Mean Field Ising and Potts Models: Tensor and Multiplex Variants

Project Report by

Sanchayan Bhowal

Advised by
Prof. Bhaswar Bhattacharya and *Prof. Somabha Mukherjee*
Co-Advised by
Prof. Yogeshwaran D

Indian Statistical Institute,
Bangalore

Statistics and Mathematics Unit

September 17, 2024

Abstract

The Ising and Potts models are fundamental tools in statistical mechanics for studying phase transitions and critical phenomena. This master's project explores variants of these models, specifically the tensor Potts model and the multiplex Ising model, within the framework of mean field theory. The project aims to provide an information theoretic interpretation of the limit theorems for the magnetization vector in the tensor Potts model under the Curie-Weiss setting. Additionally, it aims to investigate the asymptotics and fluctuations of the partition function in the multiplex Ising model. Finally it discusses a sufficient condition for the joint consistency of the MPLE estimators of inverse temperatures.

1 Introduction

Spin glass models were introduced to study complex materials within condensed matter physics and statistical mechanics, particularly disordered magnetic systems. These models provide valuable insights into materials that deviate from idealized structures, such as crystals, by considering interactions among spins located at various nodes within a network.

In mean-field models, interactions occur over a complete graph which simplifies the analysis. While this approach allows for tractable mathematical formulations, real-world materials often interact through a lattice structure, where interactions are restricted to neighboring sites, making lattice models more accurate but less analytically tractable.

2 Tensor Curie Weiss Potts Model

2.1 Model Description

For integers $p \geq 2$ and $q \geq 2$, the p-tensor Potts model is a discrete probability distribution on the set $[q]^N$ (here and afterwards, for a positive integer m, we will use $[m]$ to denote the set $\{1, 2, \ldots, m\}$) for some positive integers q and N, given by:

$$\mathbb{P}_{\beta,h,N}(\boldsymbol{X}) := \frac{1}{q^N Z_N(\beta, h)} \exp\left(\beta \sum_{1\leq i_1,\ldots,i_p\leq N} J_{i_1,\ldots,i_p} \mathbb{1}_{X_{i_1}=\ldots=X_{i_p}} + h\sum_{i=1}^{N} \mathbb{1}_{X_i=1}\right) \quad (\boldsymbol{X} \in [q]^N), \tag{2.1}$$

where $\beta > 0$, $h \geq 0$ and $\boldsymbol{J} := ((J_{i_1,\ldots,i_p}))_{i_1,\ldots,i_p\in[N]}$ is a symmetric tensor. The p-tensor Curie-Weiss Potts model is obtained by taking $J_{i_1,\ldots,i_p} := N^{1-p}$ for all $(i_1, \ldots, i_p) \in [N]^p$, whence model (2.1) takes the form:

$$\mathbb{P}_{\beta,h,N}(\boldsymbol{X}) := \frac{1}{q^N Z_N(\beta, h)} \exp\left(\beta N \sum_{r=1}^{q} \bar{X}_{\cdot r}^{p} + Nh\bar{X}_{\cdot 1}\right) \quad (\boldsymbol{X} \in [q]^N) \tag{2.2}$$

where $\bar{X}_{\cdot r} := N^{-1} \sum_{i=1}^{N} X_{i,r}$ with $X_{i,r} := \mathbb{1}_{X_i = r}$. The variables p and q are called the *interaction order* and the *number of states/colors* of the Potts model. A sufficient statistic for the exponential family (2.2) is the empirical magnetization vector:

$$\bar{\boldsymbol{X}}_N := \left(\bar{X}_{\cdot 1}, \ldots, \bar{X}_{\cdot q}\right)^{\top} .$$

Note that $\bar{\boldsymbol{X}}_N$ is a probability vector, i. e. has non-negative entries adding to 1. The parameter space is given by,

$$\Theta := \{(\beta, h) : \beta > 0, h \geq 0\} = (0, \infty) \times [0, \infty) .$$

2.2 Asymptotics of the Magnetization Vector

In this section, we state the main results regarding the asymptotics of the magnetization vector which was derived in [1]. For this, we need a few definitions and notations. For $p, q \geq 2$ and $(\beta, h) \in \Theta$, the *negative free energy* function $H_{\beta,h} : \mathcal{P}_q \to \mathbb{R}$ is defined as:

$$H_{\beta,h}(\boldsymbol{t}) := \beta \sum_{r=1}^{q} t_r^p + h t_1 - \sum_{r=1}^{q} t_r \log t_r$$

where $\mathcal{P}_q$ denotes the set of all q-dimensional probability vectors. We start by showing that the magnetization vector concentrates around the set $\mathcal{M}_{\beta,h}$ of all global maximizers of the function $H_{\beta,h}$.

Theorem 2.1. *Let $\beta_N \to \beta$ and $h_N \to h$. Then, under $\mathbb{P}_{\beta_N, h_N, N}$, the empirical magnetization $\bar{\boldsymbol{X}}_N$ satisfies a large deviation principle with speed N and rate function $-H_{\beta,h} + \sup H_{\beta,h}$. Consequently, for a point $\boldsymbol{t} \in \mathbb{R}^q$ and a set $A \subseteq \mathbb{R}^q$, if we define $d(\boldsymbol{t}, A) := \inf_{\boldsymbol{a} \in A} \|\boldsymbol{t} - \boldsymbol{a}\|_2$, then for every $\varepsilon > 0$, there exists a constant $C_{q,\varepsilon} > 0$ depending only on q and ε, such that:*

$$\mathbb{P}_{\beta_N, h_N, N}\left(d(\bar{\boldsymbol{X}}_N, \mathcal{M}_{\beta,h}) \geq \varepsilon\right) \leq e^{-C_{q,\varepsilon} N}$$

for all large N.

Theorem 2.1 is proved in Appendix A[1]. It enables us to derive a law of large numbers of the magnetization vector towards the set $\mathcal{M}_{\beta,h}$ of global maximizers of $H_{\beta,h}$.

2.3 Information Theoretic interpretation

Entropy is a measure of randomness in a system. In the context of the Ising model at high temperatures, where the inverse temperature β is small, the magnetic moments at each vertex exhibit a higher degree of randomness, approaching a uniform distribution. Conversely, as the temperature approaches zero, the system becomes more ordered, with reduced chaos. This suggests the presence of a transition between these two behaviors. Interestingly, this phase transition also occurs at a critical value of β, denoted as β'. Define entropy of a discrete random variable Y distributed as $\mathbf{P}$,

$$\mathbf{H}[Y] = \mathbb{E}\left[\ln \frac{1}{\mathbf{P}(X)}\right].$$

Hence for tensor Curie Weiss Potts model,

$$\mathbf{H}_{\beta,h,N}[\boldsymbol{X}] = -\mathbb{E}_{\beta,h,N}\left[\beta N \|\bar{\boldsymbol{X}}_N\|_p^p + Nh\bar{X}_{\cdot 1}\right] + N\log q + \log Z_N(\beta,h).$$

Corollary 2.2. *Suppose $\beta, h \in \Theta$,*

- *if $H_{\beta,h}$ has unique maximizer $\boldsymbol{m}_*$,*

$$\lim_{N\to\infty} \frac{1}{N}\mathbf{H}_{\beta,h,N}[\boldsymbol{X}] = \mathbf{H}[Z],$$

where Z is a random variable which takes value in $[q]$ and with distribition $\boldsymbol{m}_*$.

- *if $H_{\beta,h}$ has multiple maximizers $\{\boldsymbol{m}_i\}_{i=1}^K$,*

$$\lim_{N\to\infty} \frac{1}{N}\mathbf{H}_{\beta,h,N}[\boldsymbol{X}] = \sum_{i=1}^K p_i \mathbf{H}[Z_i],$$

where Z_i is a random variable which takes value in $[q]$ and with distribition $\boldsymbol{m}_i$ for $i = 1, \ldots, K$ and $0 < p_i < 1$ and $\sum_{i=1}^K p_i = 1$.

Now, let us focus on the case where $h = 0$. For large N, the entropy undergoes a phase transition. At high temperatures, the magnetization vector $m_* = \left(\frac{1}{q}, \cdots, \frac{1}{q}\right)$, which implies that the partition function Z is uniformly distributed, resulting in maximum entropy. Writing $m_*(\beta) = x_{s(\beta)}$, as β increases, $s(\beta)$ tends to 1, causing $\mathbf{H}[Z]$ as a function of β to approach 0. Additionally, a discontinuous phase transition occurs at $\beta_c(p,q)$.

Even when $h > 0$, it is straightforward to verify that $\mathbf{H}[Z]$ is a decreasing function of β. Moreover, the phase transition becomes continuous for $h \geq \tilde{h}_{p,q}$.

3 Multiplex Ising Model

3.1 Model Description

To define this model, let $G_p = \mathcal{G}(N,p)$ and $G_q = \mathcal{G}(N,q)$ be two realizations of a directed Erdős-Rényi graph with loops, so for all $(i,j) \in [N]$ the directed edge (i,j) is present with probability p and q respectively independent of all other edges. We denote by $(\varepsilon_{i,j}^{(1)})$ and $(\varepsilon_{i,j}^{(2)})$ the indicator variable which equals 1 if the edge (i,j) is present in the graph G_p and

G_q respectively. That is, $\varepsilon^{(1)}_{i,j} \sim$ Bernoulli (p) and $\varepsilon^{(2)}_{i,j} \sim$ Bernoulli (q). We do not assume that $\varepsilon^{(1)}_{i,j}$ and $\varepsilon^{(2)}_{i,j}$ are independent. The joint distribution is given by,

$$p_{11} := \mathbb{P}(\varepsilon^{(1)}_{i,j} = 1, \varepsilon^{(2)}_{i,j} = 1) \quad p_{10} := \mathbb{P}(\varepsilon^{(1)}_{i,j} = 1, \varepsilon^{(2)}_{i,j} = 0)$$
$$p_{01} := \mathbb{P}(\varepsilon^{(1)}_{i,j} = 0, \varepsilon^{(2)}_{i,j} = 1) \quad p_{00} := \mathbb{P}(\varepsilon^{(1)}_{i,j} = 0, \varepsilon^{(2)}_{i,j} = 0)$$

Define, $\boldsymbol{p} = (p_{11}, p_{10}, p_{01}, p_{00})$ as the joint probability vector. We consider the set of graphs (G_p, G_q) as the multiplex network on an N vertices.

The Hamiltonian of the Ising model on the Multiplex network is a function $H := H_N : \{-1, +1\}^N \to \mathbb{R}$, which can be written as

$$H(\sigma) = \frac{-1}{2N}\left(\frac{\beta_1}{p}\sum_{i,j}^{N} \varepsilon^{(1)}_{i,j}\sigma_i\sigma_j + \frac{\beta_2}{q}\sum_{i,j}^{N} \varepsilon^{(2)}_{i,j}\sigma_i\sigma_j\right)$$

where $\beta_1, \beta_2 > 0$ are inverse temperatures associated to the graphs G_p and G_q respectively. We use $\boldsymbol{\beta} = (\beta_1, \beta_2)$ to denote the tuple. We define the corresponding Gibbs measure as:

$$\mu_{\boldsymbol{\beta},h}(\sigma) = \frac{1}{Z_N(\beta_1, \beta_2)} e^{-H(\sigma)} \tag{3.1}$$

where $Z_N(\beta_1, \beta_2)$ is the partition function:

$$Z_N(\beta_1, \beta_2) := \sum_{\sigma \in \{-1,+1\}^N} e^{-H(\sigma)}$$

3.2 Asymptotics of magnetization and partition function

Here we derive the law of large numbers type result for t-+he partition function of the multiplex Ising model.

Theorem 3.1. *Suppose* $0 < \beta_1 + \beta_2 < 1$*, then*

$$\frac{Z_N(\beta_1, \beta_2)}{\mathbb{E}Z_N(\beta_1, \beta_2)} \xrightarrow{p} 1$$

where $\frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}}{2} = \frac{(1-p)\beta_1^2}{8p} + \frac{(p_{11}-pq)\beta_1\beta_2}{4pq} + \frac{(1-q)\beta_2^2}{8q}$.

Theorem 3.2. *Suppose* $0 < \beta_1 + \beta_2 < 1$*, then*

$$\frac{\sqrt{N}}{\lambda_{\boldsymbol{\beta},\boldsymbol{p}}}\left(\frac{Z_N(\boldsymbol{\beta})}{\mathbb{E}Z_N(\boldsymbol{\beta})} - 1\right) \xrightarrow{d} \mathcal{N}(0, 1).$$

4 Inference in Multiplex Ising model

Let us now look at statistical estimation of Multiplex Ising models on other graphs.

To define this model, let G_1 and G_2 be two weighted random graphs without self loops. We denote the adjacency matrices of G_1 and G_2 by $A^{(1)}$ and $A^{(2)}$ respectively.

The Hamiltonian of the Ising model on the Multiplex network is a function $H := H_N : \{-1,+1\}^N \to \mathbb{R}$, which can be written as

$$H(\sigma) = \frac{-1}{2N}\left(\beta_1 \sigma^T A^{(1)} \sigma + \beta_2 \sigma^T A^{(2)} \sigma\right) \tag{4.1}$$

where $\beta_1, \beta_2 > 0$ are inverse temperatures associated to the graphs G_1 and G_2 respectively.

Before proceeding to estimation let us check the conditions for identifiable.

Theorem 4.1. *Consider the Gibbs measure,*

$$\mu_{\boldsymbol{\beta}}(\sigma) = \frac{1}{Z_N(\beta_1,\beta_2)} e^{-H(\sigma)}$$

The models $\mu_{\boldsymbol{\beta}}$ indexed by $\boldsymbol{\beta}$ is identifiable iff,

$$c_1 A^{(1)} \neq c_2 A^{(2)} \tag{4.2}$$

for all $c_1, c_2 \in \mathbb{R}$ such that $c_1 c_2 \neq 0$.

We will now introduce the pseudo-likelihood estimator.

Definition 4.1. For any $i \in [n]$ we have

$$\mu_{\boldsymbol{\beta}}(\sigma_i = 1 | \sigma_j, j \neq i) = \frac{e^{\beta_1 m_i(\sigma) + \beta_2 m_i'(\sigma)}}{e^{\beta_1 m_i(\sigma) + \beta_2 m_i'(\sigma)} + e^{-\beta_1 m_i(\sigma) - \beta_2 m_i'(\sigma)}},$$

where $m_i(\boldsymbol{x}) := \sum_{j=1}^N A_N^{(1)}(i,j) x_j$ and $m_i'(\boldsymbol{x}) := \sum_{j=1}^N A_N^{(2)}(i,j) x_j$. Define the pseudo-likelihood as the product of the one dimensional conditional distributions:

$$\prod_{i=1}^N \mu_{\boldsymbol{\beta}}(\sigma_i = \sigma_i | \sigma_j, j \neq i) = 2^{-N} \exp\left\{ \sum_{i=1}^N \Big(\beta_1 \sigma_i m_i(\sigma) + \beta_2 \sigma_i m_i'(\sigma) - \log\cosh(\beta_1 m_i(\sigma) + \beta_2 m_i'(\sigma)) \Big) \right\}$$

If the pseudo-likelihood attains maximum at $(\beta_1, \beta_2) \in \mathbb{R}^2$, denote it by $\boldsymbol{\beta} = (\hat{\beta}_1, \hat{\beta}_2)$. This is the pseudo-likelihood estimator for the parameter vector (β_1, β_2).

The coupling matrix A_N has non-negative entries and is completely known. We will also assume the following two conditions

$$\max_{i\in[n]} \sum_{j=1}^N A_{i,j}^{(1)} \leq \gamma, \quad \max_{i\in[n]} \sum_{j=1}^N A_{i,j}^{(2)} \leq \gamma, \tag{4.3}$$

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{i,j=1}^{N} A^{(1)}_{i,j} > 0 \quad \liminf_{N\to\infty} \frac{1}{N}\sum_{i,j=1}^{N} A^{(2)}_{i,j} > 0. \tag{4.4}$$

For notational convenience, we borrow the following notations from [5],

$$\overline{\boldsymbol{x}} := \sum_{i=1}^{N} x_i \tag{4.5}$$

$$\overline{\boldsymbol{xy}} := \sum_{i=1}^{N} x_i y_j \tag{4.6}$$

where $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$.

Theorem 4.2. *Suppose σ is an observation from the multiplex Ising model (4.1), where the coupling matrices $A^{(1)}$ and $A^{(2)}$ satisfy (4.3) and (4.4), and $\boldsymbol{\beta} \in \Theta$. Set*

$$T_N(\sigma) := \frac{1}{N^2}\left(\overline{m(\sigma)^2} \cdot \overline{m'(\sigma)^2} - \overline{m(\sigma)m'(\sigma)}^2\right),$$

where,

$$m(\sigma) = A^{(1)}\sigma \quad m'(\sigma) = A^{(2)}\sigma$$

Further, if $\frac{1}{T_N(\sigma)} = o_p(\sqrt{N})$, then

$$||\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}|| = O_p\left(\frac{1}{\sqrt{N}T_N(\sigma)}\right).$$

In particular, $\hat{\boldsymbol{\beta}}_N$ is jointly consistent for $\boldsymbol{\beta}$ in Θ.

5 Acknowledgements

I would like to express my sincere gratitude to my advisors, Professor Bhaswar Bhattacharya and Professor Somabha Mukherjee, for their invaluable ideas and inputs throughout the course of this project. I am also thankful to Professor Yogeshwaran for the helpful discussions and insightful comments. A special thanks to Pratichi for suggesting a clean argument for the identifiability of the model.

6 Proof of Theorems

6.1 Proof of Corollary 2.2

Proof. Let $H_{\beta,h}$ has one maximizer $\boldsymbol{m}_* = (m_1, \ldots, m_q)$. From the proof of central limit theorem in [1],

$$\frac{1}{N}\log Z_N(\beta, h) \to H_{\beta,h}(\boldsymbol{m}_*)$$

In the proof of central limit theorem in [1], we also show that $\bar{\boldsymbol{X}}_N$ is uniformly integrable. Hence, putting these two things together,

$$-\mathbb{E}_{\beta,h,N}\left[\beta N\|\bar{\boldsymbol{X}}_N\|_p^p + Nh\bar{X}_{\cdot 1}\right] + N\log q + \log Z_N(\beta,h) \to -\sum_{i=1} m_i \log m_i.$$

If $H_{\beta,h}$ has more than one maximizer then, we need to break the expectation into conditionals i.e.,

$$\mathbb{E}_{\beta,h,N}\left[\beta N\|\bar{\boldsymbol{X}}_N\|_p^p + Nh\bar{X}_{\cdot 1}\right] = \sum_{i=1}^{q} \mathbb{E}_{\beta,h,N}\left[\beta N\|\bar{\boldsymbol{X}}_N\|_p^p + Nh\bar{X}_{\cdot 1} \mid \bar{\boldsymbol{X}}_N \in B(\boldsymbol{m}_i,\varepsilon)\right]\mathbb{P}\left(\bar{\boldsymbol{X}}_N \in B(\boldsymbol{m}_i,\varepsilon)\right),$$

where $\varepsilon > 0$ is small enough such that $B(\boldsymbol{m}_i,\varepsilon)$ are disjoint. Now, by Lemma A.8 of [2],

$$\mathbb{P}\left(\bar{\boldsymbol{X}}_N \in B(\boldsymbol{m}_i,\varepsilon)\right) \to p_i.$$

and,

$$\frac{1}{N}\log Z_N(\beta,h) \to H_{\beta,h}(\boldsymbol{m}_1).$$

The above equations complete the proof. □

6.2 Proof of Theorem 3.1

In this section we will prove the law of large numbers type result for partition function. The proof is primarily motivated from [5].

Define,

$$F(\boldsymbol{p},x,y) := \log(p_{00} + p_{10}e^x + p_{01}e^y + p_{11}e^{x+y}).$$

In order to get rid of subscripts let us rename $a := p_{11}$, hence we get,

$$F(\boldsymbol{p},x,y) = \log(1-p-q+a+(p-a)e^x+(q-a)e^y+ae^{x+y}).$$

Remark 6.1. Upon Taylor expansion we get that,

$$\begin{aligned}F(\boldsymbol{p},x,y) =& px + qy + \frac{p(1-p)}{2}x^2 + (a-pq)xy + \frac{q(1-q)}{2}y^2\\ &\frac{2p^3-3p^2+p}{6}x^3 + \frac{2p^2q-pq-2ap+a}{2}x^2y + \frac{2q^2p-pq-2aq+a}{2}xy^2 + \frac{2q^3-3q^2+q}{6}y^3\\ &\frac{-6p^4+12p^3-7p^2+p}{24}x^4 + \frac{-6p^3q+6p^2q+6ap^2-6ap-pq+a}{6}x^3y+\\ &\frac{-2a^2+8apq-2ap-2aq+a-p^2\left(6q^2-2q\right)+p\left(2q^2-q\right)}{4}x^2y^2+\\ &\frac{-6q^3p+6q^2p+6aq^2-6aq-pq+a}{6}xy^3 + \frac{-6q^4+12q^3-7q^2+q}{24}y^4 + \dots\end{aligned}$$

Remark 6.2. The following results follow easily using the above Taylor expansion,

$$\frac{F(\boldsymbol{p},x,y)+F(\boldsymbol{p},-x,-y)}{2} = \frac{p(1-p)}{2}x^2+(a-pq)xy+\frac{q(1-q)}{2}y^2+O((p+q+a)^2(x+y)^4) \tag{6.1}$$

$$\frac{F(\boldsymbol{p},x,y)-F(\boldsymbol{p},-x,-y)}{2} = px+qy+O((p+q+a)(x+y)^3) \tag{6.2}$$

Lemma 6.1.

$$\mathbb{E}[e^{-H(\sigma)}] = \exp\left(\frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}}{2}+\frac{\beta_1+\beta_2}{2N}\overline{\sigma}^2+O\left(\frac{1}{N^2}\right)\left(\frac{\overline{\sigma}^2}{N}+1\right)\right),$$

where the O-term is uniform in $\sigma \in \{-1,1\}^N$.

Proof. Define $\gamma_1=\frac{\beta_1}{2Np}, \gamma_2=\frac{\beta_2}{2Nq}$. We need to study

$$\begin{aligned}\mathbb{E}e^{-H(\sigma)} &= \mathbb{E}\left[e^{\gamma_1\sum_{i,j=1}^N \varepsilon^{(1)}_{i,j}\sigma_i\sigma_j+\gamma_2\sum_{i,j=1}^N\varepsilon^{(2)}_{i,j}\sigma_i\sigma_j}\right]\\ &= \prod_{i,j=1}^N \mathbb{E}\left[e^{\gamma_1\varepsilon^{(1)}_{i,j}\sigma_i\sigma_j+\gamma_2\varepsilon^{(2)}_{i,j}\sigma_i\sigma_j}\right]\\ &= \prod_{i,j=1}^N F(\boldsymbol{p},\gamma_1\sigma_i\sigma_j,\gamma_2\sigma_i\sigma_j).\end{aligned}$$

Defining $f(x)=f(x;\boldsymbol{p},\gamma_1,\gamma_2)=\log\left(p_{00}+p_{10}e^{\gamma_1 x}+p_{01}e^{\gamma_2 x}+ae^{x(\gamma_1+\gamma_2)}\right)$, we can write

$$\mathbb{E}e^{H(\sigma)} = \exp\left(\sum_{i,j=1}^N f\left(\sigma_i\sigma_j\right)\right).$$

Observe that the argument of f, i.e. $\sigma_i\sigma_j$, can only take the two values ± 1. For these values we can linearize the function f, i.e. we can write

$$f(x)=a_0+a_1x, \quad x\in\{-1,+1\},$$

where $a_0=a_0(p,\gamma)$ and $a_1=a_1(p,\gamma)$ are given by

$$\begin{aligned}a_0 &= \frac{f(1)+f(-1)}{2} = \frac{F(\boldsymbol{p},\gamma_1,\gamma_2)+F(\boldsymbol{p},-\gamma_1,-\gamma_2)}{2}\\ a_1 &= \frac{f(1)-f(-1)}{2} = \frac{F(\boldsymbol{p},\gamma_1,\gamma_2)-F(\boldsymbol{p},-\gamma_1,-\gamma_2)}{2}\end{aligned}$$

From here we obtain

$$\mathbb{E}e^{H(\sigma)} = \exp\left(\sum_{i,j=1}^N f\left(\sigma_i\sigma_j\right)\right) = \exp\left(\sum_{i,j=1}^N\left(a_0+a_1\sigma_i\sigma_j\right)\right) = \exp\left(N^2a_0+a_1\overline{\sigma}^2\right)$$

Hence, by (6.1) and using $\gamma_1 = \frac{\beta_1}{2Np}, \gamma_2 = \frac{\beta_2}{2Nq}$ we get that,

$$N^2 a_0 = \frac{F(\boldsymbol{p}, x, y) + F(\boldsymbol{p}, -x, -y)}{2} \quad = \frac{(1-p)\beta_1^2}{8p} + \frac{(a-pq)\beta_1\beta_2}{4pq} + \frac{(1-q)\beta_2^2}{8q} + O(\frac{1}{N^2})$$

$$\overline{\sigma}^2 a_1 = \frac{\beta_1 + \beta_2}{2N}\overline{\sigma}^2 + O\left(\frac{\overline{\sigma}^2}{N^3}\right).$$

Taking everything together, we obtain the required statement. □

Lemma 6.2.

$$\mathbb{E}[e^{-H(\sigma)}e^{-H(\tau)}] = \exp\left(\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}\left(\frac{\overline{\sigma\tau}^2}{N^2} + 1\right) + \frac{\beta_1+\beta_2}{2N}(\overline{\sigma}^2 + \overline{\tau}^2) + O\left(\frac{1}{N^2}\right)\left(\frac{\overline{\sigma}^2 + \overline{\tau}^2}{N} + 1\right)\right),$$

where the O-term is uniform in $\sigma \in \{-1,1\}^N$.

Proof.

$$\begin{aligned}
\mathbb{E}e^{-H(\sigma)}e^{-H(\tau)} &= \mathbb{E}\left[e^{\gamma_1 \sum_{i,j=1}^N \varepsilon^{(1)}_{i,j}(\sigma_i\sigma_j + \tau_i\tau_j) + \gamma_2 \sum_{i,j=1}^N \varepsilon^{(2)}_{i,j}(\sigma_i\sigma_j + \tau_i\tau_j)}\right] \\
&= \prod_{i,j=1}^N \mathbb{E}\left[e^{\gamma_1 \varepsilon^{(1)}_{i,j}(\sigma_i\sigma_j + \tau_i\tau_j) + \gamma_2 \varepsilon^{(2)}_{i,j}(\sigma_i\sigma_j + \tau_i\tau_j)}\right] \\
&= \exp\left(\sum_{i,j=1}^N f\left(\sigma_i\sigma_j + \tau_i\tau_j\right)\right).
\end{aligned}$$

Note that $f\left(\sigma_i\sigma_j + \tau_i\tau_j\right)$ is a function of arguments $x_1 := \sigma_i\sigma_j$ and $x_2 := \tau_i\tau_j$. Now, $x_1, x_2 \in -1, 1$. Hence, we can write f as follows,

$$f(x_1 + x_2) = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2.$$

Solving the equations we get,

$$\begin{aligned}
b_0 = b_{12} &= \frac{f(2) + f(-2)}{4} = \frac{F(\boldsymbol{p}, 2\gamma_1, 2\gamma_2) + F(\boldsymbol{p}, -2\gamma_1, -2\gamma_2)}{4} \\
b_1 = b_2 &= \frac{f(2) - f(-2)}{4} = \frac{F(\boldsymbol{p}, 2\gamma_1, 2\gamma_2) - F(\boldsymbol{p}, -2\gamma_1, -2\gamma_2)}{4}.
\end{aligned}$$

From here we obtain

$$\mathbb{E}e^{-H(\sigma)}e^{-H(\tau)} = \exp\left(N^2 b_0 + b_1\overline{\sigma}^2 + b_2\overline{\tau}^2 + b_{12}\overline{\sigma\tau}^2\right)$$

Hence, by (6.1) and (6.2),

$$\begin{aligned}
N^2 b_0 &= \frac{(1-p)\beta_1^2}{4p} + \frac{(a-pq)\beta_1\beta_2}{2pq} + \frac{(1-q)\beta_2^2}{4q} + O\left(\frac{1}{N^2}\right) \\
\overline{\sigma}^2 b_1 &= \frac{\beta_1+\beta_2}{2N}\overline{\sigma}^2 + O\left(\frac{\overline{\sigma}^2}{N^3}\right) \\
\overline{\tau}^2 b_2 &= \frac{\beta_1+\beta_2}{2N}\overline{\tau}^2 + O\left(\frac{\overline{\tau}^2}{N^3}\right) \\
\overline{\sigma\tau}^2 b_{12} &= \frac{\overline{\sigma\tau}^2}{N^2}\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}} + O\left(\frac{1}{N^2}\right).
\end{aligned}$$

Taking everything together, we obtain the required statement. □

Lemma 6.3. *Suppose* $0 < \beta_1 + \beta_2 < 1$, *then*

$$\mathbb{E}Z_N(\beta_1, \beta_2) \sim \exp\left(\frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}}{2}\right) 2^N \mathbb{E}[e^{\frac{\beta_1+\beta_2}{2}\xi^2}]$$

Proof. Note that $|\overline{\sigma}| \le N$. Hence, by Lemma 6.1,

$$\mathbb{E}[e^{-H(\sigma)}] \sim \exp\left(\frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}}{2} + \frac{\beta_1+\beta_2}{2N}\overline{\sigma}^2\right).$$

Therefore, we can write $\mathbb{E}Z_N(\beta_1, \beta_2)$ as,

$$\begin{aligned}
\mathbb{E}Z_N(\beta_1, \beta_2) &= \sum_{\sigma \in \{-1,1\}^N} \mathbb{E}e^{-H(\sigma)} \\
&\sim \sum_{\sigma \in \{-1,1\}^N} \exp\left(\frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}}{2} + \frac{\beta_1+\beta_2}{2N}\overline{\sigma}^2\right). \\
&\sim \exp\left(\frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}}{2}\right) \sum_{\sigma \in \{-1,1\}^N} e^{\frac{\beta_1+\beta_2}{2N}\overline{\sigma}^2}. \\
&\sim \exp\left(\frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}}{2}\right) 2^N \mathbb{E}_\zeta\left[e^{\frac{\beta_1+\beta_2}{2N}(\sum_{i=1}^N \zeta_i)^2}\right],
\end{aligned}$$

where $\zeta_i \sim Uniform\{-1, 1\}$ are independently distributed. Now, by de Moivre Laplace CLT,

$$\frac{\sum_{i=1}^N \zeta_i}{\sqrt{N}} \xrightarrow{d} \xi,$$

where $\xi \sim \mathcal{N}(0, 1)$. Hence, continuous mapping Theorem,

$$\sigma_N = e^{\frac{\beta_1+\beta_2}{2N}(\sum_{i=1}^N \zeta_i)^2} \xrightarrow{d} g(\xi)e^{\frac{\beta_1+\beta_2}{2}\xi^2}.$$

In [4], they prove uniform integrability of such σ_N. Hence, we get

$$\mathbb{E}_\zeta\left[e^{\frac{\beta_1+\beta_2}{2N}(\sum_{i=1}^N \zeta_i)^2}\right] \to \mathbb{E}\left[e^{\frac{\beta_1+\beta_2}{2}\xi^2}\right].$$

This proves our claim. □

Proof of Theorem 3.1. Now, observe that,

$$\begin{aligned}
\frac{1}{4^N}\mathbb{E}[Z_N(\beta_1, \beta_2)^2] &= \frac{1}{4^N}\sum_{\sigma,\tau} \exp\left(\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}\left(\frac{\overline{\sigma\tau}^2}{N^2} + 1\right) + \frac{\beta_1+\beta_2}{2N}(\overline{\sigma}^2 + \overline{\tau}^2) + O\left(\frac{1}{N^2}\right)\left(\frac{\overline{\sigma}^2 + \overline{\tau}^2}{N} + 1\right)\right) \\
&= \mathbb{E}_0\left[\exp\left(\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}\left(\frac{\overline{\sigma\tau}^2}{N^2} + 1\right) + \frac{\beta_1+\beta_2}{2N}(\overline{\sigma}^2 + \overline{\tau}^2) + O\left(\frac{1}{N}\right)\right)\right]
\end{aligned}$$

where $\mathbb{E}_0$ is with respect to random variables σ, τ which are i.i.d Rademacher with mean 0. Note that $\sigma\tau$ is independent of σ and τ. By dominated convergence theorem,

$$\mathbb{E}_0 \exp\left(\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}\left(\frac{\overline{\sigma\tau}^2}{N^2}\right)\right) \to 1.$$

Also, by uniform integrability of σ_N as proved in [4],

$$\begin{aligned}\mathbb{E}_0\left[\exp\left(\frac{\beta_1+\beta_2}{2N}(\overline{\sigma}^2+\overline{\tau}^2)\right)\right] &= \left(\mathbb{E}_\zeta\left[e^{\frac{\beta_1+\beta_2}{2N}(\sum_{i=1}^N \zeta_i)^2}\right]\right)^2 \\ &\to \mathbb{E}\left[e^{\frac{\beta_1+\beta_2}{2}\xi^2}\right]^2.\end{aligned}$$

where ζ is Rademacher with mean 0. Hence,

$$\mathbb{E}[Z_N(\beta_1,\beta_2)^2] \sim \exp\left(\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}\right) 4^N \mathbb{E}[e^{\frac{\beta_1+\beta_2}{2}\xi^2}]^2.$$

Therefore using the above computations we conclude that,

$$\begin{aligned}&\frac{\mathbb{E}Z_N(\beta_1,\beta_2)^2}{(\mathbb{E}Z_N(\beta_1,\beta_2))^2} \to 1 \\ \Rightarrow &\frac{\mathrm{Var}\left(Z_N(\beta_1,\beta_2)\right)}{(\mathbb{E}Z_N(\beta_1,\beta_2))^2} \to 0 \\ \Rightarrow &\frac{Z_N(\beta_1,\beta_2)}{\mathbb{E}Z_N(\beta_1,\beta_2)} \xrightarrow{p} 1\end{aligned}$$

for all $0 < \beta_1 + \beta_2 < 1$. □

6.3 Proof of Theorem 3.2

We define the following quantity $X(\sigma)$ as

$$X(\sigma) := e^{-H(\sigma)} - \mathbb{E}e^{-H(\sigma)}\left(1 - H(\sigma) - \frac{\beta_1+\beta_2}{2N}\overline{\sigma}^2\right). \tag{6.3}$$

Remark 6.3. Note that $\mathbb{E}X(\sigma) = 0$ for each $\sigma \in \{-1,1\}^N$.

If we show that $\sum_\sigma X(\sigma)$ has fewer fluctuations than we can formally write,

$$Z_N(\boldsymbol{\beta}) = \sum_\sigma e^{-H(\sigma)} \approx \sum_\sigma \mathbb{E}\left[e^{-H(\sigma)}\right]\left(1 - H(\sigma) - \frac{\beta_1+\beta_2}{2N}\overline{\sigma}^2\right).$$

We will follow techniques introduced in [6]. Hence, let us now move to deriving an expression for $\mathrm{Var}\sum_\sigma X(\sigma)$.

$$\mathrm{Var}\sum X(\sigma) = \sum_{\sigma,\tau}\mathrm{Cov}(X(\sigma), X(\tau))$$

Now, we will break the $\mathrm{Cov}(X(\sigma), X(\tau))$ as follows,

$$\begin{aligned}\mathrm{Cov}(X(\sigma), X(\tau)) = &\mathrm{Cov}(\exp(-H(\sigma)), \exp(-H(\tau))) \\ &+ \mathbb{E}e^{-H(\sigma)}\mathbb{E}e^{-H(\tau)}\,\mathrm{Cov}(H(\sigma), H(\tau)) \\ &+ \mathbb{E}e^{-H(\sigma)}\,\mathrm{Cov}\left(e^{-H(\tau)}, H(\sigma)\right) \\ &+ Ee^{-H(\tau)}\,\mathrm{Cov}\left(e^{-H(\sigma)}, H(\tau)\right)\end{aligned}$$

Lemma 6.4.

$$\mathrm{Cov}\left(e^{-H(\sigma)}, e^{-H(\tau)}\right) = \mathbb{E}e^{-H(\sigma)}\mathbb{E}e^{-H(\tau)}\left(e^{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}\frac{\overline{\sigma\tau}^2}{N^2}+O\left(\frac{1}{N^2}\right)\left(\frac{\overline{\sigma}^2+\overline{\tau}^2}{N}+1\right)} - 1\right)$$

Proof. Follows from Lemma 6.2 trivially. □

Lemma 6.5.

$$\mathrm{Cov}(H(\sigma), H(\tau)) = \frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}\overline{\sigma\tau}^2}{N^2}.$$

Proof. Let us first break down the sum in four parts,

$$\begin{aligned}\mathrm{Cov}(H(\sigma), H(\tau)) &= \mathrm{Cov}\left(\gamma_1\sum_{i,j}\sigma_i\sigma_j\varepsilon^{(1)}_{i,j} + \gamma_2\sum_{i,j}\sigma_i\sigma_j\varepsilon^{(2)}_{i,j}, \gamma_1\sum_{i,j}\tau_i\tau_j\varepsilon^{(1)}_{i,j} + \gamma_2\sum_{i,j}\tau_i\tau_j\varepsilon^{(2)}_{i,j}\right) \\ &= \mathrm{Cov}\left(\gamma_1\sum_{i,j}\sigma_i\sigma_j\varepsilon^{(1)}_{i,j}, \gamma_1\sum_{i,j}\tau_i\tau_j\varepsilon^{(1)}_{i,j}\right) + \mathrm{Cov}\left(\gamma_2\sum_{i,j}\sigma_i\sigma_j\varepsilon^{(2)}_{i,j}, \gamma_2\sum_{i,j}\tau_i\tau_j\varepsilon^{(2)}_{i,j}\right) \\ &+ \mathrm{Cov}\left(\gamma_1\sum_{i,j}\sigma_i\sigma_j\varepsilon^{(1)}_{i,j}, \gamma_2\sum_{i,j}\tau_i\tau_j\varepsilon^{(2)}_{i,j}\right) + \mathrm{Cov}\left(\gamma_2\sum_{i,j}\sigma_i\sigma_j\varepsilon^{(2)}_{i,j}, \gamma_1\sum_{i,j}\tau_i\tau_j\varepsilon'_{i,j}\right).\end{aligned}$$

Consider each term separately,

$$\mathrm{Cov}\left(\gamma_1\sum_{i,j}\sigma_i\sigma_j\varepsilon^{(1)}_{i,j}, \gamma_1\sum_{i,j}\tau_i\tau_j\varepsilon^{(1)}_{i,j}\right) = \gamma_1^2\left(p^2\sum_{(i,j)\neq(i',j')}\sigma_i\sigma_j\tau_{i'}\tau_{j'} + p\sum_{(i,j)}\sigma_i\sigma_j\tau_i\tau_j - p^2\overline{\sigma}^2\overline{\tau}^2\right)$$

$$\mathrm{Cov}\left(\gamma_2\sum_{i,j}\sigma_i\sigma_j\varepsilon^{(2)}_{i,j}, \gamma_2\sum_{i,j}\tau_i\tau_j\varepsilon^{(2)}_{i,j}\right) = \gamma_2^2\left(q^2\sum_{(i,j)\neq(i',j')}\sigma_i\sigma_j\tau_{i'}\tau_{j'} + q\sum_{(i,j)}\sigma_i\sigma_j\tau_i\tau_j - q^2\overline{\sigma}^2\overline{\tau}^2\right)$$

$$\mathrm{Cov}\left(\gamma_1\sum_{i,j}\sigma_i\sigma_j\varepsilon^{(1)}_{i,j}, \gamma_2\sum_{i,j}\tau_i\tau_j\varepsilon^{(2)}_{i,j}\right) = \gamma_1\gamma_2\left(pq\sum_{(i,j)\neq(i',j')}\sigma_i\sigma_j\tau_{i'}\tau_{j'} + a\sum_{(i,j)}\sigma_i\sigma_j\tau_i\tau_j - pq\overline{\sigma}^2\overline{\tau}^2\right).$$

Putting everything together we get that,

$$\begin{aligned}\mathrm{Cov}(H(\sigma), H(\tau)) &= \gamma_1^2(1-p)p\overline{\sigma\tau}^2 + \gamma_2^2(1-q)q\overline{\sigma\tau}^2 + 2\gamma_1\gamma_2(a-pq)\overline{\sigma\tau}^2 \\ &= \left[\frac{\beta_1^2}{4N^2}\frac{(1-p)}{p} + \frac{\beta_2^2}{4N^2}\frac{(1-q)}{q} + 2\frac{\beta_1\beta_2}{4N^2}\left(\frac{a-pq}{pq}\right)\right]\overline{\sigma\tau}^2 \\ &= \frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}}{N^2}\overline{\sigma\tau}^2\end{aligned}$$

$\square$

Lemma 6.6.

$$\mathbb{E}e^{-H(\sigma)}\operatorname{Cov}\left(e^{-H(\tau)}, H(\sigma)\right) = -\mathbb{E}e^{-H(\sigma)}Ee^{-H(\tau)}\left[\frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}\overline{\sigma\tau}^2}{N^2} + O\left(\frac{\overline{\sigma}^2}{N^3}\right)\right]$$

$$Ee^{-H(\tau)}\operatorname{Cov}\left(e^{-H(\sigma)}, H(\tau)\right) = -\mathbb{E}e^{-H(\sigma)}Ee^{-H(\tau)}\left[\frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}\overline{\sigma\tau}^2}{N^2} + O\left(\frac{\overline{\tau}^2}{N^3}\right)\right]$$

Putting everything together we get the following,

Corollary 6.1.

$$\operatorname{Cov}(X(\sigma), X(\tau)) = \mathbb{E}e^{-H(\sigma)}Ee^{-H(\tau)}\left(e^{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}\frac{\overline{\sigma\tau}^2}{N^2} + O\left(\frac{1}{N^2}\right)\left(\frac{\overline{\sigma}^2+\overline{\tau}^2}{N}+1\right)} - 1 - \frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}\overline{\sigma\tau}^2}{N^2} + O\left(\frac{\overline{\sigma}^2+\overline{\tau}^2}{N^3}\right)\right)$$

Lemma 6.7. $\operatorname{Var}\left(\sum_\sigma X(\sigma)\right) = \mathbb{E}\left[Z_N(\beta)\right]^2\left(O\left(\frac{1}{N^2}\right) + o\left(\frac{1}{N}\right)\right)$

Proof. Define,

$$S_N := \left\{(\sigma,\tau) : \overline{\sigma}^2 \le N^{6/5}, \overline{\tau}^2 \le N^{6/5}, \overline{\sigma\tau}^2 \le N^{6/5}\right\}$$

$$\operatorname{Var}\left(\sum_\sigma X(\sigma)\right) = \sum_{(\sigma,\tau)\in S_N}\operatorname{Cov}(X(\sigma), X(\tau)) + \sum_{(\sigma,\tau)\in S_N^c}\operatorname{Cov}(X(\sigma), X(\tau))$$

Consider the first sum,

$$\begin{aligned}
\sum_{(\sigma,\tau)\in S_N}\operatorname{Cov}(X(\sigma), X(\tau)) &= \sum_{(\sigma,\tau)\in S_N}\mathbb{E}e^{-H(\sigma)}\mathbb{E}e^{-H(\tau)}\left(O\left(\frac{1}{N^2p^3}\right) + O\left(\left(\frac{\overline{\sigma\tau}^2}{N^2}c_{\beta,p,\tau}\right)^2\right) + O\left(\frac{|\sigma^2+|\,\tau|^2}{N^3}\right.\right. \\
&\le \sum_{(\sigma,\tau)\in S_N}\mathbb{E}e^{-H(\sigma)}\mathbb{E}e^{-H(\tau)}\left(O\left(\frac{1}{N^2p^3}\right) + O\left(\frac{N^{2/5}}{N^2}\right)\right) \\
&= \sum_{(\sigma,\tau)\in S_N}\mathbb{E}e^{-H(\sigma)}\mathbb{E}e^{-H(\tau)}\left(O\left(\frac{1}{N^2}\right) + o\left(\frac{1}{N}\right)\right)
\end{aligned}$$

Let us now move to the second sum,

$$\begin{aligned}
V_N(k,l,m) &:= \{(\sigma,\tau)|\overline{\sigma} = k, \overline{\tau} = l, \overline{\sigma\tau} = m\} \\
v_N(k,l,m) &= |V_N(k,l,m)|
\end{aligned}$$

Taking σ,τ to be Uniform $\{-1,1\}^N$. Therefore, covariance matrix of $(\sigma,\tau,\sigma\tau)$ is identity. Now, it follows from Local Central Limit Theorem that,

$$v_N(k,l,m) \le C2^{2N}N^{-3/2}e^{-\frac{k^2+l^2+m^2}{2N}}$$

Now by triangle inequality,

$$
\begin{aligned}
&\left|\sum_{(\sigma,\tau)\in V_N(k,l,m)} \operatorname{Cov}(X(\sigma),X(\tau))\right| \\
&\leq \sum_{(\sigma,\tau)\in V_N(k,l,m)} \mathbb{E}e^{-H(\sigma)}\mathbb{E}e^{-H(\tau)}\left(e^{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}\frac{\overline{\sigma\tau}^2}{N^2}+O\left(\frac{1}{N^2}\right)\left(\frac{\overline{\sigma}^2+\overline{\tau}^2}{N}+1\right)}+1+\frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}\overline{\sigma\tau}^2}{N^2}+O\left(\frac{\overline{\sigma}^2+\overline{\tau}^2}{N^3}\right)\right) \\
&= \sum_{(\sigma,\tau)\in V_N(k,l,m)} \mathbb{E}e^{-H(\sigma)}\mathbb{E}e^{-H(\tau)}\left(e^{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}\frac{m^2}{N^2}+O\left(\frac{1}{N^2}\right)\left(\frac{k^2+l^2}{N}+1\right)}+1+\frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}m^2}{N^2}+O\left(\frac{k^2+l^2}{N^3}\right)\right) \\
&\lesssim 2^{2N}N^{-3/2}e^{-\frac{k^2+l^2+m^2}{2N}}\ e^{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}+(\beta_1+\beta_2)\frac{k^2+l^2}{2N}+O\left(\frac{1}{N^2}\right)\left(\frac{\overline{\sigma}^2+\overline{\tau}^2}{N}+1\right)}\ e^{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}\frac{m^2}{N^2}+O\left(\frac{1}{N^2}\right)\left(\frac{k^2+l^2}{N}+1\right)}\ e^{\frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}m^2}{N^2}+O\left(\frac{k^2+l^2}{N^3}\right)}
\end{aligned}
$$

Now, $\beta_1+\beta_2<1$, hence we can get a δ independent of k,l,m such that,

$$
\left|\sum_{(\sigma,\tau)\in V_N(k,l,m)} \operatorname{Cov}(X(\sigma),X(\tau))\right| \leq Ce^{-\delta\frac{k^2+l^2+m^2}{N}}.
$$

Therefore, by Theorem 6.3,

$$
\begin{aligned}
\sum_{\sigma,\tau\in S_N^C} \operatorname{Cov}(X(\sigma),X(\tau)) &\leq 2^{2N}CN^{-3/2}e^{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}}\sum_{\substack{(k,l,m)\\ k,l,m>N^{6/5}}} e^{-\delta\frac{k^2+l^2+m^2}{N}} \\
&\leq 2^{2N}e^{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}}e^{-\delta'(Np)^{1/5}} \\
&\lesssim \mathbb{E}\left[Z_N(\beta)\right]^2 e^{-\delta'(Np)^{1/5}} \\
&= \mathbb{E}\left[Z_N(\beta)\right]^2 o\left(\frac{1}{N}\right),
\end{aligned}
$$

where the second inequality follows from Riemann sum approximation. □

Proof of Theorem 3.2. Define,

$$
W_N(\beta) := \sum_{\sigma}\mathbb{E}e^{-H(\sigma)}\left(-H(\sigma)-\frac{\beta_1+\beta_2}{2N}\overline{\sigma}^2\right)
$$

$$
\begin{aligned}
\operatorname{Var} W_N(\beta) &= \sum_{\sigma,\tau}\mathbb{E}e^{-H(\sigma)}\mathbb{E}e^{-H(\tau)}\operatorname{Cov}(H(\sigma),H(\tau)) \\
&= \sum_{\sigma,\tau} e^{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}+(\beta_1+\beta_2)\frac{\overline{\sigma}^2+\overline{\tau}^2}{2N}+o(1)}\frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}\overline{\sigma\tau}^2}{N^2} \\
&\sim e^{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}}\frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}}{N}\sum_{\sigma,\tau}\frac{\overline{\sigma\tau}^2}{N}e^{(\beta_1+\beta_2)\frac{\overline{\sigma}^2+\overline{\tau}^2}{2N}}
\end{aligned}
$$

Now, considering σ, τ to be independently distributed as Uniform $\{-1,1\}^N$. Hence, $(\frac{\overline{\sigma}}{\sqrt{N}}, \frac{\overline{\tau}}{\sqrt{N}}, \frac{\overline{\sigma\tau}}{\sqrt{N}})$ converges in distribution to three-dimensional standard Normal distribution. Again from [4], the random variables $Y_N = \frac{\overline{\sigma\tau}^2}{N} e^{(\beta_1+\beta_2)\frac{\overline{\sigma}^2+\overline{\tau}^2}{2N}}$ are uniformly integrable. So,

$$\lim_{N\to\infty} \frac{1}{2^{2N}} \sum_{\sigma,\tau} \frac{\overline{\sigma\tau}^2}{N} e^{(\beta_1+\beta_2)\frac{\overline{\sigma}^2+\overline{\tau}^2}{2N}} = \mathbb{E}_{\xi_1,\xi_2,\xi_3}\left(\xi_3^2 \exp\left(e^{(\beta_1+\beta_2)\frac{\xi_1^2+\xi_2^2}{2}}\right)\right) = \frac{1}{1-\beta_1-\beta_2},$$

where (ξ_1, ξ_2, ξ_3) is distribution as three-dimensional standard Normal distribution. Using Lemma 6.3,

$$\begin{aligned} \operatorname{Var} W_N(\beta) &\sim e^{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}} \frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}}{N} 4^N \sum_{\sigma,r} \frac{1}{2^{2N}N} e^{-\frac{\beta_1+\beta_2}{2}\frac{\overline{\sigma}^2+\overline{\tau}^2}{N}} \\ &\sim e^{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}} \frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}}{N} 4^N \frac{1}{1-\beta_1-\beta_2} \\ &\sim \frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}}{N} \left(\mathbb{E} Z_N(\beta)\right)^2. \end{aligned}$$

Now, define,

$$\sigma_{ij} = \sum_{\sigma} \left(\gamma_1 \varepsilon^{(1)}_{i,j} \sigma_i \sigma_j + \gamma_2 \varepsilon^{(2)}_{i,j} \sigma_i \sigma_j\right) \mathbb{E} e^{-H(\sigma)}.$$

Hence, we get that $W_N(\beta) = \sum_{i,j} \sigma_{i,j} - \mathbb{E}\left[\sigma_{i,j}\right]$. By Theorem 6.3,

$$|\sigma_{i,j}| \leq \sum_{\sigma} \left|\gamma_1 \sum_{i,j=1}^N \varepsilon^{(1)}_{i,j} \sigma_i\sigma_j + \gamma_2 \sum_{i,j=1}^N \varepsilon^{(2)}_{i,j}\sigma_i\sigma_j\right| \mathbb{E} e^{-H(\sigma)} \lesssim \frac{2^N}{N} e^{\frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}}{2}}.$$

Hence,

$$\sigma_{ij} = o\left(\frac{2^N}{\sqrt{N}} e^{\frac{\lambda^2_{\boldsymbol{\beta},\boldsymbol{p}}}{2}}.\right) \Rightarrow |\sigma_{ij}| = o\left(\sqrt{\operatorname{Var} W_N(\beta)}\right)$$

Take $s_N = \sqrt{\operatorname{Var} W_N(\beta)}$ Therefore, we get that,

$$\frac{1}{s_N^{2+\delta}} \sum_{i,j=1}^N \mathbb{E}\left[|\sigma_{i,j} - \mathbb{E}\sigma_{ij}|^{2+\delta}\right] = o(1) \frac{1}{s_N^2} \sum_{i,j=1}^N \mathbb{E}\left[|\sigma_{i,j} - \mathbb{E}\sigma_{ij}|^2\right] = o(1)$$

By Lyaponov CLT,

$$\begin{aligned} &\frac{W_N(\beta)}{\sqrt{\operatorname{Var} W_N(\beta)}} \to \mathcal{N}(0,1) \\ \Rightarrow &\frac{\sqrt{N}}{\lambda_{\boldsymbol{\beta},\boldsymbol{p}}} \frac{W_N(\beta)}{\mathbb{E} Z_N(\beta)} \sim \mathcal{N}(0,1) \end{aligned} \tag{6.4}$$

Now, $\sum_\sigma X(\sigma) = Z_N(\beta) - \mathbb{E} Z_N(\beta) - W_N(\beta)$

$$\left|\sqrt{N}\frac{\sum_\sigma X(\sigma)}{\mathbb{E}Z_N(\beta)}\right| = O\left(\frac{1}{N}\right) + o(1) = o(1) \tag{6.5}$$

Combining (6.4) and (6.5), we get the required result. □

6.4 Proof of Theorems in Section 4

Proof of Theorem 4.1. Let us take the model to be identifiable. Assume that the condition (4.2) does not hold. Hence, there exists c_1 and c_2 such that,

$$c_1 A^{(1)} = c_2 A^{(2)}.$$

Now, take,

$$\boldsymbol{\beta} = \begin{bmatrix} c_1 + \varepsilon \\ \varepsilon \end{bmatrix} \quad \boldsymbol{\beta}' = \begin{bmatrix} \varepsilon \\ c_2 + \varepsilon \end{bmatrix}$$

This shows that $\mu_{\boldsymbol{\beta}} = \mu_{\boldsymbol{\beta}'}$ which implies that $\boldsymbol{\beta} = \boldsymbol{\beta}'$. Hence, $c_1 = c_2 = 0$ which contradicts the assumption. Hence, (4.2) is satisfied. On the other hand assume (4.2) holds. Assume that the model is not identifiable. Hence, there exists $\boldsymbol{\beta} \neq \boldsymbol{\beta}'$ such that $\mu_{\boldsymbol{\beta}} = \mu_{\boldsymbol{\beta}'}$. Therefore,

$$\sigma^T(\beta_1 A^{(1)} + \beta_2 A^{(2)})\sigma = \sigma(\beta_1' A^{(1)} + \beta_2' A^{(2)})\sigma,$$

for any $\sigma \in 1, -1^N$. Take $c_i = \beta_i - \beta_i'$ for $i = 1, 2$,

$$\sigma^T(c_1 A^{(1)} + c_2 A^{(2)})\sigma = 0.$$

. Define, $B = c_1 A^{(1)} + c_2 A^{(2)}$. Since, B is a traceless matrix, we have $\sigma^T B \sigma = 2\sum_{i>j} b_{ij}\sigma_i\sigma_j$. We will prove that $B = 0$ by induction on n. For $n = 2$ it is easy to show. Assume it to be true for $n - 1$,

$$\sum_{i>j} b_{ij}\sigma_i\sigma_j = 0 \implies \sum_{n>i>j} b_{ij}\sigma_i\sigma_j + \sum_{j=1}^{n-1} b_{nj}\sigma_n\sigma_j \quad = 0$$

Now take consider two equations one with $\sigma_n = 1$ and another with $\sigma_n = -1$,

$$\sum_{n>i>j} b_{ij}\sigma_i\sigma_j + \sum_{j=1}^{n-1} b_{nj}\sigma_j = 0, \quad \sum_{n>i>j} b_{ij}\sigma_i\sigma_j - \sum_{j=1}^{n-1} b_{nj}\sigma_j = 0$$

Now, we have that $\sum_{n>i>j} b_{ij}\sigma_i\sigma_j + \sum_{j=1}^{n-1} b_{nj}\sigma_j = 0$. By induction hypothesis, this implies $b_{ij} = 0$ for $n > i > j$. And,

$$\sum_{j=1}^{n-1} b_{nj}\sigma_j = 0,$$

for all $\sigma \in \{1, -1\}^N$. This means $B\sigma = 0$ for any such $\sigma \in \{1, -1\}^N$. Note that the vectors in $\{1, -1\}^N$ are linearly independent. This shows that $B = 0$. Hence, $c_1 A^{(1)} + c_2 A^{(2)} = 0$, which is a contradiction to our assumption. Hence, the model is identifiable which completes the proof. □

The proof of the following lemma follows verbatim from [3].

Lemma 6.8. *Suppose $\sigma = (\sigma_1, \cdots, \sigma_N)$ is an observation from the Ising model (4.1), where the coupling matrix A_N satisfies (4.3) and (4.4).*

Then,

$$\begin{aligned}
&\limsup_{N\to\infty} \frac{1}{N}\mathbb{E}\left[\sum_{i=1}^{N}\left(\sigma_i - \tanh\left(\beta_1 m_i(\sigma) + \beta_2 m_i'(\sigma)\right)\right) m_i(\sigma)\right]^2 < \infty \\
&\limsup_{N\to\infty} \frac{1}{N}\mathbb{E}\left[\sum_{i=1}^{N}\left(\sigma_i - \tanh\left(\beta_1 m_i(\sigma) + \beta_2 m_i'(\sigma)\right)\right) m_i'(\sigma)\right]^2 < \infty
\end{aligned} \tag{6.6}$$

Proof. Assume that $|\gamma| = 1$. Define a function $F : \{-1,1\}^N \times \{-1,1\}^N \to \mathbb{R}$ as

$$F(\tau, \tau') = \frac{1}{2}\sum_{i=1}^{N}\left(m_i(\tau) + m_i(\tau')\right)(\tau_i - \tau_i')$$

Note that F is antisymmetric, that is, $F(\tau,\tau') \equiv -F(\tau',\tau)$.

Now choose a coordinate I uniformly at random, and replace the I th coordinate of σ by a sample drawn from the conditional distribution of σ_I given $(\sigma_j)_{j\neq I}$. Call the resulting vector σ'. Then (σ, σ') is an exchangeable pair of random variables. Observe that

$$F(\sigma,\sigma') = m_I(\sigma)(\sigma_I - \sigma_I')$$

Now let

$$\begin{aligned}
f(\sigma) := \mathbb{E}\left(F(\sigma,\sigma') \mid \sigma\right) &= \frac{1}{N}\sum_{i=1}^{N} m_i(\sigma)\left(\sigma_i - \mathbb{E}\left(\sigma_i \mid (\sigma_j)_{j\neq i}\right)\right) \\
&= \frac{1}{N}\sum_{i=1}^{N} m_i(\sigma)\left(\sigma_i - \tanh\left(\beta_1 m_i(\sigma) + \beta_2 m_i'(\sigma)\right)\right)
\end{aligned}$$

Then $\mathbb{E}\left(f(\sigma)^2\right) = \mathbb{E}\left(f(\sigma)F(\sigma,\sigma')\right)$. Since (σ,σ') is an exchangeable pair, therefore

$$\mathbb{E}\left(f(\sigma)F(\sigma,\sigma')\right) = \mathbb{E}\left(f(\sigma')F(\sigma',\sigma)\right)$$

Again, since F is antisymmetric, we have $\mathbb{E}\left(f(\sigma')F(\sigma',\sigma)\right) = -\mathbb{E}\left(f(\sigma')F(\sigma,\ \sigma')\right)$. Combining, we have

$$\begin{aligned}
\mathbb{E}\left(f(\sigma)^2\right) &= \mathbb{E}\left(f(\sigma)F(\sigma,\sigma')\right) = -\mathbb{E}\left(f(\sigma')F(\sigma,\sigma')\right) \\
&= \frac{1}{2}\mathbb{E}\left(\left(f(\sigma) - f(\sigma')\right)F(\sigma,\sigma')\right)
\end{aligned} \tag{6.7}$$

For any $1 \le j \le N$ and $\tau \in \{-1,1\}^N$, let

$$\tau^{(j)} := (\tau_1, \ldots, \tau_{j-1}, -\tau_j, \tau_{j+1}, \ldots, \tau_N)$$

and

$$p_j(\tau) := \frac{e^{-\beta_1 \tau_j m_j(\tau) - \beta_2 \tau_j m_j'(\tau)}}{e^{\beta_1 m_j(\tau) + \beta_2 m_j'(\tau)} + e^{-\beta_1 m_j(\tau) - \beta_2 m_j(\tau)}} = \mathbb{P}\left\{\sigma_j' = -\tau_j \mid \sigma = \tau, I = j\right\}$$

Then

$$\begin{aligned}
&\mathbb{E}\left(\left(f(\sigma) - f\left(\sigma'\right)\right) F\left(\sigma, \sigma'\right) \mid \sigma\right) \\
&\quad = \frac{1}{N} \sum_{j=1}^{N} \left(f(\sigma) - f\left(\sigma^{(j)}\right)\right) F\left(\sigma, \sigma^{(j)}\right) p_j(\sigma) \\
&\quad = \frac{1}{N} \sum_{j=1}^{N} \left(f(\sigma) - f\left(\sigma^{(j)}\right)\right) \left(2 m_j(\sigma) \sigma_j\right) p_j(\sigma)
\end{aligned} \tag{6.8}$$

Now, for ease of notation, we define the functions a_i and b_{ij} as

$$a_i(\tau) := \tau_i - \tanh\left(\beta_1 m_i(\tau) + \beta_2 m_i'(\tau)\right) \tag{6.9}$$

and

$$b_{ij}(\tau) := \tanh\left(\beta_1 m_i(\tau) + \beta_2 m_i'(\tau)\right) - \tanh\left(\beta_1 m_i(\tau^{(j)}) + \beta_2 m_i'(\tau^{(j)})\right) \tag{6.10}$$

Then $f(\tau) = N^{-1} \sum_i m_i(\tau) a_i(\tau)$, and hence

$$\begin{aligned}
&f(\sigma) - f\left(\sigma^{(j)}\right) \\
&\quad = \frac{1}{N} \sum_{i=1}^{N} \left(m_i(\sigma) - m_i\left(\sigma^{(j)}\right)\right) a_i(\sigma) + \frac{1}{N} \sum_{i=1}^{N} m_i\left(\sigma^{(j)}\right) \left(a_i(\sigma) - a_i\left(\sigma^{(j)}\right)\right) \\
&\quad = \frac{2\sigma_j}{N} \sum_{i=1}^{N} A_{ij}^{(1)} a_i(\sigma) + \frac{2 m_j(\sigma) \sigma_j}{N} - \frac{1}{N} \sum_{i=1}^{N} m_i\left(\sigma^{(j)}\right) b_{ij}(\sigma)
\end{aligned}$$

Let T_{1j}, T_{2j} and T_{3j} be the three terms on the last line. Using (6.7) and (6.8), we see that

$$\mathbb{E}\left(f(\sigma)^2\right) = \frac{1}{N} \sum_{j=1}^{N} \left(T_{1j} + T_{2j} + T_{3j}\right) m_j(\sigma) \sigma_j p_j(\sigma) \tag{6.11}$$

Now, since $\sum_i a_i(\sigma)^2 \leq 4N$ and

$$\sum_j m_j(\sigma)^2 p_j(\sigma)^2 \leq \sum_j m_j(\sigma)^2 = \|A^{(1)} \sigma\|^2 \leq N$$

it follows that

$$\left|\frac{1}{N}\sum_{j=1}^{N} T_{1j} m_j(\sigma)\sigma_j p_j(\sigma)\right| = \frac{2}{N^2}\left|\sum_{i,j=1}^{N} A_{ij}^{(1)} a_i(\sigma) m_j(\sigma) p_j(\sigma)\right|$$
$$\leq \frac{2}{N^2}\sqrt{4N}\sqrt{N} = \frac{4}{N}$$

Next, let us look at the T_2-term. We have

$$\frac{1}{N}\sum_{j=1}^{N} T_{2j} m_j(\sigma)\sigma_j p_j(\sigma) = \frac{2}{N^2}\sum_{j=1}^{N} m_j(\sigma)^2 p_j(\sigma)$$
$$\leq \frac{2}{N^2}\sum_{j=1}^{N} m_j(\sigma)^2 \leq \frac{2}{N}$$

Finally, let us bound the T_3-term. Take any i and let e_i be the i th coordinate vector in $\mathbb{R}^N$. Then

$$\sum_{j=1}^{n} (A_{ij}^{(1)})^2 = \left\|e_i^t A^{(1)}\right\|^2 \leq \|e_i\|^2 \|A^{(1)}\|^2 \leq 1$$

Thus, if we let J_2 be the matrix $\left((A_{ij}^{(1)})^2\right)_{1\leq i,j\leq N}$, then by the well-known result that the L^2 operator norm of a symmetric matrix is bounded by its L^∞ operator norm, we get

$$\|J_2\| \leq \max_{1\leq i\leq N}\sum_{j=1}^{N} (A_{ij}^{(1)})^2 \leq 1$$

Now let $h(x) := \tanh(\beta_1 x + \beta_2 y)$. It is easy to verify that $\|\nabla^2 h\| \leq \beta_1^2 + \beta_2^2$. Therefore,

$$\left|h\left(m_i(\sigma)\right) - h\left(m_i\left(\sigma^{(j)}\right)\right) - \left(m_i(\sigma) - m_i\left(\sigma^{(j)}\right)\right) h'\left(m_i(\sigma)\right)\right|$$
$$\leq \frac{\beta_1^2+\beta_2^2}{2}\left(\left(m_i(\sigma) - m_i\left(\sigma^{(j)}\right)\right)^2 + \left(m_i'(\sigma) - m_i'\left(\sigma^{(j)}\right)\right)^2\right)$$

Let $c_i(\sigma) := h_x\left(m_i(\sigma)\right)$, $d_i(\sigma) := h_y\left(m_i(\sigma)\right)$. Note that $m_i(\sigma) - m_i\left(\sigma^{(j)}\right) = 2A_{ij}^{(1)}\sigma_j$ and $m_i'(\sigma) - m_i'\left(\sigma^{(j)}\right) = 2A_{ij}^{(2)}\sigma_j$. So the above inequality can be rewritten as

$$\left|b_{ij}(\sigma) - 2A_{ij}^{(1)}\sigma_j c_i(\sigma) - 2A_{ij}^{(2)}\sigma_j d_i(\sigma)\right| \leq 2\left(\beta_1^2+\beta_2^2\right)\left(\left(A_{ij}^{(1)}\right)^2 + \left(A_{ij}^{(2)}\right)^2\right).$$

Finally, note that $|c_i(\sigma)| \leq \beta_1$ and $|d_i(\sigma)| \leq \beta_2$. Using all this information and the bounds on the operator norms of J and J_2, we see that for any $x, y \in \mathbb{R}^n$,

$$\left|\sum_{i,j} x_i y_j b_{ij}(\sigma)\right| \leq \left|\sum_{i,j} x_i y_j \left(2A_{ij}^{(1)}\sigma_j c_i(\sigma) + 2A_{ij}^{(2)}\sigma_j d_i(\sigma)\right)\right|$$

$$
\begin{aligned}
&+\left|\sum_{i,j} x_i y_j \left(b_{ij}(\sigma) - 2A^{(1)}_{ij}\sigma_j c_i(\sigma) - 2A^{(2)}_{ij}\sigma_j d_i(\sigma)\right)\right| \\
\leq& 2\left(\sum_i \left(x_i\left(c_i(\sigma)+d_i(\sigma)\right)\right)^2\right)^{1/2}\left(\sum_j (y_j\sigma_j)^2\right)^{1/2} + \sum_{i,j}|x_i y_j|\, 2\left(\beta_1^2+\beta_2^2\right)\left(\left(A^{(1)}_{ij}\right)^2 + \left(A^{(2)}_{ij}\right)^2\right) \\
\leq& (2(\beta_1+\beta_2) + 2(\beta_1^2+\beta_2^2))\|x\|\|y\|.
\end{aligned}
$$

Again, it is clear from the definition (6.10) of b_{ij} that $|b_{ij}(\sigma)| \leq 2\beta_1\left|A^{(1)}_{ij}\right| + 2\beta_2\left|A^{(2)}_{ij}\right|$. Thus,

$$
\left|\sum_{i,j} x_i y_j A^{(1)}_{ij} b_{ij}(\sigma)\right| \leq \sum_{i,j}|x_i y_j|\left(2\beta_1\left|A^{(1)}_{ij}\right| + 2\beta_2\left|A^{(2)}_{ij}\right|\right)(A^{(1)}_{ij}) \leq 2c(\beta_1+\beta_2)\|x\|\|y\|,
$$

where c is some universal constant. Applying these inequalities to the T_3-term in (13), we get

$$
\begin{aligned}
\left|\frac{1}{N}\sum_{j=1}^N T_{3j} m_j(\sigma)\sigma_j p_j(\sigma)\right| &= \left|\frac{1}{N^2}\sum_{i,j=1}^N m_i\left(\sigma^{(j)}\right) b_{ij}(\sigma) m_j(\sigma)\sigma_j p_j(\sigma)\right| \\
&= \left|\frac{1}{N^2}\sum_{i,j=1}^N \left(m_i(\sigma) - 2A^{(1)}_{ij}\sigma_j\right) b_{ij}(\sigma) m_j(\sigma)\sigma_j p_j(\sigma)\right| \\
&\leq \frac{C_{\boldsymbol{\beta}}}{N}
\end{aligned}
$$

Thus, we have computed upper bounds for all terms in (6.11). Combining, we have

$$
\limsup \mathbb{E}\left(Nf(\sigma)^2\right) \leq \infty
$$

The corresponding statement for $m_i'(\sigma)$ can be proved similarly. This completes the proof. □

Proof of Theorem 4.2.

$$
\begin{aligned}
Q(\boldsymbol{\beta} \mid \sigma) =& \frac{\partial}{\partial\beta_1}\tilde{P}_L = \sum_{i=1}^N \sigma_i m_i(\sigma) - m_i(\sigma)\tanh\left(\beta_1 m_i(\sigma) + \beta_2 m_i'(\sigma)\right) \\
R(\boldsymbol{\beta} \mid \sigma) =& \frac{\partial}{\partial\beta_2}\tilde{P}_L = \sum_{i=1}^N \sigma_i m_i'(\sigma) - m_i'(\sigma)\tanh\left(\beta_1 m_i(\sigma) + \beta_2 m_i'(\sigma)\right)
\end{aligned}
$$

Setting

$$\widetilde{PL}_N(\boldsymbol{\beta}|\sigma) := \sum_{i=1}^{N} \left(\beta_1\sigma_i m_i(\sigma) + \beta_2\sigma_i m_i'(\sigma) - \log\cosh(\beta_1 m_i(\sigma) + \beta_2 m_i'(\sigma))\right) \tag{6.12}$$

note that $PL_N(\boldsymbol{\beta}|\sigma) = \nabla\widetilde{PL}_N(\boldsymbol{\beta}|\sigma)$. Differentiating the function $(\boldsymbol{\beta}) \mapsto \widetilde{PL}_N(\boldsymbol{\beta}|\sigma)$ twice we get the negative Hessian matrix given by

$$H_N(\boldsymbol{\beta}|\sigma) = \begin{bmatrix} \sum_{i=1}^N m_i(\sigma)^2\theta_i(\boldsymbol{\beta}|\sigma) & \sum_{i=1}^N m_i(\sigma)m_i'(\sigma)\theta_i(\boldsymbol{\beta}|\sigma) \\ \sum_{i=1}^N m_i(\sigma)m_i'(\sigma)\theta_i(\boldsymbol{\beta}|\sigma) & \sum_{i=1}^N m_i'(\sigma)^2\theta_i(\boldsymbol{\beta}|\sigma) \end{bmatrix}, \tag{6.13}$$

where $\theta_i(\boldsymbol{\beta}|\sigma) := \operatorname{sech}^2\left(\beta_1 m_i(\sigma) + \beta_2 m_i'(\sigma)\right)$. The determinant of the Hessian is given by

$$\begin{aligned}
\det H_N(\boldsymbol{\beta} \mid x) = & \sum_{i,j=1}^N m_i(x)^2 m_j'(x)^2\theta_i(\boldsymbol{\beta} \mid x)\theta_j(\boldsymbol{\beta} \mid x) - \sum_{i,j=1}^N m_i(x)m_i'(x)m_j(x)m_j'(x)\theta_i(\boldsymbol{\beta} \mid x)\theta_j(\boldsymbol{\beta} \mid x) \\
= & \frac{1}{2}\Bigg(\sum_{i,j=1}^N m_i(x)^2 m_j'(x)^2\theta_i(\boldsymbol{\beta} \mid x)\theta_j(\boldsymbol{\beta} \mid x) - \sum_{i,j=1}^N m_i(x)m_i'(x)m_j(x)m_j'(x)\theta_i(\boldsymbol{\beta} \mid x)\theta_j(\boldsymbol{\beta} \mid x) \\
& + \sum_{i,j=1}^N m_j(x)^2 m_i'(x)^2\theta_j(\boldsymbol{\beta} \mid x)\theta_i(\boldsymbol{\beta} \mid x) - \sum_{i,j=1}^N m_j(x)m_j'(x)m_i(x)m_i'(x)\theta_j(\boldsymbol{\beta} \mid x)\theta_i(\boldsymbol{\beta} \mid x)\Bigg) \\
= & \sum_{i,j=1}^N \left(m_i(x)m_j'(x) - m_j(x)m_i'(x)\right)^2\theta_j(\boldsymbol{\beta} \mid x)\theta_i(\boldsymbol{\beta} \mid x) \\
\geqslant & \operatorname{sech}^4((\beta_1+\beta_2)|\gamma|)\sum_{i,j=1}^N \left(m_i(x)m_j'(x) - m_j(x)m_i'(x)\right)^2 \\
= & sech^4((\beta_1+\beta_2)|\gamma|)N^2 T_N(\sigma)
\end{aligned}$$

which gives

$$|H_N(\boldsymbol{\beta}|\sigma)| = \lambda_N(\boldsymbol{\beta}|\sigma)\mu_N(\boldsymbol{\beta}|\sigma) \geq \operatorname{sech}^4((\beta_1+\beta_2)|\gamma|)N^2 T_N(\sigma). \tag{6.14}$$

Recall the 2×2 matrix $H_N(\boldsymbol{\beta}|\sigma)$ as defined in (6.13), and denote $\lambda_N(\boldsymbol{\beta}|\sigma) \geq \mu_N(\boldsymbol{\beta}|\sigma)$ to be its eigenvalues. We start by giving a lower bound to the minimum eigenvalue $\mu_N(\boldsymbol{\beta}|\sigma)$. To this effect, note that

$$\lambda_N(\boldsymbol{\beta}|\sigma) + \mu_N(\boldsymbol{\beta}|\sigma) = \operatorname{tr} H_N(\boldsymbol{\beta}|\sigma) = \sum_{i=1}^N \theta_i(\boldsymbol{\beta}|\sigma)(m_i(\sigma)^2 + m_i'(\sigma)^2) \leq 2n\gamma^2,$$

which along with (6.14) gives

$$\mu_N(\boldsymbol{\beta}|\sigma) \geq \frac{\lambda_N(\boldsymbol{\beta}|\sigma)\mu_N(\boldsymbol{\beta}|\sigma)}{\lambda_N(\boldsymbol{\beta}|\sigma) + \mu_N(\boldsymbol{\beta}|\sigma)} = \frac{\det H_N(\boldsymbol{\beta}|\sigma)}{\operatorname{tr} H_N(\boldsymbol{\beta}|\sigma)} \geq \frac{\operatorname{sech}^4((\beta_1+\beta_2)|\gamma|)}{2\gamma^2} N T_N(\sigma). \tag{6.15}$$

Armed with this estimate, we now complete the proof of the Theorem. To this effect, setting $(\beta_1(t), \beta_2(t)) = (t\hat{\beta}_{N,1} + (1-t)\beta_1, t\hat{\beta}_{N,2} + (1-t)\beta_2)$, define a function $g_N : [0,1] \to \mathbb{R}$ by

$$g_N(t) := (\hat{\beta}_{N,1} - \beta_1)Q_N\left(\beta_1(t), \beta_2(t)|\sigma\right) + (\hat{\beta}_{N,2} - \beta_2)R_N\left(\beta_1(t), \beta_2(t)|\sigma\right),$$

and note that

$$|g_N(1) - g_N(0)| = |(\hat{\beta}_{N,1} - \beta_1)Q_N(\beta_1, \beta_2|\sigma) + (\hat{\beta}_{N,2} - \beta_2)R_N(\beta_1, \beta_2|\sigma)| = O_p(\sqrt{N}Y_N), \quad (6.16)$$

where $Y_N := ||\hat{\beta}_{N,1} - \beta_1, \hat{\beta}_{N,2} - \beta_2||_2$, and we use Cauchy-Schwarz inequality along with (6.6) of Lemma 6.8. Also, we have

$$g_N'(t) = (\hat{\beta}_{N,1} - \beta_1, \hat{\beta}_{N,2} - \beta_2)H_N(\beta_1(t), \beta_2(t)|\sigma)(\hat{\beta}_{N,1} - \beta_1, \hat{\beta}_{N,2} - \beta_2)^\top \geq \mu_N(\beta_1(t), \beta_2(t)|\sigma)Y_N^2,$$

In particular we have $g_N'(t) \geq 0$ for all $t \in (0,1)$. Further, using (6.15) we get the existence of $r, s > 0$ such that

$$\inf_{(\boldsymbol{\beta}) \in \Theta : ||\beta - \beta_1, B - \beta_2|| \leq r} \mu_N(\boldsymbol{\beta}|\sigma) \geq sNT_N(\sigma).$$

Noting that $||\beta_1(t) - \beta_1, \beta_2(t) - \beta_2||_2 = tY_N$ gives

$$\int_0^1 g_N'(t)dt \geq \int_0^{\min(1, \frac{r}{Y_N})} g_N'(t)dt \geq \min\left(1, \frac{r}{Y_N}\right) snT_N(\sigma)W_N^2, \quad (6.17)$$

which along with (6.16) gives $\min(Y_N, r) = O_p(\frac{1}{\sqrt{N}T_N(\sigma)})$. Since $r > 0$ is fixed, it follows that $Y_N = o_p(1)$, and so $(\hat{\beta}_{N,1}, \hat{\beta}_{N,2})$ converges in probability to (β_1, β_2). This shows that $Y_N < r$ with probability tending to 1, which on using (6.17) gives $\int_0^1 g_N'(t)dt \geq sNT_N(\sigma)Y_N^2$. Along with (6.16) this gives $nT_N(\sigma)Y_N = O_p(\sqrt{N})$, which is the claimed bound. □

References

[1] Bhowal, S., and Mukherjee, S. Limit Theorems and Phase Transitions in the Tensor Curie-Weiss Potts Model. *arXiv preprint arXiv:2307.01052* (2023).

[2] Bhowal, S., and Mukherjee, S. Rates of Convergence of the Magnetization in the Tensor Curie-Weiss Potts Model. *arXiv preprint arXiv:2406.15907* (2024).

[3] Chatterjee, S. Estimation in spin glasses: A first step. *The Annals of Statistics 35*, 5 (2007), 1931 – 1946.

[4] Ellis, R. S. *Entropy, large deviations, and statistical mechanics.* Springer, 2007.

[5] Kabluchko, Z., Löwe, M., and Schubert, K. Fluctuations of the magnetization for Ising models on dense Erdős–Rényi random graphs. *Journal of Statistical Physics 177* (2019), 78–94.

[6] Kabluchko, Z., Löwe, M., and Schubert, K. Fluctuations for the partition function of Ising models on Erdös–Rényi random graphs. In *Annales de l'Institut Henri Poincare (B) Probabilites et statistiques* (2021), vol. 57, Institut Henri Poincaré, pp. 2017–2042.