The randomly dilute Curie-Weiss model with Gaussian, bounded and sub-Gaussian couplings

by

Luka van der Heiden

Luka van der Heiden 5176514

Author: Luka van der Heiden
Supervisor: Elena Pulvirenti, Jos Thijssen
Project Duration: March 2022 - November, 2022
Faculty: Faculty of Electrical engineering mathematics and computersience, Delft
Faculty of applied science, Delft

TU Delft

Summary

The behaviour of the dilute Curie-Weiss model has been analysed for Gaussian, bounded and sub-Gaussian couplings. This model is an abstraction of the Curie-Weiss model which is a model of a spin configuration where instead of a constant coupling between spins the spin coupling is a realisation of a random variable. To analyse this behaviour first the behaviour of the standard Curie-Weiss model is depicted. A notable part from this behaviour is the presence of a phase transition., indicating that this model can be used for a study of phase transitions. After having analysed the behaviour of the standard Curie-Weiss model theorems, stating the closeness between the standard Curie-Weiss model and the randomly dilute Curie-Weiss model for Gaussian, bounded and sub-Gaussian couplings, will be proven. Indicating that the models behave approximately the same way. These theorems will be in the form of a bound for the on randomly dilute Curie-Weiss model as an exponential multiplied by the standard Curie-Weiss model with the probability of this bound being a sub-Gaussian distribution.

Contents

1 Introduction

An important branch of physics is the statistical mechanics. In statistical mechanics the behaviour of physical systems is studied not by looking at the system itself but by looking at the combined behaviour of all the individual particles that make up the system. So instead of looking at a clump of gas and describing its pressure and volume, we look at it via all the individual molecules in the gas. This branch of physics has produced many important results. Most notably it goes hand in hand with solid state physics this gave rise to very important theories that are at the basis of modern computers.

In this paper we look at a model that is used to model magnets from the viewpoint of statistical mechanics, i.e. the Curie-Weiss model. This model is a first step to understanding how magnets behave at high and low temperatures.

The Curie-Weiss model is constructed from the spin state of all the individual atoms in a magnet. In this model there are two interactions that are taken into account. The magnetic interaction between different atoms and the interaction between atoms and an external magnetic field. The behaviour of the system will be determined from these two interactions and known laws from statistical mechanics.

An important part that will be noticed from this behaviour is the presence of a phase transition. A phase transition is an transition in our model from one physical phase to another. In the Curie-Weiss model we will see that this transition is from paramagnetism to ferromagnetism. However in other systems a phase transition may be from a liquid to a solid state or some other phase transition. As the Curie-Weiss model is a basic model in which a phase transition is present. Studying this model could help us understand other phase transitions better.

An important fault in the Curie-Weiss model is that it has some unjustifiable assumptions. One of these assumptions is that in the Curie-Weiss model all the interactions between different atoms have the same strength. However since most atoms have different distances between each other it is not a well grounded assumption. So in this project we will look at an extension of the Curie-Weiss model in which an abstraction of this assumption is considered. The abstraction that we will consider is the randomly dilute Curie-Weiss model. In this model we will look at a system similar to the Curie-Weiss model, where we can set the interaction strengths between different atoms. By setting this interaction strengths one can for example take this distance into account and greatly improve the accuracy of the model. However setting all the interaction values to a specific value makes the model very difficult to solve. Thus we will consider the case where the interactions are random variables. In this paper we prove a result which establishes that the randomly dilute Curie-Weiss model behaves like the corresponding system where the random couplings are replaced with their averages.

By obtaining this result we can determine if the presence of the phase transition. Remains for the more general Curie-Weiss model. And thereby we get to know how good of an approximation the Curie-Weiss model is for the actual physical system. And this result could help understand other systems that contain a phase transition.

To compare these two models we will first take a closer look at the behaviour of the Curie-Weiss model. After having seen how the Curie-Weiss model behaves we will prove the closeness of the Randomly Dilute Curie-Weiss model to the standard Curie-Weiss model.

After having proven the closeness of the dilute to the non dilute model for Gaussian random variables we will prove in a similar way the closeness for bounded and sub-Gaussian random variables.

With these results we can show that the presence of a phase transition is not that of a special occurrence and can be expected in more models that somewhat resemble the Curie-Weiss model

2

Some basic concepts of Statistical Mechanics

In this report we study a model that arose from statistical physics. In this chapter a short discussion of the basic concepts of statistical mechanics will be included, together with an important result form statistical mechanics from which the Curie-Weiss model will be constructed.This discussion will be based on [5].

Statistical mechanics is a branch of physics where statistical properties are used to derive properties of matter. These properties are mostly thermodynamical properties such as temperature and entropy. In statistical mechanics these macroscopic properties are derived from a microscopic state, to which a probability is assigned. By calculating macroscopic properties from these microscopic states we can link them, thereby creating a probability measure for macroscopic properties.

From this probability measure averages can be calculated as approximations of the real value.

Now to use this framework of statistical mechanics we need to have a way to give a probability value to each micro state. The probabilistic ensemble we will use for the Curie-Weiss model will be the Canonical ensemble, which will be explained in the next section.

To explain the Canonical ensemble we will have to use two definitions from statistical mechanics. Firstly we will define the thermodynamical entropy, as follows

$$S = k \log(\Omega). \tag{2.1}$$

Here S is the entropy, k is the Boltzmann constant and Ω is the amount of micro states with which the overall system behaves the same thermodynamically. Thus the entropy can be computed by counting the amount of possible micro states. This counting is a technique that is often used in statistical mechanics and we will also use it later in this report. This entropy satisfies the following thermodynamic identity.

$$dS = \frac{1}{T}(dU + PdV - \mu dN), \tag{2.2}$$

where we describe the change in entropy as a function of T temperature, dU change in energy, P pressure, dV change in volume,μ chemical potential and dN change in number of particles in the system.

2.1. The canonical ensemble

The Canonical ensemble is a statistical ensemble where we assume that the system is in equilibrium with a large reservoir. This reservoir will be held at a fixed temperature and the reservoir and the system will be confined in a fixed volume. Also both the reservoir and our system will have a fixed number of particles.

To define a probability measure we will see that the probability of finding the system in a state is proportional to the amount of micro states that the reservoir can have if the system is in that state, or in formulas

$$\frac{P(state1)}{P(state2)} = \frac{\Omega_R(state1)}{\Omega_R(state2)} \tag{2.3}$$

where $\Omega_R(statei)$ is the amount of possible states in the reservoir corresponding to state i in the system, for $i \in \{1, 2\}$

Using the definition of entropy we can obtain a proportion related to entropy as follows

$$\frac{P(state1)}{P(state2)} = \frac{e^{S_R(state1)/k}}{e^{S_R(state2)/k}} = e^{(S_R(state1)-S_R(state2))/k}. \tag{2.4}$$

Now for the canonical ensemble we assume that the amount of particles in the model that we study is fixed and that the change in volume is negligible compared to the change in energy. Under these assumptions we can integrate the change in entropy to obtain the following

$$S_R(state1) - S_R(state2) = \frac{U_R(state1) - U_R(state2)}{T}. \tag{2.5}$$

By noticing that the system and the reservoir are only in contact with each other and no other parts, we can conclude that the difference of the energy of the reservoir and the difference of the energy of the system must compensate each other to comply to the conservation of energy. After some rewriting we then obtain

$$\frac{P(state1)}{e^{U(state1)/kT}} = \frac{P(state2)}{e^{U(state2)/kT}}. \tag{2.6}$$

From this equation the distribution can be obtained, by observing that both sides must be a constant. From this we have the distribution up to a constant, so that by normalising this distribution we obtain the final distribution

$$P(state) = \frac{1}{Z} e^{E(state)\beta}, \tag{2.7}$$

where Z is the normalisation constant and β is a conventional way to write $\frac{1}{kT}$. Most importantly E is the energy of our system in this particular state. This energy is sometimes called the Hamiltonian of the system. Now that we have obtained a distribution for an abstract thermodynamic system we can use this for the model we want to study.

3

The Curie-Weiss model

3.1. The spin model

In this chapter we will explain and discuss the behaviour of the standard Curie-Weiss model. The Curie-Weiss model is a model from statistical physics used to describe magnets. This explanation will be detailed such that later generalizations of the model will be understood too.

In the Curie-Weiss model to model the magnet each individual atom of the magnet is considered as a small magnetic dipole with magnetic dipole moment σ_i. As a simplification we will assume that the magnetic moment of each of these magnets is restricted to two options. Namely pointing up and pointing down. This assumption can be justified by noticing that by choosing our frame of reference these are the stable states for each atom. Now, by choosing the right units of measure we have that for each atom its dipole moment is either $+1$ or -1, and therefore

$$\forall i, \sigma_i \in \{-1, +1\}.$$

With these assumptions we can fully describe the state of our model as a spin configuration. A spin configuration of N atoms is given by $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, .., \sigma_N)^T$.

By taking into account the interactions of these atoms we can introduce an Hamiltonian. By using this Hamiltonian we can use the canonical ensemble to obtain a probability measure for any configuration of the system

Each atom has 2 types of interactions a magnetic interaction with all the other atoms. With strength $\frac{1}{N}$ and an interaction with a constant external magnetic fields. From this the following Hamiltonian can be introduced,

$$H_N^{CW}(\boldsymbol{\sigma}) = -\frac{1}{N} \sum_{1 \leq i < j \leq N} \sigma_i \sigma_j - h \sum_{i=1}^{N} \sigma_i, \tag{3.1}$$

here h is the external magnetic field. From this Hamiltonian the following Gibbs measure can be obtained for any $\boldsymbol{\sigma} \in S_N$ where $S_N = \{-1, +1\}^N$ is the set in which the spin configurations are defined

$$\mu_{\beta,N}^{CW}(\boldsymbol{\sigma}) = \frac{e^{-\beta H_N^{CW}(\boldsymbol{\sigma})}}{Z_{\beta,N}^{CW}}. \tag{3.2}$$

Here $Z_{\beta,N}^{CW}$ is the normalization constant, sometimes referred to as the partition function, and it implies that

$$\sum_{\boldsymbol{\sigma} \in S_N} \mu_{\beta,N}^{CW}(\boldsymbol{\sigma}) = 1 \tag{3.3}$$

3.2. Mean magnetisation

In the above section the standard Curie-Weiss was constructed. An important part of this construction is the assumption that all interactions between two atoms have interaction strength $\frac{1}{N}$. This assumption makes the Curie-Weiss model a relatively easy model to solve. However by making this assumption we disregard the distance between two atoms . This makes the assumption not physically realistic. One can also see this model as being of N atoms on a fully connected graph depicted in Picture 3.1.

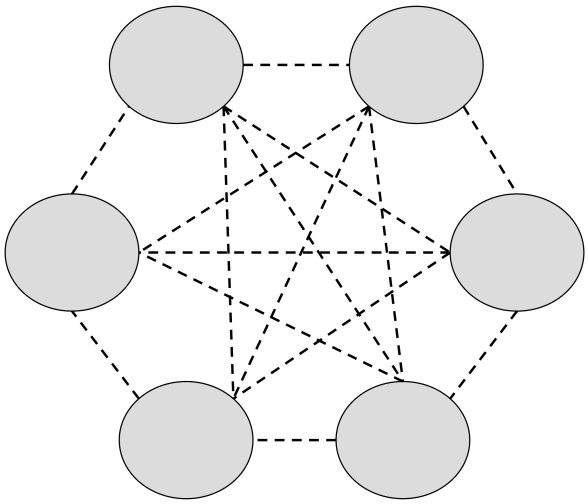

Figure 3.1: A fully connected graph as which the Curie-Weiss model can be seen since all interaction strengths are equal. Source:[6]

As seen in the previous section with this assumption we have the following Gibbs measure for the Curie-Weiss model

$$\mu^{CW}_{\beta,N}(\boldsymbol{\sigma}) = \frac{1}{Z^{CW}_{\beta,N}} \exp\Big(-\beta(-\frac{1}{N}\sum_{1\leq i<j\leq N}\sigma_i\sigma_j - h\sum_{i=1}^{N}\sigma_i)\Big). \tag{3.4}$$

To see how this model behaves we will not look at all the spin configurations but at the magnetisation defined as the empirical average of all the spins

$$m_N(\boldsymbol{\sigma}) = \frac{1}{N}\sum_{i=1}^{N}\sigma_i. \tag{3.5}$$

The magnetisation belongs to the set $\Gamma_N = \{-1, -1+\frac{2}{N}, ..., 1-\frac{2}{N}, 1\}$. Indeed. if all the spins are pointing down $m = -1$, and a new possible element of Γ_N can be reached by flipping one spin from down to up, thereby adding $\frac{2}{N}$ to the magnetisation. Until all spins are pointing up and have a magnetisation of 1.

From this the Hamiltonian of the Curie-Weiss model can be greatly simplified. Because we no longer have to look at the entire spin configuration but only at the magnetisation. And the magnetisation lives on a much simpler set. Thus we have

$$H^{CW}_N(\boldsymbol{\sigma}) = -N\big(\frac{1}{2}m(\boldsymbol{\sigma})^2 + hm(\boldsymbol{\sigma})\big). \tag{3.6}$$

Now the probability distribution according to m can be written as

$$Q^{CW}_{\beta,N}(M) = \mu^{CW}_{\beta,N}(m\in M) = \sum_{m\in M}\sum_{\boldsymbol{\sigma}:m(\boldsymbol{\sigma})=m}\frac{e^{-\beta H_N(\boldsymbol{\sigma})}}{Z^{CW}_{\beta,N}}. \tag{3.7}$$

Now if M has only 1 element with a combinatorial argument one can write the following mesoscopic measure

$$Q^{CW}_{\beta,N}(m) = \binom{N}{\frac{1+m}{2}N} \frac{e^{\beta N(\frac{1}{2}m^2+hm)}}{Z^{CW}_{\beta,N}}. \tag{3.8}$$

Here it is used that by having a magnetisation of m an amount of $\frac{1+m}{2}N$ spins must be pointing up. This is because m can be written as,

$$m = \frac{\#\uparrow - \#\downarrow}{N} = \frac{\#\uparrow - (N - \#\uparrow)}{N}, \tag{3.9}$$

where # $\uparrow$ is the amount of spins pointing up and # $\downarrow$ is the amount of spins pointing down. By rearranging we indeed get that the amount of spins pointing up is given by $\frac{1+m}{2}N$. Therefore there are $\binom{N}{\frac{1+m}{2}N}$ spin configurations that have magnetisation m.

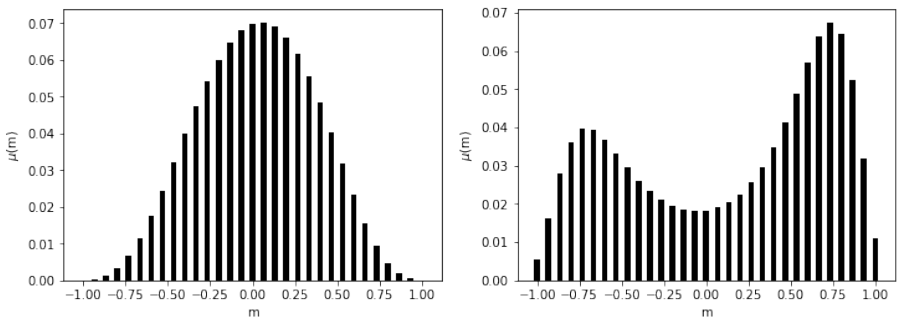

Figure 3.2: plot of the distribution of the Curie-Weiss model for 30 particles with an external field of 0.01. Left $\beta = 0.8$ and right $\beta = 1.2$. one can clearly see the existence of 1 local maximum on the left and two local maxima on the right. Indicating the phase transition at $\beta = 1$. These plots are obtained by numerically calculating equation (3.7).

Now we have obtained a formula for the distribution of the Curie-Weiss model depending on the magnetisation, thereby reducing the degrees of freedom. We can see how it behaves. For this we will have a heuristic argument in this section and in the next section we will look at this distribution as N goes to infinity, since magnets consist of many atoms.

In Figure 3.2 the distribution is plotted for 30 atoms with an external field of h=0.01 and 2 different values of β namely $\beta = 0.8$ and $\beta = 1.2$. One can clearly observe that for the smaller beta only one local maximum is present while for the bigger beta 2 local maxima are present. This difference between the distributions points at a phase transition occurring at $\beta = 1$ The specifics of this behaviour will be discussed in Section 3.3, while an interpretation of this transition will be given in Section 3.4

3.3. Phase transition

In the previous section we introduced the distribution of the magnetisation of the Curie-Weiss model for an arbitrary amount of particles. However if we take the limit as N goes to infinity the behaviour of this model becomes more obvious.

According to Stirling's approximation (see Appendix A.1), we can write the distribution of (3.7) as an exponential function, such that no summation terms are present.

$$Q^{CW}_{\beta,N}(m) = \frac{e^{-N\,j(m)}}{Z^{CW}_{\beta,N}}. \tag{3.10}$$

Here, according to Stirling's approximation, j which we from now on we will call the free energy of the Curie-Weiss model, is defined as

$$j(m) = -\frac{\beta m^2}{2} - \beta hm + \frac{1-m}{2}\log\frac{1-m}{2} + \frac{1+m}{2}\log\frac{1+m}{2}. \tag{3.11}$$

Now as we take the limit of N to infinity, the distribution will concentrate around the global minima of the free energy. This happens because the probability of any other magnetisation decreases exponentially compared to the probability of the global minima, when we take the limit of N to infinity.

In order to find the places of these minima we will have to differentiate the free energy and set it equal to zero. After some calculations which we will reserve for the Appendix (see A.2) it is obtained that the places where the derivative of the free energy is zero are the solutions of the following equation,

$$\tanh(\beta m + \beta h) = m. \tag{3.12}$$

Depending on the value of beta this equation will have 1 or 3 solutions .

For $\beta \leq 1$ there will be one solution of the above equation corresponding with 1 local and also a global minimum of the free energy. For $\beta > 1$ there will be 2 solutions corresponding with 2 local minima and 1 local maximum.

Now from equation (3.10) the distribution as $N \to \infty$ will concentrate around the global minima of the free energy. From the four graphs in figure 3.3 we can understand how the global minima behave for four situations.

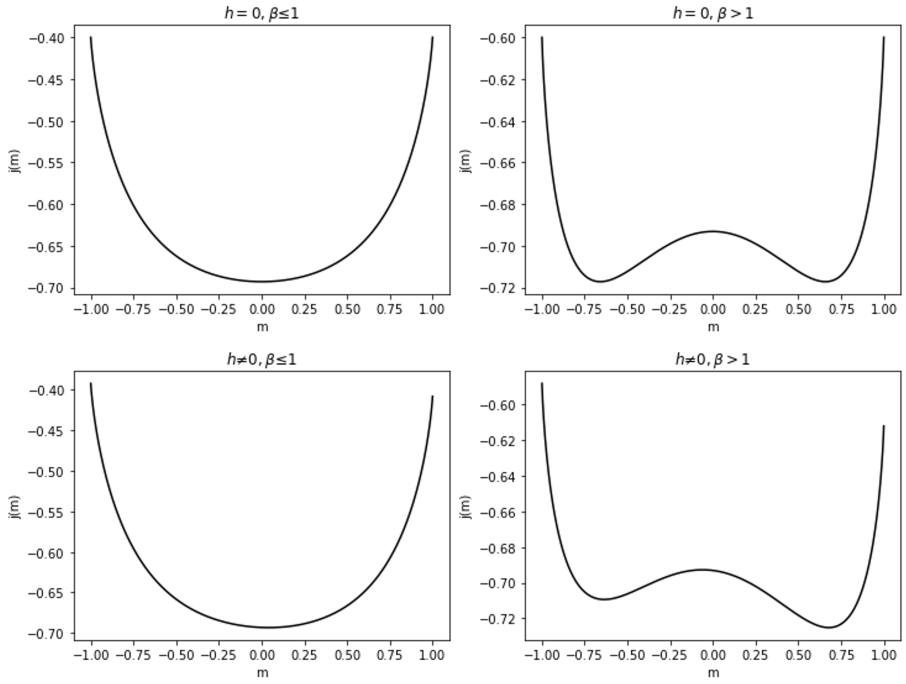

Figure 3.3: Plots of the free energy of the magnetisation.Here four cases are plotted. Top left $h = 0, \beta \leq 1$, top right $h = 0, \beta > 1$, bottom left $h \neq 0, \beta \leq 1$ and bottom right left $h \neq 0, \beta > 1$

We can see that for three of the four situations there is only one global minimum. Here in the limit N to infinity the distribution is a delta distribution at this minimum.

For the situation $h = 0$ and $\beta > 1$ since the two local minima have by symmetry the same free energy, the distribution will be two delta peaks at these two minima.

The clear distinction between these behaviours depicts different phases. The transition of one to another phase is called a phase transition. The presence of this phase transition in the Curie-Weiss model makes the Curie-Weiss model an interesting model to study.

3.4. Physical interpretation

Now that a mathematical depiction of the Curie-Weiss model is described a physical interpretation of the result can be given. As a piece of metal has many atoms this can be considered as the Curie-Weiss model in the limit of large N. Now since $\beta = \frac{1}{Tk_b}$ we can see that for large Temperature the metal does not have it's own magnetisation, but does react to a external magnetic field. So in this case we have paramagnetic behaviour.

This happens because at large temperature the direction of the spins of each atom oscillates very quick. Therefore all the interactions between the atoms can be neglected. However some will line up with an external magnetic field therefore shifting the minimum a bit.

At lower temperature these oscillations become less energetic and the interactions become more important making it that most of the spins line up and showing ferromagnetic behaviour.

The presence of these two phases also indicates a phase transition. This physically manifests as the existence of a temperature where below this temperature we have ferromagnetic behaviour and above this temperature paramagnetic behaviour. The transition between these phases should behave somewhat the same as the standard phase transitions we know of, such as liquid-vapour phase transition. Since the Curie-Weiss model is mathematically a treatable model and still shows a phase transition, the study of the Curie-Weiss model could teach us things about other phase transitions.

4

The dilute Curie-Weiss model with Gaussian random variables

The randomly dilute Curie-Weiss model is a generalisation of the Curie-Weiss model described in the previous section. In this section we will first introduce the randomly dilute Curie-Weiss model and after that we will prove a theorem stating the closeness of the randomly dilute Curie-Weiss model to the standard Curie-Weiss.The proof of this theorem will generally follow the same strategy used to proof the same closeness for Bernoulli random variables in "Bovier-Marello-Pulvirenti" [2]. An extra addition in this report compared to "Bovier-Marello-Pulvirenti" is that in this report all constants will be calculated exactly.

4.1. The dilute model

The randomly dilute Curie-Weiss model is a modification of the Curie-Weiss model, where instead of setting the interaction strengths J_{ij} equal to 1 we will consider them as random variables. In our case we will look at i.d.d. Gaussian random variables.

Let again $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_N)^T \in S_N$ be the spin configuration. Let all the interaction strengths $(J_{ij})_{1 \leq i < j \leq N}$ in the triangular array of interaction strengths be such that $J_{ij} \sim \mathcal{N}(A, B^2)$ and all of them are independent. The Hamiltonian of the randomly dilute Curie-Weiss model can be defined as

$$H_N^{RDCW}(\boldsymbol{\sigma}) = -\frac{1}{AN} \sum_{1 \leq i < j \leq N} J_{ij}\sigma_i\sigma_j - h\sum_{i=1}^{N} \sigma_i, \tag{4.1}$$

and its mean is therefore

$$\begin{aligned} \mathbb{E}_J[H_N^{RDCW}(\boldsymbol{\sigma})] &= -\frac{1}{AN} \sum_{1 \leq i < j \leq N} \mathbb{E}_J[J_{ij}]\sigma_i\sigma_j - h\sum_{i=1}^{N} \sigma_i \\ &= -\frac{1}{AN} \sum_{1 \leq i < j \leq N} A\sigma_i\sigma_j - h\sum_{i=1}^{N} \sigma_i = H_N^{CW}(\boldsymbol{\sigma}) \end{aligned} \tag{4.2}$$

where $\mathbb{E}_J$ is used to represent the average due to the random variables J_{ij}. From this we can split the Hamiltonian into the known Curie-Weiss model part and a random part.

$$H_N^{RDCW}(\boldsymbol{\sigma}) = \mathbb{E}_J[H_N^{RDCW}(\boldsymbol{\sigma})] + \Delta_{N,A}(\boldsymbol{\sigma}) \tag{4.3}$$

where setting $\hat{J}_{ij} = J_{ij} - A$ implies that

$$\Delta_{N,A}(\boldsymbol{\sigma}) = -\frac{1}{AN} \sum_{1 \leq i < j \leq N} \hat{J}_{ij}\sigma_i\sigma_j \tag{4.4}$$

From this a distribution for the randomly dilute Curie-Weiss model can be introduced in the same way as it was introduced in the previous chapter,

$$Q^{RDCW}(m) = \frac{e^{-\beta \tilde{H}_N(m)}}{Z^{RDCW}} \Big(\sum_{\substack{\boldsymbol{\sigma} \in S_N \\ m(\boldsymbol{\sigma})=m}} e^{-\beta \Delta(\sigma)} \Big) \tag{4.5}$$

Where Z^{RDCW} is a normalisation constant such that.

$$\sum_{m \in \Gamma_N} Q^{RDCW}(m) = 1 \tag{4.6}$$

This is the mesoscopic measure of the randomly dilute Curie-Weiss model, which has now been split into a deterministic part that behaves like the Curie-Weiss model, and a random part. In the following parts of this chapter we will proof a theorem stating the closeness of the randomly dilute Curie-Weiss to the Curie-Weiss model, thereby proving that they behave in similar ways.

4.2. The main result

Theorem 3.1. For any $s > 0$,$\beta > 0$,$h \geq 0$ any function $g(m) : \Gamma_N \to [0, \infty)$ and positive constants k_1, k_2. Then as N goes asymptotically to infinity the following bounds hold.

$$\mathbb{P}_J \Bigg(\sum_{m \in \Gamma_N} g(m) Q^{RDCW}(m) \leq e^{\alpha + s} \frac{Z^{CW}}{Z^{RDCW}} \Big(\sum_{m \in \Gamma_N} g(m) Q^{CW}(m) \Big) (1 + \mathsf{o}(1)) \Bigg) \geq 1 - k_1 e^{-k_2 s^2} \tag{4.7}$$

$$\mathbb{P}_J \Bigg(\sum_{m \in \Gamma_N} g(m) Q^{RDCW}(m) \geq e^{\kappa - s} \frac{Z^{CW}}{Z^{RDCW}} \Big(\sum_{m \in \Gamma_N} g(m) Q^{CW}(m) \Big) (1 + \mathsf{o}(1)) \Bigg) \geq 1 - k_1 e^{-k_2 s^2} \tag{4.8}$$

Where α, κ, Z^{RDCW} and Z^{CW} are defined in respectively (4.24),(4.40), (4.6) and (3.3) thus

$$\begin{aligned} \alpha =& \frac{\beta^2}{4A^2} B^2 \\ \kappa =& \max_{\eta \in (0,1)} \big(\alpha + \log(\eta) - 2 \frac{2\beta\Delta}{A} \sqrt{\alpha + \log(2) - \log(1-\eta)} \big). \end{aligned} \tag{4.9}$$

In our result the function $g(m)$ is an arbitrary function that can be used to determine what information we want from the eventual distribution Q^{RDCW}.

So now our main theorem tells that if we want to know any information about the randomly dilute Curie-Weiss we can compare it to the Curie-Weiss with a correction term of $\frac{e^{\alpha} Z^{CW}_{N,\beta}(1+o(1))}{Z^{RDCW}_{N,\beta}}$ which is a lower bound. If we increase this lower bound with a factor e^s for an arbitrary positive s the probability that our bound is correct increases with $k_1 e^{k_2 * s^2}$ where k_1 and k_2 are positive constants.

The same reasoning as above can be said for the lower bound with factors $\frac{e^{\kappa} Z^{CW}_{N,\beta}(1+o(1)}{Z^{RDCW}_{N,\beta}}$ and e^{-s}

Thus any information about the randomly dilute Curie-Weiss model can be bounded by the same information for the standard Curie-Weiss model together with a correction term. By increasing this bound exponentially in variable s the probability that the bound is correct behaves as a sub-Gaussian distribution in s.

4.3. Concentration inequality

To prove the theorem we will first introduce the following notation.

$$\mathcal{Z} = \sum_{m\in\Gamma_N} g(m) \sum_{\substack{\boldsymbol{\sigma}\in S_N \\ m(\boldsymbol{\sigma})=m}} e^{-\beta\Delta(\boldsymbol{\sigma})} \tag{4.10}$$

$$F = \log(\mathcal{Z}) \tag{4.11}$$

Where here $g : \Gamma_N \to [0,\infty)$ is a function with which we can choose what information we want from the distribution.

With this we can split $\mathcal{Z}$ into two parts, a deterministic part and and a random part that will take values close to 1,

$$\mathcal{Z} = e^{\mathbb{E}(F)} e^{F-\mathbb{E}(F)}. \tag{4.12}$$

Now that the randomly dilute Curie-Weiss model has been split in a random and non-random part the random part can be bounded by a concentration inequality.In order to do this from [1] we will recall the following theorem [5.6].

Theorem 3.2 Let $X = (X_1, ..., X_n)$ be a vector of n independent standard normal random variables. Let $f : \mathbb{R} \to \mathbb{R}^n$ be an L-Lipschitz function. Then for all $t > 0$.

$$\mathbb{P}\big(f(X) - \mathbb{E}(f(X)) \geq t\big) \leq e^{\frac{-t^2}{2L^2}} \tag{4.13}$$

This theorem will be used on the function F defined in (4.11) and dependent on N^2 i.d.d. standard Gaussian random variables $X = (X_{11}, ..., X_{1N}, ..., X_{N1}, ..., X_{NN})$

$$F(X) = \log\Big(\sum_{m\in\Gamma_N} g(m) \sum_{\substack{\boldsymbol{\sigma}\in S_N \\ m(\boldsymbol{\sigma})=m}} \exp\big(\beta\frac{B}{AN} \sum_{1\leq i<j\leq N} X_{ij}\sigma_i\sigma_j\big)\Big). \tag{4.14}$$

Hereby the following probabilistic bounds can be determined,

$$\begin{aligned} 1-\mathbb{P}\big(F(X) - \mathbb{E}(F(X)) \leq t\big) &\leq e^{\frac{-t^2}{2L^2}} \\ \mathbb{P}\big(F(X) - \mathbb{E}(F(X)) \leq t\big) &\geq 1 - e^{\frac{-t^2}{2L^2}} \\ \mathbb{P}\big(\exp\big[F(X) - \mathbb{E}(F(X))\big] \leq e^t\big) &\geq 1 - e^{\frac{-t^2}{2L^2}}. \end{aligned} \tag{4.15}$$

Since F is L-Lipschitz, $-F$ is L-Lipschitz as well, and the following bound can be obtained.

$$\begin{aligned} 1-\mathbb{P}\big(\mathbb{E}(F(X)) - F(X) \leq t\big) &\leq e^{\frac{-t^2}{2L^2}} \\ \mathbb{P}\big(F(X) - \mathbb{E}(F(X)) \geq -t\big) &\geq 1 - e^{\frac{-t^2}{2L^2}} \\ \mathbb{P}\big(\exp\big[F(X) - \mathbb{E}(F(X))\big] \geq e^{-t}\big) &\geq 1 - e^{\frac{-t^2}{2L^2}}. \end{aligned} \tag{4.16}$$

The Lipschitz-constant L will be determined in Section 4.6

4.4. Upper bound on the deterministic term

Now that we have bounds for the random part of $\mathcal{Z}$, we are left with the deterministic part, $e^{\mathbb{E}(F)}$, of which we will construct a lower and an upper bound. For the upper bound first $\mathbb{E}(\mathcal{Z})$ will be calculated and then by Jensen's inequality an upper bound will be obtained in the form

$$\mathbb{E}(\log(\mathcal{Z})) \leq \log(\mathbb{E}(\mathcal{Z})). \tag{4.17}$$

To calculate this upper bound we will have to use the moment generating function of $\hat{J}_{ij}$, defined as follows

$$M_{\hat{J}_{ij}}(x) := \mathbb{E}(e^{x\hat{J}_{ij}}). \tag{4.18}$$

Performing a Taylor expansion of the moment generating function we obtain the following

$$\begin{aligned} M_{\hat{J}_{ij}}(x) &= M_{\hat{J}_{ij}}(0) + xM'_{\hat{J}_{ij}}(0) + \frac{x^2}{2}M''_{\hat{J}_{ij}}(0) + o(x^2) \\ &= M_{\hat{J}_{ij}}(0) + x\mathbb{E}(\hat{J}_{ij}) + \frac{x^2}{2}\mathbb{E}(\hat{J}_{ij}^2) + o(x^2) \\ &= 1 + 0 + \frac{x^2}{2}B^2 + o(x^2) \end{aligned} \tag{4.19}$$

Since $M_{\hat{J}_{ij}}(x)$ is always positive awe can use the expansion $\log(1+x) = x + o(x)$ in order to rewrite the moment generating function as

$$M_{\hat{J}_{ij}}(x) = \exp\big(\log(M_{\hat{J}_{ij}}(x)\big) = \exp\big(\frac{x^2}{2}B^2 + o(x^2)\big). \tag{4.20}$$

By using the fact that all the $\hat{J}_{ij}$ are independent, the expectation of $\mathcal{Z}$ can be written as a product.

$$\begin{aligned} \mathbb{E}(\mathcal{Z}) &= \sum_{m\in\Gamma_N} g(m) \sum_{\substack{\boldsymbol{\sigma}\in S_N \\ m(\boldsymbol{\sigma})=m}} \mathbb{E}\big(\exp(\frac{\beta}{AN}\sum_{1\le i<j\le N}\hat{J}_{ij}\sigma_i\sigma_j\big) \\ &= \sum_{m\in\Gamma_N} g(m) \sum_{\substack{\boldsymbol{\sigma}\in S_N \\ m(\boldsymbol{\sigma})=m}} \prod_{1\le i<j\le N} \mathbb{E}\big(\exp(\frac{\beta}{AN}\hat{J}_{ij}\sigma_i\sigma_j\big) \end{aligned} \tag{4.21}$$

In this equation the expectation can be recognised as the moment generating function $M_{\hat{J}_{ij}}(\frac{\beta}{AN}\sigma_i\sigma_j)$. Now since $\sigma_i, \sigma_j \in \{-1, +1\}$ a result independent of any random variables is obtained,

$$\mathbb{E}(\mathcal{Z}) = \sum_{m\in\Gamma_N} g(m) \sum_{\substack{\boldsymbol{\sigma}\in S_N \\ m(\boldsymbol{\sigma})=m}} \prod_{1\le i<j\le N} \exp\left(\frac{\frac{\beta^2}{A^2N^2}}{2}B^2 + o(\frac{\beta^2}{A^2N^2})\right). \tag{4.22}$$

Using the facts that Π multiplies over $\frac{N(N-1)}{2}$ entries and that the second $\sum$ sums over $\binom{N}{\frac{1+m}{2}N}$ entries as seen in the previous chapter, we can come to the following bound.

$$\begin{aligned} \mathbb{E}(\mathcal{Z}) &= \sum_{m\in\Gamma_N} g(m)\binom{N}{\frac{1+m}{2}N} \exp\left(\frac{\beta^2}{2A^2N^2}B^2 + o(\frac{1}{N^2})\right)^{\frac{N(N-1)}{2}} \\ &= \sum_{m\in\Gamma_N} g(m)\binom{N}{\frac{1+m}{2}N} \exp\left(\frac{\beta^2N(N-1)}{4A^2N^2}B^2 + o(1)\right) \end{aligned} \tag{4.23}$$

Now since we used some expensions in this derivation this holds asymptotically for $\frac{\beta}{AN} \to 0$ and this case can be accomplished by letting $N \to \infty$. Then the following bound is obtained,

$$\begin{aligned} \mathbb{E}(\log(\mathcal{Z})) \le \log(\mathbb{E}(\mathcal{Z})) &= \log\big(\sum_{m\in\Gamma_N} g(m)\binom{N}{\frac{1+m}{2}N}\big) + \frac{\beta^2}{4A^2}B^2 + o(1) \\ &= \log\big(\sum_{m\in\Gamma_N} g(m)\binom{N}{\frac{1+m}{2}N}\big) + \alpha + o(1), \end{aligned} \tag{4.24}$$

where $\alpha = \frac{\beta^2}{4A^2}B^2$ such that later results can be written down easier.

4.5. Lower bound

An important tool in the calculation to obtain a lower bound is the Paley-Zygmound inequality. It states that for any non negative random variable X and any constant $\eta \in (0, 1)$ the following holds.

$$\mathbb{P}\big(X \ge \eta\mathbb{E}(X)\big) \ge (1-\eta)^2\frac{\mathbb{E}(X)^2}{\mathbb{E}(X^2)}. \tag{4.25}$$

This inequality will be used on the random variable $\mathcal{Z}$ and together with the concentration inequality, two events will be constructed of which the sum of their probabilities will be bigger then 1. This will

imply that there is an overlap in these two events. From the fact that we know that this overlap exists a lower bound for $\mathcal{Z}$ will be constructed.

Using the Payley-Zygmound inequality, we obain for $\mathcal{Z}$

$$\begin{aligned}\mathbb{P}\big(\mathcal{Z} \geq \eta\mathbb{E}(\mathcal{Z})\big) &= \mathbb{P}\big(\log(\mathcal{Z}) \geq \log(\eta\mathbb{E}(\mathcal{Z})\big)\\ =&\mathbb{P}\big(\log(\mathcal{Z}) \geq \log(\eta) + \log(\mathbb{E}(\mathcal{Z})\big) \geq (1-\eta)^2\frac{\mathbb{E}(\mathcal{Z})^2}{\mathbb{E}(\mathcal{Z}^2)}\end{aligned} \tag{4.26}$$

Where for the last part we have to calculate an upper bound for $\mathbb{E}(\mathcal{Z}^2)$ which will be done with a similar calculation as for the upper bound of $\mathbb{E}(\mathcal{Z})$.

$$\begin{aligned}\mathbb{E}(\mathcal{Z}^2) &= \mathbb{E}\Big(\sum_{m,m'\in\Gamma_N} g(m)g(m') \sum_{\substack{\boldsymbol{\sigma}^{(1)},\boldsymbol{\sigma}^{(2)}\in S_N\\ m(\boldsymbol{\sigma}^{(1)})=m,\\ m(\boldsymbol{\sigma}^{(2)})=m'}} \exp\big(\sum_{1\leq i<j\leq N}\frac{\beta}{NA}\hat{J}_{ij}(\sigma_i^{(1)}\sigma_j^{(1)}+\sigma_i^{(2)}\sigma_j^{(2)})\big)\Big)\\ &= \sum_{m,m'\in\Gamma_N} g(m)g(m') \sum_{\substack{\boldsymbol{\sigma}^{(1)},\boldsymbol{\sigma}^{(2)}\in S_N\\ m(\boldsymbol{\sigma}^{(1)})=m,\\ m(\boldsymbol{\sigma}^{(2)})=m'}} \mathbb{E}\Big(\exp\big(\sum_{1\leq i<j\leq N}\frac{\beta}{NA}\hat{J}_{ij}(\sigma_i^{(1)}\sigma_j^{(1)}+\sigma_i^{(2)}\sigma_j^{(2)})\big)\Big)\\ &= \sum_{m,m'\in\Gamma_N} g(m)g(m') \sum_{\substack{\boldsymbol{\sigma}^{(1)},\boldsymbol{\sigma}^{(2)}\in S_N\\ m(\boldsymbol{\sigma}^{(1)})=m,\\ m(\boldsymbol{\sigma}^{(2)})=m'}} \prod_{1\leq i<j\leq N} \mathbb{E}\Big(\exp\big(\frac{\beta}{NA}\hat{J}_{ij}(\sigma_i^{(1)}\sigma_j^{(1)}+\sigma_i^{(2)}\sigma_j^{(2)})\big)\Big).\end{aligned} \tag{4.27}$$

Here we can again recognise the part that looks the same as the moment generating function $M_{\hat{J}_{ij}}(\frac{\beta}{AN}(\sigma_i^{(1)}\sigma_j^{(1)}+\sigma_i^{(2)}\sigma_j^{(2)})$. Substituting this into the equation and by using that $(\sigma_i^{(1)}\sigma_j^{(1)}+\sigma_i^{(2)}\sigma_j^{(2)})^2$ is at most 4, we obtain

$$\begin{aligned}\mathbb{E}(\mathcal{Z}^2) = &\sum_{m,m'\in\Gamma_N} g(m)g(m') \sum_{\substack{\boldsymbol{\sigma}^{(1)},\boldsymbol{\sigma}^{(2)}\in S_N\\ m(\boldsymbol{\sigma}^{(1)})=m,\\ m(\boldsymbol{\sigma}^{(2)})=m'}} \prod_{1\leq i<j\leq N}\\ &\exp\left(\frac{\beta^2B^2}{2N^2A^2}((\sigma_i^{(1)}\sigma_j^{(1)}+\sigma_i^{(2)}\sigma_j^{(2)})^2 + o(\frac{\beta^2}{N^2A^2}(\sigma_i^{(1)}\sigma_j^{(1)}+\sigma_i^{(2)}\sigma_j^{(2)})^2)\right)\\ \leq &\sum_{m,m'\in\Gamma_N} g(m)g(m') \sum_{\substack{\boldsymbol{\sigma}^{(1)},\boldsymbol{\sigma}^{(2)}\in S_N\\ m(\boldsymbol{\sigma}^{(1)})=m,\\ m(\boldsymbol{\sigma}^{(2)})=m'}} \prod_{1\leq i<j\leq N} \exp\left(\frac{\beta^2B^22}{N^2A^2} + o(\frac{4\beta^2}{A^2N^2})\right).\end{aligned} \tag{4.28}$$

Using the same cardinality rules as used in the proof for the upperbbound of $\mathbb{E}(\mathcal{Z})$, we can rewrite this as

$$\mathbb{E}(\mathcal{Z}^2) \leq \sum_{m,m'\in\Gamma_N} g(m)g(m')\binom{N}{\frac{1+m}{2}N}\binom{N}{\frac{1+m'}{2}N}\exp\left(\frac{\beta^2B^2(N-1)}{NA^2}+o(1)\right). \tag{4.29}$$

In this we can recognise the expression for $\mathbb{E}(\mathcal{Z})$ found in (4.23), and using the constant α defined in (4.24), we have

$$\begin{aligned}\mathbb{E}(\mathcal{Z}^2) &\leq \sum_{m\in\Gamma_N}\binom{N}{\frac{1+m}{2}N}\mathbb{E}(\mathcal{Z})\exp\left(\frac{3\beta^2B^2(N-1)}{4NA^2}+o(1)\right)\\ &= \mathbb{E}(\mathcal{Z})^2\exp\left(\frac{\beta^2B^2(N-1)}{2NA^2}+o(1)\right)\\ &= \mathbb{E}(\mathcal{Z})^2\exp(2\alpha+o(1)).\end{aligned} \tag{4.30}$$

Now plugging this in the Payley-Zygmound inequality the following probabilistic bound is found

$$\mathbb{P}\bigg(\log(\mathcal{Z}) \geq \log(\eta) + \log(\mathbb{E}(\mathcal{Z}))\bigg) \geq \frac{(1-\eta)^2}{\exp(2\alpha + o(1))}. \tag{4.31}$$

Thus we have constructed the first probabilistic bound and we will now construct the second probabilistic bound. This bound will be constructed from the concentration inequality namely from the following inequality from equations (4.15)

$$\mathbb{P}\bigg(\log(\mathcal{Z}) \leq \mathbb{E}(\log(\mathcal{Z})) + t\bigg) \geq 1 - e^{-\frac{t^2}{2L^2}}. \tag{4.32}$$

If the probabilities from (4.31) and (4.32) added together are greater then 1 the events overlap. This overlap would imply

$$\log(\eta) + \log(\mathbb{E}(\mathcal{Z})) \leq \log(\mathcal{Z}) \leq \mathbb{E}(\log(\mathcal{Z})) + t \tag{4.33}$$

which is contained in:

$$\log(\eta) + \log(\mathbb{E}(\mathcal{Z}) \leq \mathbb{E}(\log(\mathcal{Z})) + t \tag{4.34}$$

which implies

$$\mathbb{E}(\log(\mathcal{Z})) \geq \log(\eta) - t + \log(\mathbb{E}(\mathcal{Z}). \tag{4.35}$$

We can notice that this is a deterministic bound. The only part that now remains is to see when this overlap occurs and what then the bound becomes since in the above expression η and t are still arbitrary constants. Since the added probability has to be greater then 1 a condition is,

$$e^{-\frac{t^2}{2L^2}} < \frac{(1-\eta)^2}{\exp(2\alpha + o(1))}. \tag{4.36}$$

Since t, η, L and $\exp(2\alpha + o(1))$ are all positive the, following condition is easily obtained.

$$t > \sqrt{2L^2(2\alpha + o(1) - 2\log(1-\eta))}. \tag{4.37}$$

Combining all the parts found above we find the following lower bound depending on $\eta \in (0,1)$

$$\mathbb{E}(\log(\mathcal{Z})) \geq \log\left(\sum_{m\in\Gamma_N} g(m)\binom{N}{\frac{1+m}{2}N}\right) + \alpha + o(1) + \log(\eta) - 2L\sqrt{\alpha - \log(1-\eta)} \tag{4.38}$$

Here we have thus constructed a lower bound that holds for all $\eta \in (0,1)$. The best bound is such that the lower bound is maximised, thus we can write the lower bound as

$$\mathbb{E}(\log(\mathcal{Z})) \geq \log\left(\sum_{m\in\Gamma_N} g(m)\binom{N}{\frac{1+m}{2}N}\right) + \kappa + o(1) \tag{4.39}$$

with

$$\kappa = \alpha + \max_{\eta\in(0,1)} \left(\log(\eta) - 2L\sqrt{\alpha - \log(1-\eta)}\right) \tag{4.40}$$

4.6. Lipschitzianity condition

To justify the use of Theorem from Section 4.3, the Lipschitz constant L has to be found for the function F in (4.14).

For this we will first consider the following part of the above mentioned function.

$$K(X) = \frac{\beta B}{AN} \sum_{1\leq i<j\leq N} X_{ij}\sigma_i\sigma_j \tag{4.41}$$

This can be seen as the inner product between the vectors $X = (X_{11}, ..., X_{NN})$ and $V = \frac{\beta B}{AN}(\sigma_1\sigma_1, ..., \sigma_N\sigma_N)$, with indices such that $1 \leq i < j \leq N$. If we have 2 random variables X, Y distributed according to X we can say that,

$$|K(X) - K(Y)| = |V \cdot X - V \cdot Y| \leq \|V\|_2 d(X, Y) \tag{4.42}$$

where in the last part we use the Cauchy Schwarz inequality i.e. $|ab| \leq \|a\|_2\|b\|_2$
From this we can perform the following calculations.

$$\begin{aligned} F(X) &= \log\Big(\sum_{m\in\Gamma_N} g(m) \sum_{\substack{\boldsymbol{\sigma}\in S_N \\ m(\boldsymbol{\sigma})=m}} \exp\big(\beta\frac{B}{AN}\sum_{1\leq i<j\leq N} X_{ij}\sigma_i\sigma_j\big)\Big) \\ &\leq \log\Big(\sum_{m\in\Gamma_N} g(m) \sum_{\substack{\boldsymbol{\sigma}\in S_N \\ m(\boldsymbol{\sigma})=m}} \exp\big(\beta\frac{B}{AN}\sum_{1\leq i<j\leq N} Y_{ij}\sigma_i\sigma_j\big)\exp\big(\|V\|_2 d(X,Y)\big)\Big) \\ &= F(Y) + \|V\|_2 d(X, Y) \end{aligned} \tag{4.43}$$

Doing the same calculations but swapping X an Y we obtain the following result.

$$|F(X) - F(Y)| \leq \|V\|_2 d(X, Y) \tag{4.44}$$

Thus F is a $\|V\|_2$-Lipschitz function and therefore to calculate the Lipschitz constant the only part left to do is calculate $\|V\|_2$. We have

$$\begin{aligned} \|V\|_2 &= \sqrt{\sum_{1\leq i<j\leq N} \big(\frac{\beta B}{AN}\sigma_i\sigma_j\big)^2} \\ &= \frac{\beta B}{AN}\sqrt{\sum_{1\leq i<j\leq N} \sigma_i^2\sigma_j^2} \\ &= \frac{\beta B}{AN}\sqrt{\frac{N(N-1)}{2}} \leq \frac{\beta B}{A\sqrt{2}} = L \end{aligned} \tag{4.45}$$

where in the last part we use that $\sigma_i^2 = 1$. Thus we have obtained the Lipschitz constant L that is used in the final result. And have thereby also shown that the **Theorem 3.2** can be used.

4.7. Proof of the main theorem

With all the results that have been proven in the above sections we can finally prove Theorem 3.1. We will first start with the first bound of the main theorem. We first recall the definition of $\mathcal{Z}$ and the form given in (4.12)

$$\sum_{m\in\Gamma_N} \bar{g}(m) \sum_{\substack{\boldsymbol{\sigma}\in S_N \\ m(\boldsymbol{\sigma})=m}} e^{\beta\Delta(\boldsymbol{\sigma})} = e^{\mathbb{E}(F)}e^{F-\mathbb{E}(F)} \tag{4.46}$$

Now multiplying both parts in the concentration inequality found in (4.16) with $e^{\mathbb{E}(F)}$ we obtain

$$\mathbb{P}\big(\sum_{m\in\Gamma_N} \bar{g}(m) \sum_{\substack{\boldsymbol{\sigma}\in S_N \\ m(\boldsymbol{\sigma})=m}} e^{\beta\Delta(\boldsymbol{\sigma})} \leq e^t e^{\mathbb{E}(F)}\big) \geq 1 - e^{\frac{-t^2}{2L^2}} \tag{4.47}$$

by using the upper bound for $\mathbb{E}(\log(\mathcal{Z})) = \mathbb{E}(F)$ we find that

$$\mathbb{P}\Big(\sum_{m\in\Gamma_N} \bar{g}(m) \sum_{\substack{\boldsymbol{\sigma}\in S_N \\ m(\boldsymbol{\sigma})=m}} e^{\beta\Delta(\boldsymbol{\sigma})} \leq e^{t+\alpha}\big(\sum_{m\in\Gamma_N} \bar{g}(m)\binom{N}{\frac{1+m}{2}N}\big)(1+o(1))\Big) \geq 1 - e^{\frac{-t^2}{2L^2}}. \tag{4.48}$$

Now by choosing $\bar{g}(m)$ to be a a specific function namely $\frac{e^{-\beta \tilde{H}_N(m)}}{Z} g(m)$ and recognising previous definitions we can finish the proof

$$\mathbb{P}\Big(\sum_{m\in\Gamma_N} \frac{e^{-\beta \tilde{H}_N(m)}}{Z} g(m) \sum_{\substack{\boldsymbol{\sigma}\in S_N \\ m(\boldsymbol{\sigma})=m}} e^{\beta\Delta(\boldsymbol{\sigma})} \leq e^{t+\alpha} \big(\sum_{m\in\Gamma_N} \frac{e^{-\beta \tilde{H}_N(m)}}{Z} g(m) \binom{N}{\frac{1+m}{2}N} \big)(1+o(1)) \Big) \geq 1 - e^{\frac{-t^2}{2L^2}}$$

$$\mathbb{P}\Big(\sum_{m\in\Gamma_N} g(m) Q^{RDCW}(m) \leq e^{t+\alpha} \frac{Z^{CW}}{Z^{RDCW}} \big(\sum_{m\in\Gamma_N} g(m) Q^{CW}(m) \big)(1+o(1)) \Big) \geq 1 - e^{\frac{-t^2}{2L^2}}, \tag{4.49}$$

with $L = \frac{\beta B}{A\sqrt{2}}$

The same procedure can be used to find the second part of the main theorem. A detail to realise is that the only part in this proof where we used that the random variable had to be Gaussian is in the concentration inequality thus this proof can be used similarly for any other random variable that has a concentration inequality that looks like the one we used.

5

Extension for bounded and sub-Gaussian random variables

In the previous chapter we compared the Curie-Weiss model to the Dilute Curie-Weiss model with Gaussian variables. In this chapter the same techniques will be used to derive a similar theorem for bounded and sub Gaussian random variables, since most of the proofs consist of the same calculations but with some extra constants. We will skip these calculations and only indicate where the extra constants come up. The places where the proofs differ will also be discussed.

5.1. Bounded random variables

In this section, $(J_{ij})_{ij}$ will be a triangular arry of bounded random variables, meaning that $\exists M > 0$ such that

$$\left| J_{ij} \right| \leq M \tag{5.1}$$

For the rest we will describe the mean and variance of J_{ij} the same way as for the Gaussian case i.e. $\mathbb{E}(J_{ij}) = A$ and $Var(J_{ij}) = B^2$.

For this random variable we can split the Curie-Weiss model the same way as for the Gaussian case obtaining

$$Q^{RDCW}(m) = \frac{e^{-\beta \tilde{H}_N(m)}}{Z^R DCW}\Big(\sum_{\substack{\boldsymbol{\sigma} \in S_N \\ m(\boldsymbol{\sigma})=m}} e^{-\beta \Delta(\sigma)} \Big) = \frac{e^{-\beta \tilde{H}_N(m)}}{Z^{RDCW}}\Big(\sum_{\substack{\boldsymbol{\sigma} \in S_N \\ m(\boldsymbol{\sigma})=m}} e^{\frac{\beta}{AN} \sum_{1 \leq i < j \leq N} (J_{ij} - A)\sigma_i \sigma_j} \Big). \tag{5.2}$$

Where Z^{RDCW} is a normalisation constant.

From this split we again have to obtain an concentration inequality and the upper and lower bounds for the deterministic parts.

5.1.1. Random part

For the concentration inequality we again will use the function,

$$F = \log(\mathcal{Z}) = \log\Big(\sum_{m \in \Gamma_N} g(m) \sum_{\substack{\boldsymbol{\sigma} \in S_N \\ m(\boldsymbol{\sigma})=m}} e^{-\beta \Delta(\boldsymbol{\sigma})} \Big). \tag{5.3}$$

From this we again can use the same calculations for the upper and lower bound as in the calculations for the upper and lower bounds where some extra factors come in.

To calculate the random part we will use the theorem 6.2 from [1] Where a function f of random variables $x_1, ..., x_n$ has the bounded difference property if for some nonnegative constants $c_1, ..., c_n$

$$\sup_{x_1,...,x_n} \left| f(x_1,...,x_n) - f(x_1,...,x_{i-1},x_i',x_{i+1},...x_n) \right| \leq c_i \qquad 1 \leq i \leq n, \tag{5.4}$$

here x_i' is a random variable that is i.i.d. to x_i

Theorem 5.1. Assume that the function f satisfies the bounded differences assumption with constants $c_1,...,c_n$ and denote

$$v = \frac{1}{4}\sum_{i=1}^{n} c_i^2. \tag{5.5}$$

Let $Z = f(X_1,...,X_n)$ where the X_i are independent. Then

$$\mathbb{P}\{Z - \mathbb{E}(Z) > t\} \leq e^{-t^2/2v} \tag{5.6}$$

Note that since the bounded differences assumption is symmetric, Z also satisfies the lower-tail inequality

$$\mathbb{P}\{f - \mathbb{E}(f) > t\} \leq e^{-t^2/2v} \tag{5.7}$$

To calculate these c_i we will rewrite the above expression as

$$\left| f(x_1,...,x_n) - f(x_1,...,x_{i-1},x_i',x_{i+1},...x_n) \right| = \left| f_X - f_{X'} \right| \tag{5.8}$$

Thus in order to apply this theorem to our case of bounded random variables we will have to prove (5.4) on the function F specified above, where F is dependent of $\hat{J}$ which is $J - A$.

The most important step in calculating this is writing the difference as an integral,

$$\left| F_{\hat{J}} - F_{\hat{J}'} \right| = \left| \int_0^1 \frac{d}{dt} F\big(\hat{J}t + (1-t)\hat{J}'\big)dt \right|, \tag{5.9}$$

where $\hat{J}'$ is an independent copy of $\hat{J}$ This statement can be easily checked by calculating the integral on the right side.

By differentiating the expression we now have left we can obtain the bounded difference constants c_i, here we will write $\Delta_{\hat{J}}(\boldsymbol{\sigma})$ to make clear the Δ depends on $\hat{J}$,

$$
\begin{aligned}
&\left| \int_0^1 \frac{d}{dt} F\big(\hat{J}t + (1-t)\hat{J}'\big)dt \right| \\
=&\left| \int_0^1 \frac{d}{dt} \log\big(\mathcal{Z}(\hat{J}t + (1-t)\hat{J}')\big)dt \right| \\
=&\left| \int_0^1 \frac{1}{\mathcal{Z}} \frac{d}{dt} \Big(\sum_{m\in\Gamma_N} g(m) \sum_{\substack{\boldsymbol{\sigma}\in S_N \\ m(\boldsymbol{\sigma})=m}} e^{-\beta\Delta_{\hat{J}t+(1-t)\hat{J}'}(\boldsymbol{\sigma})}\Big)dt \right| \\
=&\left| \int_0^1 \frac{1}{\mathcal{Z}} \sum_{m\in\Gamma_N} g(m) \sum_{\substack{\boldsymbol{\sigma}\in S_N \\ m(\boldsymbol{\sigma})=m}} \frac{d}{dt} e^{-\beta\Delta_{\hat{J}t+(1-t)\hat{J}'}(\boldsymbol{\sigma})}dt \right| \\
=&\left| \int_0^1 \frac{1}{\mathcal{Z}} \sum_{m\in\Gamma_N} g(m) \sum_{\substack{\boldsymbol{\sigma}\in S_N \\ m(\boldsymbol{\sigma})=m}} \frac{d}{dt} e^{\frac{\beta}{AN}\sum_{1\le i<j\le N}(t\hat{J}_{ij}+(1-t)\hat{J}'_{ij})\sigma_i\sigma_j} dt \right| \\
=&\left| \int_0^1 \sum_{m\in\Gamma_N} g(m) \sum_{\substack{\boldsymbol{\sigma}\in S_N \\ m(\boldsymbol{\sigma})=m}} \Big(\sum_{1\le i<j\le N} \frac{\beta}{AN}(\hat{J}_{ij}-\hat{J}'_{ij})\sigma_i\sigma_j \frac{1}{\mathcal{Z}} e^{\frac{\beta}{AN}\sum_{1\le i<j\le N}(t\hat{J}_{ij}+(1-t)\hat{J}'_{ij})\sigma_i\sigma_j} dt \right| \\
=&\left| \sum_{1\le i<j\le N} \frac{\beta}{AN}(\hat{J}_{ij}-\hat{J}'_{ij}) \int_0^1 \sum_{m\in\Gamma_N} g(m) \sum_{\substack{\boldsymbol{\sigma}\in S_N \\ m(\boldsymbol{\sigma})=m}} \sigma_i\sigma_j \frac{1}{\mathcal{Z}} e^{\frac{\beta}{AN}\sum_{1\le i<j\le N}(t\hat{J}_{ij}+(1-t)\hat{J}'_{ij})\sigma_i\sigma_j} dt \right|
\end{aligned}
\tag{5.10}
$$

After this long calculation to rewrite our bounded difference constant we can now start to identify some parts in our last expression and simplify them. Firstly by construction$\hat{J}_{ij} - \hat{J}'_{ij}$ is only non zero for one choice of ij. This makes us able to remove one of the summation signs thereby obtaining the following,

$$
\left| F_{\hat{J}} - F_{\hat{J}'} \right| = \left| \frac{\beta}{AN}(J_{kl} - \hat{J}_{kl}) \int_0^1 \sum_{m\in\Gamma_N} g(m) \sum_{\substack{\boldsymbol{\sigma}\in S_N \\ m(\boldsymbol{\sigma})=m}} \sigma_k\sigma_l \frac{1}{\mathcal{Z}} e^{-\beta\Delta_{t\hat{J}_{ij}+(1-t)\hat{J}'_{ij}}} dt \right|, \tag{5.11}
$$

where kl is the entry where J and $\hat{J}$ differ.
By using the triangle inequality we obtain

$$
\left| F_{\hat{J}} - F_{\hat{J}'} \right| \le \frac{\beta}{AN} \left| \hat{J}_{kl} - \hat{J}'_{kl} \right| \left| \int_0^1 \sum_{m\in\Gamma_N} g(m) \sum_{\substack{\boldsymbol{\sigma}\in S_N \\ m(\boldsymbol{\sigma})=m}} \sigma_k\sigma_l \frac{1}{\mathcal{Z}} e^{-\beta\Delta_{t\hat{J}_{ij}+(1-t)\hat{J}'_{ij}}} dt \right|. \tag{5.12}
$$

For the final part we can look at the definition of $\mathcal{Z}$ and from this it is obvious that the part inside the integral is clearly less then 1 and thus the entire absolute value of the integral is less then one,

$$
\left| F_{\hat{J}} - F_{\hat{J}'} \right| \le \frac{\beta}{AN} \left| \hat{J}_{kl} - \hat{J}'_{kl} \right|. \tag{5.13}
$$

This derivation has been done without using the bounded condition. If we use the bounded condition stated at the start of this chapter we obtain the following bounded difference constant,

$$
c_i = \frac{\beta}{AN} 2M. \tag{5.14}
$$

These c_i give us the following

$$
v = \frac{M^2\beta^2 N(N-1)}{A^2N^2 2} \le \frac{M^2\beta^2}{A^2 2} \tag{5.15}
$$

So now we can use **Theorem 5.1** thereby obtaining a concentration inequality for bounded random variables

$$\mathbb{P}\{f - \mathbb{E}(f) < -t\} \leq e^{-t^2A^2/M^2\beta^2}, \tag{5.16}$$

and

$$\mathbb{P}\{f - \mathbb{E}(f) > t\} \leq e^{-t^2A^2/M^2\beta^2}. \tag{5.17}$$

The rest of the proof of the closeness for the randomly dilute Curie-Weiss model with bounded random variables to the Curie-Weiss model will follow the same path as has been used for the randomly dilute Curie-Weiss model with Gaussian random variables. However in the rest of the proof we will replace L with $\frac{\beta M}{A\sqrt{2}}$.

5.2. Sub-Gaussian

Just as in the previous section we have proved closeness for bounded random variables, in this section we will prove the same closeness for Sub-Gaussian random variables. This proof will use a result that looks a lot like Theorem 5.1. However this theorem uses some properties from sub-Gaussian random variables that will be discussed first.

5.2.1. Sub Gaussian random variables and the Sub-Gaussian diameter

In this section we will introduce a definition of sub-Gaussian random variables. After introducing the sub-Gaussian random variables we will discuss the sub-Gaussian diameter, an important property of random variables from which we will construct a concentration inequality. This and the following section is based on [4] and [3]

Definition A random Variable is a sub-Gaussian random variable of class $\mathcal{G}(\Delta^2)$ if the following holds

$$\mathbb{E}(e^{\lambda(X-\mathbb{E}(X))}) \leq e^{\frac{\Delta^2\lambda^2}{2}} \qquad \forall\lambda \in \mathbb{R}. \tag{5.18}$$

This is implies that

$$\mathbb{P}\big(X - \mathbb{E}(X) > t\big) \leq e^{\frac{-t^2}{2\Delta^2}} \qquad t > 0 \tag{5.19}$$

and

$$\mathbb{P}\big(X - \mathbb{E}(X) < -t\big) \leq e^{\frac{-t^2}{2\Delta^2}} \qquad t > 0 \tag{5.20}$$

This implication can be proven using the Markov inequality which state that for non negative random variables Y we have that $\mathbb{P}(Y > a) \leq \frac{\mathbb{E}(Y)}{a}$. Therefore we have

$$\mathbb{P}\big((X - \mathbb{E}(X)) > t\big) = \mathbb{P}\big(e^{\lambda(X-\mathbb{E}(X))} > e^{\lambda t}\big) \leq \frac{\mathbb{E}(e^{\lambda(X-\mathbb{E}(X))})}{e^{\lambda t}} \leq e^{\frac{\Delta^2\lambda^2}{2} - \lambda t} \tag{5.21}$$

Since this equation holds for all λ we can optimise the equation over λ. This bound is optimized when

$$\frac{d}{d\lambda} e^{\frac{\Delta^2\lambda^2}{2} - \lambda t} = 0 \tag{5.22}$$

which results in

$$\lambda = \frac{t}{\Delta^2}. \tag{5.23}$$

From this we obtain the final bound

$$\mathbb{P}\big((X - \mathbb{E}(X)) > t\big) \leq e^{\frac{t^2}{2\Delta^2} - \frac{t^2}{\Delta^2}} = e^{\frac{-t^2}{2\Delta^2}}. \tag{5.24}$$

The case with the minus sign can be proved in the same way.

Now we will discuss the sub-Gaussian diameter a property of random variables that will be used to construct a concentration inequality.

A random variable is of class $\mathcal{SG}(\Delta_{sg}^2)$ if the random variable $\Xi(X) = \epsilon|X - X'|$ is of class $\mathcal{G}(\Delta_{sg}^2)$, where $\epsilon = 2\xi - 1$ with $\xi \sim Ber(\frac{1}{2})$, and X' is a independent copy of X.

The sub-Gaussian diameter of a random variable is the minimal Δ_{sg} such that the variable is of class $\mathcal{SG}(\Delta_{sg}^2)$.

In order to use the sub-Gaussian diameter we need to calculate an upper bound of the sub-Gaussian diameter of a sub-Gaussian random variable of class $\mathcal{G}(\Delta^2)$.

$$\begin{aligned}\mathbb{E}(e^{\lambda(\epsilon|X-X'|-\mathbb{E}(\epsilon|X-X'|)}) &= \mathbb{E}(e^{\lambda\epsilon|X-X'|-\frac{1}{2}\mathbb{E}(|X-X'|)-\frac{1}{2}\mathbb{E}(-|X-X'|)}) \\ &= \mathbb{E}(e^{\lambda\epsilon|X-X'|}) \\ &= \frac{1}{2}\mathbb{E}(e^{\lambda|X-X'|}) + \frac{1}{2}\mathbb{E}(e^{-\lambda|X-X'|})\end{aligned} \tag{5.25}$$

Now since X' is an independent copy of X we can split each of these 2 terms

$$\begin{aligned}\mathbb{E}(e^{\lambda|X-X'|}) &= \mathbb{P}(X > X')\mathbb{E}(e^{\lambda(X-X')}) + \mathbb{P}(X = X')\mathbb{E}(e^0) + \mathbb{P}(X < X')\mathbb{E}(e^{-\lambda(X-X')}) \\ &= 2\mathbb{P}(X > X')\mathbb{E}(e^{\lambda(X-X')}) + \mathbb{P}(X = X') \\ &= 2\mathbb{P}(X > X')\mathbb{E}(e^{\lambda(X-\mathbb{E}(X)-(X'-\mathbb{E}(X'))}) + \mathbb{P}(X = X') \\ &= 2\mathbb{P}(X > X')\mathbb{E}(e^{\lambda(X-\mathbb{E}(X))})\mathbb{E}(e^{-\lambda(X'-\mathbb{E}(X'))}) + \mathbb{P}(X = X') \\ &\leq 2\mathbb{P}(X > X')e^{\Delta^2\lambda^2} + \mathbb{P}(X = X')\end{aligned} \tag{5.26}$$

Where in the fourth row we used independence and in the final row we used the that our variable is sub-Gaussian. By noticing that our for sub-Gaussian random variables since they are bounded by a Gaussian $\mathbb{P}(X = X') = 0$. From this we get that $2\mathbb{P}(X > X') = 1$ thus by adding this to both terms we get

$$\mathbb{E}(e^{\lambda(\epsilon|X-X'|-\mathbb{E}(\epsilon|X-X'|)} \leq e^{\Delta^2\lambda^2}. \tag{5.27}$$

hereby obtaining a bound for the sub-Gaussian diameter of $\Delta_{sg} \leq \sqrt{2}\Delta$

5.2.2. Concentration inequality

Now that we have found an upper bound for the sub-Gaussian diameter of sub-Gaussian random variable, we can construct a concentration inequality. To construct this inequality we will use the following theorem. This section is again based on [3] and [4]

Theorem 5.2 for $c_1, ..., c_N \geq 0$, Let X_i be random variables taking values in $\mathcal{R}$ of class $\mathcal{SG}(\Delta_i^2)$ and let $f : \mathbb{R}^N \to \mathbb{R}$ satisfying the Lipschitz condition

$$|f(x_1, ..., x_N) - f(x_1', ..., x_N')| \leq \sum_{i=1}^{N} c_i|x_i - x_i'|. \tag{5.28}$$

Then for t>0

$$\mathbb{P}\big(|f(X) - \mathbb{E}(f(X))| > t\big) \leq 2\exp\big(\frac{-t^2}{2\Sigma_{i=1}^N c_i^2\Delta_i^2}\big) \tag{5.29}$$

holds.

Now by equation (5.13) we find that $c_i = \frac{\beta}{AN}$ and we know that sub-Gaussian random variables from class $\mathcal{G}(\Delta)$ has a sub-Gaussian diameter of Δ. Thereby we obtain the following probabilistic bound

$$\mathbb{P}\big(|f(X) - \mathbb{E}(f(X))| > t\big) \leq 2\exp\big(\frac{-t^2}{4\beta^2\Delta^2/A^2}\big) \tag{5.30}$$

Now from this bound on the absolute value of our function we can create an upper and lower tail bound by observing that the set we want to know the probability for is a subset of a set we already know the probability for. We get

$$\mathbb{P}\big(f(X) - \mathbb{E}(f(X)) > t\big) \leq \exp\big(\frac{-t^2}{4\beta^2\Delta^2/A^2}\big). \tag{5.31}$$

and in the same way we can use this for the case with the minus sign,

$$\mathbb{P}\big(f(X) - \mathbb{E}(f(X)) < -t\big) \leq \exp\big(\frac{-t^2}{4\beta^2\Delta^2/A^2}\big). \tag{5.32}$$

In all of these cases the eventual bound is of a single exponential. Sometimes a factor C can appear in front of this exponential. This factor alters the calculations for the lower bound a bit. then instead we obtain.

$$\kappa = \alpha + \max_{\eta\in(0,1)} \big(\log(\eta) - 2L\sqrt{\alpha + \log(C) - \log(1-\eta)}\big) \tag{5.33}$$

6

Conclusion

In this paper we have proved the closeness for the random dilute Curie-Weiss model to the Curie-Weiss model for Gaussian, bounded and sub-Gaussian couplings. Hereby we have demonstrated that the standard Curie-Weiss model is an acceptable approximation of the dilute Curie-Weiss model. And that we can apply its behaviour to actual magnets.

This closeness is stated in the following theorem, in terms of mesoscopic measures and averages of observables

Theorem The following bounds hold for any $s > 0$, where $Q(m)^{RDCW}$ is the mesoscopic distribution of the randomly dilute Curie-Weiss model and $Q(m)^{CW}$ is the mesoscopic distribution of the Curie-Weiss model. and Z^{RDCW}, Z^{CW} are their respective normalisation constants.

$$\mathbb{P}_J\Big(\sum_{m\in\Gamma_N} g(m)Q^{RDCW}(m) \leq e^{\alpha+s}\frac{Z^{CW}}{Z^{RDCW}}\big(\sum_{m\in\Gamma_N} g(m)Q^{CW}(m)\big)(1+\ o(1))\Big) \geq 1-k_1e^{-k_2s^2} \quad (6.1)$$

$$\mathbb{P}_J\Big(\sum_{m\in\Gamma_N} g(m)Q^{RDCW}(m) \geq e^{\kappa-s}\frac{Z^{CW}}{Z^{RDCW}}\big(\sum_{m\in\Gamma_N} g(m)Q^{CW}(m)\big)(1+o(1))\Big) \geq 1-k_1e^{-k_2s^2} \quad (6.2)$$

Where α, κ, k_1 and k_2 depend on the form of the coupling.
For Gaussian couplings with mean A and variance B^2 we have the following

$$\begin{aligned}\alpha =& \frac{\beta^2}{4A^2}B^2\\ \kappa =& \max_{\eta\in(0,1)}\big(\alpha+\log(\eta)-\frac{\beta B\sqrt{2}}{A}\sqrt{\alpha-\log(1-\eta)}\big)\\ k_1 =& 1\\ k_2 =& \frac{\beta^2B^2}{A^2}.\end{aligned} \quad (6.3)$$

For couplings bounded between $[-M, M]$ with mean A and variance B^2 we have the following

$$\begin{aligned}\alpha =& \frac{\beta^2}{4A^2}B^2\\ \kappa =& \max_{\eta\in(0,1)}\big(\alpha+\log(\eta)-\frac{\beta M\sqrt{2}}{A}\sqrt{\alpha-\log(1-\eta)}\big)\\ k_1 =& 1\\ k_2 =& \frac{\beta^2M^2}{A^2}.\end{aligned} \quad (6.4)$$

For couplings of class $\mathcal{G}(\Delta^2)$ with mean A and variance B^2 we have the following

$$
\begin{aligned}
\alpha =& \frac{\beta^2}{4A^2} B^2 \\
\kappa =& \max_{\eta \in (0,1)} \left(\alpha + \log(\eta) - 2 \frac{2\beta\Delta}{A} \sqrt{\alpha - \log(1-\eta)} \right) \\
k_1 =& 1 \\
k_2 =& \frac{4\beta^2 M^2}{A^2}.
\end{aligned}
\tag{6.5}
$$

References

[1] Massart Boucheron Lugosi. *Concentration Inequalities*. first. CLARENDON PRESS, 2012.

[2] Anton Bovier, Saeda Marello, and Elena Pulvirenti. “Metastability for the dilute Curie-Weiss model with Glauber dynamics”. English. In: *Electron. J. Probab.* 26 (2021). Id/No 47, p. 38. ISSN: 1083-6489. DOI: 10.1214/21-EJP610.

[3] Elena Pulvirenti Johan L. A. Dubbeldam Vicente Lenz and Martin Slowik. “”Metastability for the dilute Curie-Weiss-Potts model with Glauber dynamics”, work in progress.” In: ().

[4] Aryeh Kontorovich. *Concentration in unbounded metric spaces and algorithmic stability*. 2013. DOI: 10.48550/ARXIV.1309.1007. URL: https://arxiv.org/abs/1309.1007.

[5] Daniel V Schroeder. *An Introduction to Thermal Physics*. Oxford University Press, 2020.

[6] *wikimedia fully connected neural network*. URL: https://commons.wikimedia.org/wiki/File:Fully_connected_neural_network.svg.

A

Appendix

A.1. Stirling's approximation

Stirling's approximation is an approximation is an approximation of a factorial stating.

$$\ln(n!) = n\ln(n) - n + \Theta(\ln(n)) \tag{A.1}$$

This approximation will be applied to $\binom{N}{\frac{1+m}{2}N}$ to obtain the following approximation.

$$\begin{aligned}
\binom{N}{\frac{1+m}{2}N} &= \frac{N!}{\left(\frac{1+m}{2}N\right)!\left(\frac{1-m}{2}N\right)!} \\
&= e^{\ln\left(\frac{N!}{\left(\frac{1+m}{2}N\right)!\left(\frac{1-m}{2}N\right)!}\right)} \\
&= e^{\ln\left(N!\right)-\ln\left(\left(\frac{1+m}{2}N\right)!\right)-\ln\left(\left(\frac{1-m}{2}N\right)!\right)} \\
&= e^{N\ln(N)-N-\frac{1+m}{2}N\ln\left(\frac{1+m}{2}N\right)+\frac{1+m}{2}-\frac{1-m}{2}N\ln\left(\frac{1-m}{2}N\right)+\frac{1-m}{2}+\Theta(\ln(N))-\Theta(\ln\left(\frac{1+m}{2}N\right))-\Theta(\ln\left(\frac{1-m}{2}N\right))} \\
&= e^{N\ln(N)-\frac{1+m}{2}N\ln\left(\frac{1+m}{2}N\right)-\frac{1-m}{2}N\ln\left(\frac{1-m}{2}N\right)+\Theta(\ln(N))-\Theta(\ln\left(\frac{1+m}{2}N\right))-\Theta(\ln\left(\frac{1-m}{2}N\right))} \\
&= e^{N\ln(N)-\frac{1+m}{2}N\ln\frac{1+m}{2}N-\frac{1-m}{2}N\ln\left(\frac{1-m}{2}N\right)} \\
&= e^{N\ln(N)-\frac{1+m}{2}N\ln\left(\frac{1+m}{2}\right)-\frac{1-m}{2}N\ln\left(\frac{1-m}{2}\right)-\frac{1+m}{2}N\ln(N)-\frac{1-m}{2}N\ln(N)} \\
&= e^{N(-\frac{1-m}{2}\ln\left(\frac{1+m}{2}\right)-\frac{1-m}{2}\ln\left(\frac{1-m}{2}\right))}
\end{aligned} \tag{A.2}$$

obtaining the result used in the paper.

A.2. Differentiating the free energy of the standard Curie-Weiss model

first we will differentiate the free energy

$$\begin{aligned}
\frac{d}{dm}j(m) &= \frac{d}{dm}\Big(-\frac{\beta m^2}{2} - \beta h m + \frac{1-m}{2}\log\big(\frac{1-m}{2}\big) + \frac{1+m}{2}\log\big(\frac{1+m}{2}\big)\Big) \\
&= -\beta m - h\beta + \frac{d}{dm}\Big(\frac{1-m}{2}\log\big(\frac{1-m}{2}\big)\Big) + \frac{d}{dm}\Big(\frac{1+m}{2}\log\big(\frac{1m}{2}\big)\Big) \\
&= -\beta m - h\beta - \frac{1}{2}\log\big(\frac{1-m}{2}\big) + \frac{1-m}{2}\frac{d}{dm}\Big(\log\big(\frac{1-m}{2}\big)\Big) + \frac{1}{2}\log\big(\frac{1+m}{2}\big) + \frac{1+m}{2}\frac{d}{dm}\Big(\log\big(\frac{1+m}{2}\big)\Big) \\
&= -\beta m - h\beta - \frac{1}{2}\log\big(\frac{1-m}{2}\big) + \frac{1-m}{2}\big(\frac{1-m}{2}\big)^{-1}\frac{-1}{2} + \frac{1}{2}\log\big(\frac{1+m}{2}\big) + \frac{1+m}{2}\big(\frac{1+m}{2}\big)^{-1}\frac{1}{2} \\
&= -\beta m - h\beta - \frac{1}{2}\log\big(\frac{1-m}{2}\big) + \frac{-1}{2} + \frac{1}{2}\log\big(\frac{1+m}{2}\big) + \frac{1}{2} \\
&= -\beta m - h\beta - \frac{1}{2}\log\big(\frac{1-m}{2}\big) + \frac{1}{2}\log\big(\frac{1+m}{2}\big) \\
&= -\beta m - h\beta + \frac{1}{2}\log\big(\frac{1+m}{2}\frac{2}{1-m}\big) \\
&= -\beta m - h\beta + \tanh^{-1}(m)
\end{aligned} \tag{A.3}$$

By setting the derivative to zero the following condition is obtained.

$$\begin{aligned}
0 &= -\beta m - h\beta + \tanh^{-1}(m) \\
\beta m + h\beta &= \tanh^{-1}(m) \\
\tanh\big(\beta m + h\beta\big) &= m
\end{aligned} \tag{A.4}$$