis-Hasting update rule. Suppose that initially $\sigma_j = -1$. Let the event that $\sigma_j = +1$ at the next time step be denoted $\mu \to \nu$ as before. Then

$$P(\mu \to \nu) = P(\sigma_j = +1) = \tag{42}$$

$$\frac{\exp(-\beta E(\ldots, \sigma_j = +1, \ldots))}{\exp(-\beta E(\ldots, \sigma_j = +1, \ldots)) + \exp(-\beta E(\ldots, \sigma_j = -1, \ldots))} \tag{43}$$

$$= \frac{\exp(-\beta E_\nu)}{\exp(-\beta E_\nu) + \exp(-\beta E_\mu)} \tag{44}$$

$$= \frac{\exp(-\beta(E_\mu + (E_\nu - E\mu)))}{\exp(-\beta(E_\mu + (E_\nu - E\mu))) + \exp(-\beta E_\mu)} \tag{45}$$

$$= \frac{\exp(-\beta(E_\mu + \Delta E))}{\exp(-\beta(E_\mu + \Delta E)) + \exp(-\beta E_\mu)} \tag{46}$$

$$= \frac{\exp(-\beta E_\mu)\exp(-\beta \Delta E)}{\exp(-\beta E_\mu)\exp(-\beta \Delta E) + \exp(-\beta E_\mu)} \tag{47}$$

$$= \frac{\exp(-\beta \Delta E)}{\exp(-\beta \Delta E) + 1} \tag{48}$$

$$= \frac{1}{1 + \frac{1}{\exp(-\beta \Delta E)}} \tag{49}$$

Following the same derivation for the probability that σ_j is initially at spin up then is flipped spin down gives the same expression. Thus, the probability of the spin flip under the heat bath update rule is

$$\alpha = \left(1 + \frac{1}{\exp(-\beta \Delta E)}\right)^{-1} \tag{50}$$

We can just replace our Metropolis-Hasting algorithm with this new acceptance probability showing that the Gibbs sampler is special case of the Metropolis-Hasting algorithm. This class of update rules that depend soley on the Boltzmann distrubtion is known as Glauber dynamics. We shall see in practice this update rule generally results in faster convergence.

4. Experiments

4.1. MCMC Convergence

First we will compare the convergence of the Metropolis-Hasting sampler to the Gibbs sampler.

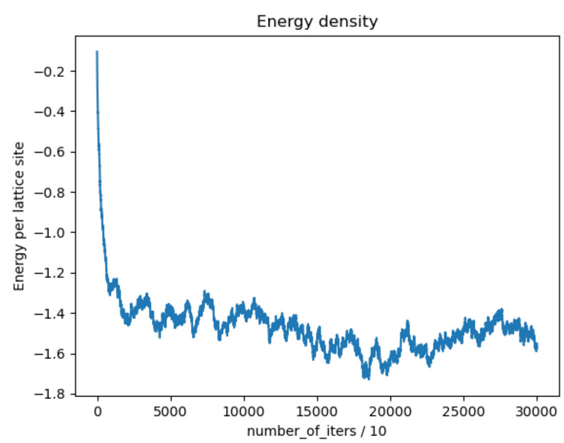

Figure 3. Convergence test of Metropolis-Hasting algorithm

4.2. Order Parameters

Now we study how the energy and the magnetization change with β and the lattice size L. Both the energy and magnetization tells us how ordered our lattice is. Recall that when our lattice has has most of the spins pointing in the same direction, these aligned spins create favorable energetic interactions leading to the lattice being in a lower energy state.

We notice that as β decreases, T increases and thus the thermal energy distrupts the orientation of the spins, causing the energy of the system to increase as β decreases. This is more appearent when we observe the plot of the magnetization.

We see that the magnetization looks like $\langle m \rangle \sim \operatorname{arcsinh}(\beta)$ as predicted by theory. Furthermore, the critical β seems to

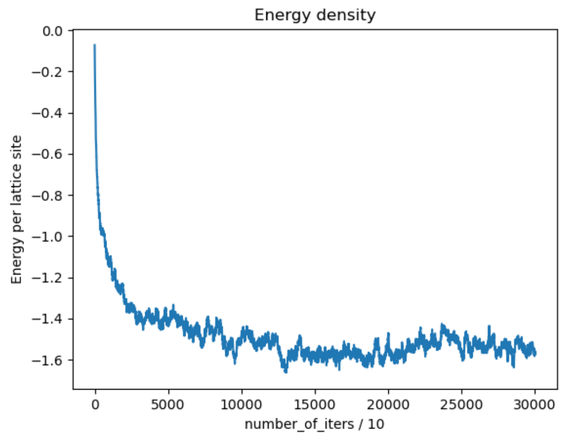

Figure 4. Convergence test of Gibbs algorithm

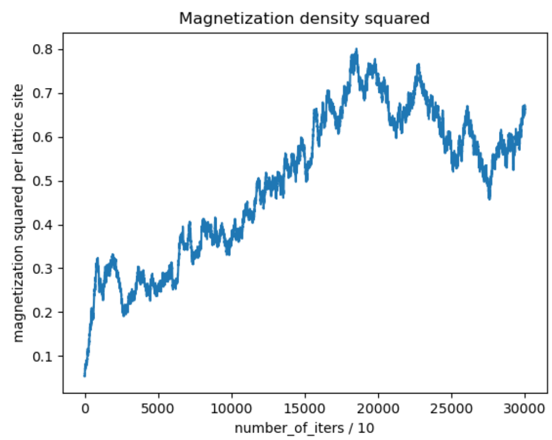

Figure 5. Convergence test of Metropolis-Hasting algorithm

be around 0.44. Comparing this experimental observation to Onsager's solution given in (24) gives

$$\beta_c = \frac{\ln(1+\sqrt{2})}{2} \approx 0.4407 \tag{51}$$

which is in very close agreement.

Lastly we also notice that as L increases, the phase transition becomes sharper. This suggests that phase transitions are an emergent phenomenon. In reality, the number of electrons is orders of magnetudes larger then this simulation giving the extremely sharp phase transitions we observe in nature. All these observations hold true for the specific heat.

The peak at $\beta = 0.44$ means that at that point, the most thermal energy is spent breaking the favorable energetic interactions of the electrons.

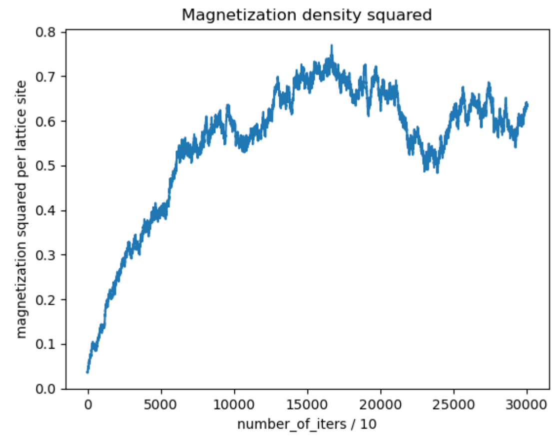

Figure 6. Convergence test of Gibbs algorithm

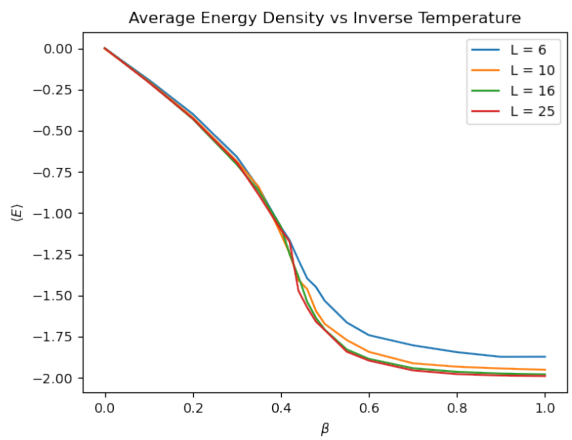

Figure 7. Average Energy Density vs Inverse Temperature

4.3. Critical Phenomena

Now we turn to special properties of the Ising model at criticality. From here on out, we use $L = 16$, $p = 0.5$ and the Gibbs sampler as before. As we outlined in section 2.2, systems at criticality exhibit long range correlations. We plot the correlation strength as a function of the distance d between two lattice points.

We see that at the critical point $\beta = 0.44$, the system exhibit much stronger correlations for large d values compared to other values at β. This is further evidence that the Ising model has a critical point at $\beta = 0.44$. When β is large, the spins are aligned together. This leads to high short range correlation but because the spins rarely fluctuate, they are stuck at their average value making the $\sigma - \langle\sigma\rangle$ small. At low β the spins fluctuate chaotically leading to little cor-

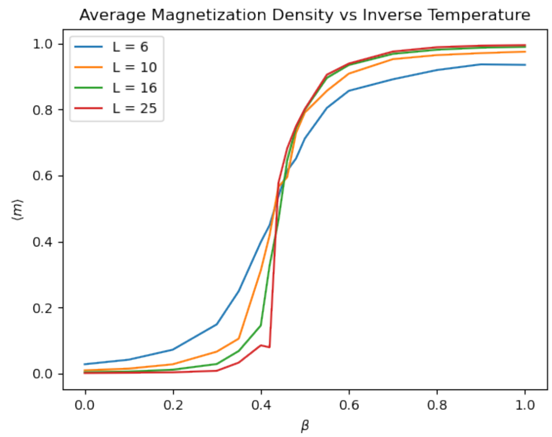

Figure 8. Average Magnetization Density vs Inverse Temperature

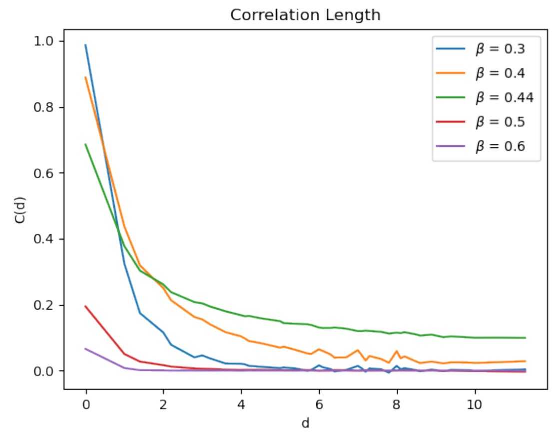

Figure 10. Correlation Length

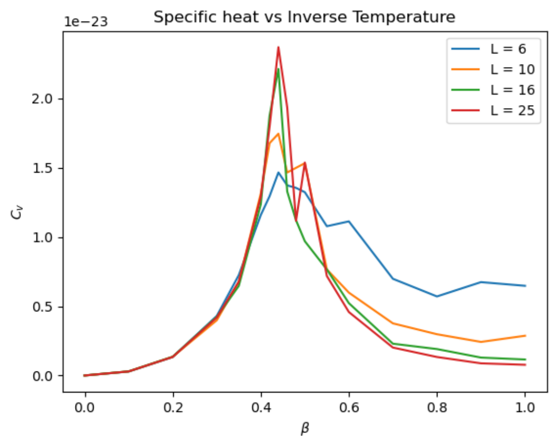

Figure 9. Specific Heat vs Inverse Temperature

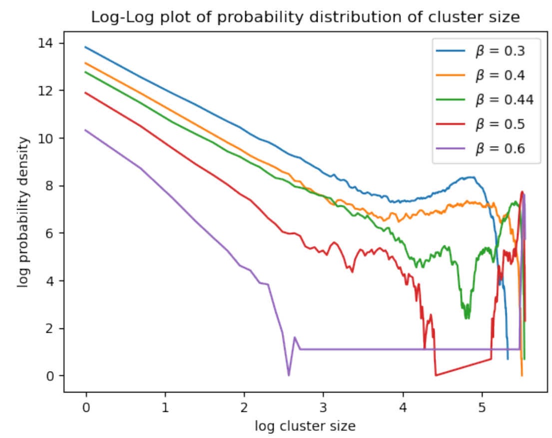

Figure 11. Distribution of cluster size

relation between spins. This emergent behavior of local interactions leading to long range communication has applications in neuroscience where local interactions between nearby neurons leads to long range communication across distance areas of the brain.

The other property of criticality that we outlined in section 2.2 is scale invariance. Using DFS to count the size of each cluster, we plotted the distrubition of cluster sizes on the log-log plot. If the clusters have no characteristic scale, the distrubition of the cluster sizes follow a power law. On a log-log plot, power laws appear as straight lines. Plotting the distribution of cluster sizes for different values of β reveals that $\beta = 0.44$ gives a curve that resembles a straight line the most.

Although all of the curves start off straight but then get more chaotic as the cluster size increases, for $\beta = 0.44$ the linear relationship holds for the longest, up to clusters of size $\log 4$. We notice that as β decreases, the frequency of small clusters increases. The chaotic behavior for large cluster sizes is likely due to the fact that the lattice is not infinite and thus large clusters are overrepresented as they can loop around the lattice and form super large clusters. Furthermore, larger clusters are more rare making estimates for their frequency more noisy.

At the critical point, many features of the Ising model follow power laws, not just the cluster size. Another example is the correlation length.

As we can see at $\beta = 0.44$, the correlation length on a log-log plot is a straight line, meaning it follows a power law. Power laws are essential to understanding critical phe-

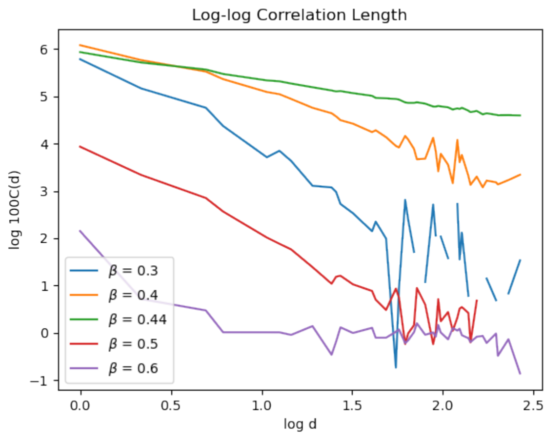

Figure 12. Log-log plot of correlation length

nomenon.

5. Discussion

6. Conclusion