

# Interaction Order Estimation in Tensor Curie–Weiss Models

Somabha Mukherjee 

Department of Statistics and Data Science, Faculty of Science, National University of Singapore, Singapore 117546, Singapore; somabha@nus.edu.sg

**Abstract:** In this paper, we consider the problem of estimating the interaction parameter  $p$  of a  $p$ -spin Curie–Weiss model at inverse temperature  $\beta$ , given a single observation from this model. We show, by a contiguity argument, that joint estimation of the parameters  $\beta$  and  $p$  is impossible, which implies that the estimation of  $p$  is impossible if  $\beta$  is unknown. These impossibility results are also extended to the more general  $p$ -spin Erdős–Rényi Ising model. The situation is more delicate when  $\beta$  is known. In this case, we show that there exists an increasing threshold function  $\beta^*(p)$ , such that for all  $\beta$ , consistent estimation of  $p$  is impossible when  $\beta^*(p) > \beta$ , and for almost all  $\beta$ , consistent estimation of  $p$  is possible for  $\beta^*(p) < \beta$ .

**Keywords:** Curie–Weiss model; joint estimation; interaction order; contiguity

## 1. Introduction

The Ising model [1] was originally introduced in Physics as a model for ferromagnetism, and has since then found numerous interesting applications in diverse areas, such as image processing [2], neural networks [3], spatial statistics [4], and disease mapping in epidemiology [5]. The classical (2-spin) Ising model is a discrete probability measure on the set of all binary strings of a fixed length, given by the following:

$$\mathbb{P}_\beta(\mathbf{x}) := \frac{\exp\{-\beta H_N(\mathbf{x})\}}{Z_N(\beta)} \quad \text{for } \mathbf{x} \in \{-1, 1\}^N \quad (1)$$

where the *Hamiltonian*  $H_N(\mathbf{x})$  is given by

$$H_N(\mathbf{x}) := - \sum_{1 \leq i, j \leq N} J_{i,j} x_i x_j,$$

$J := ((J_{i,j}))_{1 \leq i, j \leq N}$  is the interaction matrix (often taken as the scaled adjacency matrix of some graph),  $\beta > 0$  denotes the interaction strength (called *inverse temperature* in the physics literature), and  $Z_N(\beta)$  is the normalizing constant, needed to ensure that the probabilities in (1) add up to 1 (referred to as the *partition function* in Physics). The *Hamiltonian*  $\sum_{1 \leq i, j \leq N} J_{i,j} x_i x_j$  of the model (1) only captures pairwise dependence between the binary variables, arising from an underlying network/interaction structure. Unfortunately, in most real-life scenarios, pairwise interactions are not enough to explain all of the complex dependencies arising in network data, and one has to take into account higher-order interactions arising from peer-group effects. Multi-body interactions are also common in many other branches of science; for example in Chemistry [6], it is known that the atoms on a crystal surface do not just interact in pairs, but in triangles, quadruples, and higher-order tuples. A natural extension of the classical 2-spin Ising model with a focus on capturing higher-order dependencies is the  $p$ -spin Ising model [7–10], where the quadratic interaction



Academic Editor: Adam Lipowski

Received: 22 January 2025

Revised: 23 February 2025

Accepted: 24 February 2025

Published: 27 February 2025

**Citation:** Mukherjee, S. Interaction Order Estimation in Tensor Curie–Weiss Models. *Entropy* **2025**, *27*, 245. <https://doi.org/10.3390/e27030245>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

term in the sufficient statistic is replaced by a multilinear polynomial of degree  $p \geq 2$ . The probability mass function of this model is given by the following:

$$\mathbb{P}_{\beta,p}(\mathbf{x}) := \frac{\exp\{-\beta H_N(\mathbf{x})\}}{Z_N(\beta,p)} \quad \text{for } \mathbf{x} \in \{-1,1\}^N \quad (2)$$

where the Hamiltonian  $H_N(\mathbf{x})$  is given by the following:

$$H_N(\mathbf{x}) := - \sum_{1 \leq i_1, \dots, i_p \leq N} J_{i_1 \dots i_p} x_{i_1} \dots x_{i_p},$$

$J := ((J_{i_1 \dots i_p}))_{(i_1, \dots, i_p) \in [N]^p}$  is the interaction tensor (often taken as the scaled adjacency tensor of a hypergraph),  $\beta > 0$  denotes the interaction strength (or inverse temperature),  $p$  is the interaction order, and  $Z_N(\beta, p)$  is the normalizing constant (or partition function).

The  $p$ -spin Ising models have applications, including their role in explaining the microscopic theory of magnetism in solid and fluid  $^3\text{He}$  films absorbed on the surface of graphite (see [11]). They also serve as important resources for analog quantum simulation and quantum computation, contribute to the mapping of fermions to artificial spin systems in quantum algorithms for quantum chemistry, and appear in spin models of cuprate superconductor Hamiltonians (see [12] and the references therein). Furthermore, they play crucial roles in error suppression schemes for quantum annealers, adiabatic topological quantum computation, and other related areas.  $p$ -spin Ising and spin glass models have also appeared in a number of classical and recent works, such as [7,8,10,13–18] and  $p$ -spin versions of the closely related Potts model have appeared in [19–21].

The estimation of the parameter  $\beta$  in the model (2), assuming that the interaction order  $p$  is known, has been studied extensively in the past (see, for example, [22–24] for the case  $p = 2$  and [9,10,16] for the case  $p \geq 3$ ). For the  $p = 2$  case,  $\sqrt{N}$ -consistency of the so called *maximum pseudolikelihood estimator (MPLE)* of  $\beta$  under some general assumptions on the model (2) was first established in [23]. These assumptions can be verified easily for Ising models on many dense graphs at low temperatures (high values of  $\beta$ ), and for Ising models on bounded-degree graphs as well as popular spin-glass models such as the Sherrington-Kirkpatrick model at all temperatures. Some of these results were extended in [22] where different rates of consistency were obtained for Ising models on sparse graphs, such as the sparse Erdős–Rényi graphs or sparse regular graphs, even at high temperatures. In addition, inestimability of  $\beta$  was established in [22] for Ising models on a sequence of dense graphs converging to a graphon for all  $\beta$  below a certain threshold, given by the inverse spectral norm of the limiting graphon. Precise central limit theorems for the *maximum likelihood estimator (MLE)* of  $\beta$  were derived for the 2-spin Ising model on complete graphs (also known as the Curie–Weiss model) in [25]. Some of these results have been extended in [9,10,16] for the case  $p \geq 3$ .

Now, suppose that one is given a sample  $X$  from a  $p$ -spin Ising model (2), with no other information on any of the parameters. A natural question is whether it is possible to estimate the interaction order  $p$  of the model. This is highly relevant from a practical perspective, as in many real-life applications such as recommender systems, strong peer-group effects exist, and two-interaction models are known to fit the data much worse than higher-order interaction models. For example, in [26], it was empirically shown that the 3-spin Ising model fitted the Last.fm music (accessed on 9 December 2020) (<http://millionsongdataset.com/lastfm/>) much better than the classical 2-spin model, thereby indicating the presence of complex peer-group effects in this dataset. This is a fan network database for a number of popular artists and bands, and the data consist of a list of binary opinions from the fan base for each artist, indicating whether each user is a fan of that artist or not as well as the user friendship network. The exact procedure involved

first estimating the parameter  $\beta$  assuming a 2-spin Ising model from the data using the *maximum pseudolikelihood estimator*, followed by simulating a number of samples from the corresponding 2-spin Ising model with the estimated  $\beta$  as the parameter. It was observed that for most artists under consideration, the true Hamiltonian from the data lies outside the 2.5% and 97.5% quantiles of the histogram of the sampled Hamiltonians, thereby suggesting misfit. When a 3-spin Ising model was fit with the triangles in the user network as the hyperedges, the true Hamiltonian actually fell comfortably within the bulk of the histogram, thereby suggesting a good fit. In a different context, in [11], the authors mentioned that experimental and theoretical data indicated the relevance of three, four, five and six-spin exchange processes over a wide density range in adsorbed  $^3\text{He}$  films.

One might, in such situations, be interested in figuring out a systematic way of determining the value of the interaction order  $p$  for which the corresponding  $p$ -tuple interaction model fits the data best. To the best of our knowledge, this reverse problem of estimating the interaction order  $p$  given a single observation from the model (2) even with known  $\beta$  has not been addressed in the literature before. This question may be quite difficult for arbitrary underlying interaction tensors  $J$ , which necessitates some convenient structural assumptions on this tensor.

In this paper, we assume that tensor  $J$  has all entries equal to  $N^{1-p}$ , which corresponds to the  $p$ -spin Curie–Weiss model [7,9,27], given by the following:

$$\mathbb{P}_{\beta,p}(\mathbf{x}) := \frac{\exp\{\beta N \bar{x}^p\}}{Z_N(\beta,p)} \quad \text{for } \mathbf{x} \in \{-1, 1\}^N \quad (3)$$

where  $\bar{x} := \frac{1}{N} \sum_{i=1}^N x_i$ . Even under this structural assumption, the possibility of estimating  $p$  depends on whether the interaction strength  $\beta$  is known or not. We will show that consistent estimation of  $p$  is impossible if  $\beta$  is unknown, which will be a consequence of our argument on the impossibility of the joint consistent estimation of  $\beta$  and  $p$ . At the heart of these impossibility results is the fact that the sufficient statistic  $\bar{X}$  in the model (3) converges to a mixture of point masses at the maximizers of a certain function  $H_{\beta,p}$ , and that these maximizers do not uniquely determine the tuple  $(\beta, p)$ . This idea is formalized by a contiguity argument between the Curie–Weiss measures and  $N$ -fold products of Rademacher distributions, with the largest maximizer of  $H_{\beta,p}$  as the mean. It should be mentioned here that the related problem of joint inestimability of  $(\beta, h)$  for the classical 2-spin Curie–Weiss model with an additional magnetic field parameter  $h$  was addressed in [24] using similar contiguity arguments. We also extend these impossibility results to the more general  $p$ -spin Erdős–Rényi Ising models.

The situation is more intricate when  $\beta$  is known. In this case, we will show that there exists a strictly increasing threshold function  $\beta^*(p)$ , such that the consistent estimation of  $p$  is impossible for  $\beta^*(p) > \beta$ . However, for *almost all*  $\beta$  (to be precise, for all but possibly countably many  $\beta$ ),  $p$  can be estimated consistently whenever  $\beta^*(p) < \beta$ . The question of exactly describing the exceptional set of countably many  $\beta$ s for which the region  $\beta^*(p) < \beta$  is inestimable for  $p$ , is still open. Finally, we want to mention that although joint consistent estimation of  $\beta$  and  $p$  is impossible using just one sample from the  $p$ -spin Curie–Weiss model (3), we still hope to do so using multiple samples.

## 2. Impossibility of Jointly Estimating $(\beta, p)$

We start by showing that joint consistent estimation of  $\beta$  and  $p$  using only one sample  $\mathbf{X} := (X_1, \dots, X_N)$  from the model (3) is, in general, impossible. Towards this, for every  $m \in [0, 1)$ , define a set:

$\Theta_m := \{(\beta, p) \in (0, \infty) \times D : m \text{ is the unique non-negative global maximizer of } H_{\beta,p}\}$

where  $D$  denotes the set of all integers  $\geq 3$ , and

$$H_{\beta,p}(x) := \beta x^p - \frac{1}{2} \{ (1+x) \log(1+x) + (1-x) \log(1-x) \} \quad \text{for } x \in [0, 1].$$

**Theorem 1.** For every  $m \in [0, 1)$ , such that  $|\Theta_m| \geq 2$ , there does not exist any sequence of estimators (measurable functions of  $\mathbf{X}$ ), which is consistent for  $(\beta, p) \in \Theta_m$  under the model (3).

The following lemma is crucial for proving Theorem 1.

**Lemma 1.** For every  $m \in [0, 1)$ , denote the distribution of a Rademacher random variable with mean  $m$  by  $\mu$ . Then, the product measure  $\mathbb{Q} := \otimes_{i=1}^N \mu$  is contiguous to  $\mathbb{P}_{\beta,p}$  for all  $m \in [0, 1)$ , and all  $(\beta, p) \in \Theta_m$ .

**Proof.** To begin with, note that on the event  $E_K := \{\sqrt{N}(\bar{X} - m) \leq K\}$ , we have:

$$\begin{aligned} \frac{\mathbb{Q}(\mathbf{X})}{\mathbb{P}_{\beta,p}(\mathbf{X})} &= \frac{Z_N(\beta, p) \prod_{i=1}^N \left(\frac{1+m}{2}\right)^{(1+X_i)/2} \left(\frac{1-m}{2}\right)^{(1-X_i)/2}}{\exp(\beta N \bar{X}^p)} \\ &= \frac{2^{-N} Z_N(\beta, p)}{\exp\left(N \left[ \beta \bar{X}^p - \frac{1}{2} \{ (1 + \bar{X}) \log(1 + m) + (1 - \bar{X}) \log(1 - m) \} \right]\right)}. \end{aligned}$$

Now, by Lemma 3.2 and Lemma 3.4 in [9], we have:

$$2^{-N} Z_N(\beta, p) = e^{NH_{\beta,p}(m)} \Theta(1).$$

Also, by Taylor expansion, we have:

$$\begin{aligned} &\beta \bar{X}^p - \frac{1}{2} \{ (1 + \bar{X}) \log(1 + m) + (1 - \bar{X}) \log(1 - m) \} - H_{\beta,p}(m) \\ &= H_{\beta,p}(\bar{X}) - H_{\beta,p}(m) + \frac{1}{2} \left\{ (1 + \bar{X}) \log\left(\frac{1 + \bar{X}}{1 + m}\right) + (1 - \bar{X}) \log\left(\frac{1 - \bar{X}}{1 - m}\right) \right\} \\ &= O\left((\bar{X} - m)^2\right) + (\bar{X} - m) \left( \frac{1 + \bar{X}}{1 + m} - \frac{1 - \bar{X}}{1 - m} \right) = O\left((\bar{X} - m)^2\right). \end{aligned}$$

Hence, we have:

$$\frac{\mathbb{Q}(\mathbf{X})}{\mathbb{P}_{\beta,p}(\mathbf{X})} = \Theta(1) e^{NO((\bar{X}-m)^2)} \leq C_K$$

for some constant  $C_K$ . Hence, for any event  $A_N \subseteq \{-1, 1\}^N$ , we have:

$$\mathbb{Q}(\mathbf{X} \in A_N, E_K) = \mathbb{E}_{\mathbb{P}_{\beta,p}} \left[ \frac{\mathbb{Q}(\mathbf{X})}{\mathbb{P}_{\beta,p}(\mathbf{X})} \mathbb{1}_{\mathbf{X} \in A_N, E_K} \right] \leq C_K \mathbb{P}(\mathbf{X} \in A_N, E_K),$$

thereby providing the following:

$$\mathbb{Q}(\mathbf{X} \in A_N) \leq C_K \mathbb{P}(\mathbf{X} \in A_N) + \mathbb{Q}(E_K^c).$$

Now, suppose that  $\mathbb{P}(\mathbf{X} \in A_N) \rightarrow 0$ . Then, for every  $K > 0$ , we have the following:

$$\limsup_{N \rightarrow \infty} \mathbb{Q}(\mathbf{X} \in A_N) \leq \limsup_{N \rightarrow \infty} \mathbb{Q}\left(\sqrt{N}(\bar{X} - m) > K\right) = \mathbb{P}(N(0, 1 - m^2) > K).$$

Taking the limit as  $K \rightarrow \infty$  throughout the above inequality, we can conclude that  $\mathbb{Q}(\mathbf{X} \in A_N) \rightarrow 0$ , which completes the proof of Lemma 1.  $\square$

With Lemma 1 in hand, we are now ready to prove Theorem 1.

**Proof of Theorem 1.** Suppose that there exists a sequence  $T_N(\mathbf{X}) \in \mathbb{R}^2$  of consistent estimators of  $(\beta, p)$  on  $\Theta_m$ . Fixing two different points  $(\beta_1, p_1)$  and  $(\beta_2, p_2) \in \Theta_m$ , we can construct disjoint neighborhoods  $\mathcal{B}_1$  and  $\mathcal{B}_2$  around them, respectively. Through the consistency of  $T_N(\mathbf{X})$  on  $\Theta_m$ , we have the following:

$$\mathbb{P}_{\beta_i, p_i}(T_N(\mathbf{X}) \in \mathcal{B}_i) \rightarrow 1 \quad \text{for } i = 1, 2.$$

From Lemma 1, we have the following:

$$\mathbb{Q}(T_N(\mathbf{X}) \in \mathcal{B}_i) \rightarrow 1 \quad \text{for } i = 1, 2,$$

which contradicts the facts that  $\mathcal{B}_1, \mathcal{B}_2$  are disjointed, and  $\mathbb{Q}$  is a probability measure.  $\square$

**Remark 1.** Our argument implies that the consistent estimation of  $p$  is impossible if  $\beta$  is unknown. Because, if there were such a consistent estimator  $\hat{p} := \hat{p}(\mathbf{X})$ , then we could choose  $(\beta_1, p_1)$  and  $(\beta_2, p_2)$  from some  $\Theta_m$ ; construct disjoint open intervals  $\mathcal{I}_1$  and  $\mathcal{I}_2$  around  $p_1$  and  $p_2$ , respectively; and argue from contiguity that  $\mathbb{Q}(\hat{p}(\mathbf{X}) \in \mathcal{I}_i) \rightarrow 1$  for  $i = 1, 2$ , which is a contradiction.

The question is how do the sets  $\Theta_m$  look for different values of  $m \in [0, 1)$ ? To answer this, let us define for each  $(\beta, p)$ ,

$$\beta^*(p) := \sup \left\{ \beta > 0 : \sup_{x \in [0,1]} H_{\beta,p}(x) = 0 \right\}.$$

It follows from [9] that for  $p \in D$ , the function  $H_{\beta,p}$  has a unique positive maximizer  $m_*(\beta, p)$  if  $\beta > \beta^*(p)$ . By convention, we define  $m_p = m_*(\beta^*(p), p) := \lim_{\beta \rightarrow \beta^*(p)^-} m_*(\beta, p)$ .

**Proposition 1.** The sequence  $\{m_p\}_{p \in D} \rightarrow 1$  as  $p \rightarrow \infty$ , which can be sorted in ascending order as  $0 < m_{p_1} \leq m_{p_2} \leq \dots < 1$ . Further, if we denote  $\Theta_m|_2$  to be the projection of  $\Theta_m$  onto the  $p$ -coordinate, then

$$\Theta_m|_2 = \begin{cases} \emptyset & \text{if } 0 < m \leq m_{p_1}, \\ \{p_1, \dots, p_k\} & \text{if } m_{p_k} < m \leq m_{p_{k+1}}, \quad k \geq 1. \end{cases}$$

The proof of Proposition 1 is technical, and is given in Appendix A. Note that for  $m \in (0, 1)$ , the set  $\Theta_m$  is uniquely determined by its projection  $\Theta_m|_2$ , because every  $p \in \Theta_m|_2$  uniquely corresponds to the element  $(\beta_p, p) \in \Theta_m$ , where  $\beta_p := p^{-1}m^{1-p} \tanh^{-1}(m)$ . Proposition 1 thus provides a complete description of the family of sets  $\Theta_m$  for all values of  $m \in (0, 1)$ , and states that although each of these sets is finite, we can choose sets as large as possible from this family. The following proposition describes the set  $\Theta_0$ .

**Proposition 2.**  $\Theta_0 = \{(\beta, p) \in (0, \infty) \times D : \beta < \beta^*(p)\}$ .

The proof of Proposition 2 follows from the following three facts proved in [9]:

1. If  $\beta < \beta^*(p)$ , then 0 is the unique global maximizer of  $H_{\beta,p}$ .
2. If  $\beta = \beta^*(p)$ , then  $H_{\beta,p}$  has two different non-negative global maximizers.

3. If  $\beta > \beta^*(p)$ , then  $H_{\beta,p}$  has a unique non-negative global maximizer, which happens to be positive.

**Remark 2.** All of the results in this section also hold almost surely for the somewhat more general  $p$ -spin Erdős–Rényi Ising model [27], where the interaction tensor  $\mathbf{J}$  in (2) is given by  $\alpha^{-1}N^{1-p}\mathbf{A}$ , with the entries  $A_{i_1,\dots,i_p}$  of the tensor  $\mathbf{A}$  being i.i.d. Bernoulli random variables with a mean  $\alpha$ , for some fixed  $\alpha \in (0, 1)$ . This follows from the fact that the Erdős–Rényi Ising and Curie–Weiss measures are mutually contiguous, which follows from Lemma 6.6 in [28].

### 3. Estimation of $p$ When $\beta$ Is Known

Throughout this section, we assume that the parameter  $\beta$  is known. To begin with, for every  $\beta > 0$ , let us define the following two sets:

$$U_\beta := \{p \geq 2 : \beta^*(p) > \beta\} \quad \text{and} \quad L_\beta := \{p \geq 2 : \beta^*(p) < \beta\}.$$

For Lemma A.1 in [10], the set  $U_\beta$  is empty if  $\beta \geq \log 2$ , and is of the form  $\{q, q + 1, \dots\}$  for some integer  $q \geq 2$  otherwise.

**Theorem 2.** When  $\beta$  is known, no sequences of estimators that are consistent with  $p \in U_\beta$  exist.

**Proof.** It follows from Proposition 2 and Lemma 1, that the  $N$ -fold product measure of the mean-0 Rademacher distribution is contiguous to  $\mathbb{P}_{\beta,p}$  for all  $p \in U_\beta$ . The proof now follows from the argument given in Remark 1.  $\square$

**Remark 3.** One can consider an extension of the model (3) by adding an external magnetic field parameter  $h$  as follows:

$$\mathbb{P}_{\beta,p}(\mathbf{x}) := \frac{\exp\{\beta N \bar{x}^p + h N \bar{x}\}}{Z_N(\beta, p)} \quad \text{for } \mathbf{x} \in \{-1, 1\}^N \tag{4}$$

It has been shown in [9] that the consistent estimation of  $h$  is always possible when  $\beta$  is known. Proposition 2 shows that this is not the case for estimating  $p$ , which is impossible if  $\beta^*(p) > \beta$ . This inestimability region also coincides with that for estimating  $\beta$  when  $p$  is known (see [10]).

Now, we turn our attention to the set  $L_\beta$ , and we break it down to the following two parts:

$$L_\beta^1 := \{p \in L_\beta : \exists q \in L_\beta, q \neq p \text{ such that } m_*(\beta, q) = m_*(\beta, p)\} \quad \text{and} \quad L_\beta^2 := L_\beta \setminus L_\beta^1.$$

where  $m_*(\beta, q)$  is defined as the largest non-negative maximizer of  $H_{\beta,q}$ .

**Theorem 3.** When  $\beta$  is known, there does not exist any sequence of estimators that is consistent for  $p \in L_\beta^1$ .

**Proof.** Once again, from Lemma 1, we know that the  $N$ -fold product of the Rademacher distribution with a mean of  $m_*(\beta, p)$  is contiguous to both  $\mathbb{P}_{\beta,p}$  and  $\mathbb{P}_{\beta,q}$ , where  $p \in L_\beta^1$  and  $p \neq q \in L_\beta$  is such that  $m_*(\beta, q) = m_*(\beta, p)$ . Clearly,  $q \in L_\beta^1$ . Once again, the rest of the proof follows from the arguments of Remark 1.  $\square$

We now show that it is possible to estimate  $p$  consistently on the set  $L_\beta^2$ , if  $\beta$  is known. Towards this, let us define the following estimator of  $p$ :

$$\hat{p}(\beta, \delta) := \arg \min_{2 \leq q \leq \log_{|\bar{X}|-1}(2\beta + \delta) + 2, \beta^*(q) < \beta} |m_*(\beta, q)^2 - \bar{X}^2| \tag{5}$$

The intuition behind constructing our estimator, is that  $\bar{X}^2 \xrightarrow{P} m_*(\beta, p)^2$  as  $N \rightarrow \infty$ ; hence, the distance between  $\bar{X}^2$  and  $m_*(\beta, q)^2$  is expected to be minimized at  $q = p$  for large  $N$ . The next theorem shows that with a high probability, this is indeed the case.

**Theorem 4.** For every  $\delta > 0$  and  $p \in L_\beta^2$ , we have

$$\mathbb{P}_{\beta,p}(\hat{p}_{\beta,\delta} = p) = 1 - e^{-CN}$$

for some constant  $C = C_{\beta,p,\delta} > 0$  not depending on  $N$ .

**Proof.** Suppose that  $X \sim \mathbb{P}_{\beta,p}$  for some  $p \in L_\beta^2$ . It follows from Lemma A.1 in [10] and the proof of Proposition 1 that as  $q \rightarrow \infty$ ,  $m_*(\beta, q) \rightarrow 1$  if  $\beta \geq \log 2$  and  $m_*(\beta, q) \rightarrow 0$  otherwise. Hence, in any case,  $m_*(\beta, p) \in (0, 1)$  is not an accumulation point of the sequence  $\{m_*(\beta, q)\}_{q \geq 2}$ . This shows that there exists  $\varepsilon > 0$ , such that  $|m_*(\beta, q)^2 - m_*(\beta, p)^2| > \varepsilon$  for all  $q \in L_\beta$  not equal to  $p$ .

Now, it follows from the proofs of Lemma 3.1 and Lemma 3.3 in [9] that

$$\mathbb{P}_{\beta,p}(|m_*(\beta, p)^2 - \bar{X}^2| > \frac{\varepsilon}{3}) \leq e^{-C_1 N}$$

for some constant  $C_1 > 0$  not depending on  $N$ . Also, it follows from the proofs of Lemma 3.1 and Lemma 3.3 in [9] and the proof of Proposition 1, that

$$\mathbb{P}_{\beta,p}(\bar{X}^2 < (2\beta + \delta)^{2/(2-p)}) \leq e^{-C_2 N} \quad \text{i.e.} \quad \mathbb{P}_{\beta,p}(p > \log_{|\bar{X}|-1}(2\beta + \delta) + 2) \leq e^{-C_2 N}$$

for some constant  $C_2 > 0$  not depending on  $N$ . Define  $E_1 := \{|m_*(\beta, p)^2 - \bar{X}^2| \leq \varepsilon/3\}$  and  $E_2 := \{p \leq \log_{|\bar{X}|-1}(2\beta + \delta) + 2\}$ . It is clear that  $\hat{p}(\beta, \delta) = p$  on the event  $E_1 \cap E_2$ . Theorem 4 now follows, as  $\mathbb{P}(E_1^c \cup E_2^c) \leq e^{-C_1 N} + e^{-C_2 N}$ .  $\square$

**Remark 4.** One can tune the parameter  $\delta > 0$  in (5) to allow for an optimal number of integers over which the minimization is to be performed, in order to ensure proper convergence of  $\hat{p}(\beta, \delta)$ . Also, in case  $\beta < \log 2$ , eventually  $\beta^*(q) > \beta$ , so one just stops at the largest  $q$ , for which  $\beta^*(q) < \beta$  in the minimization (5).

In the reverse problem of estimating  $\beta$  with known  $p$ , consistent estimation is possible for all  $\beta > \beta^*(p)$  (see [9]). In contrast, a further challenging inestimability region  $L_\beta^1$  arises in our problem of estimating  $p$  with known  $\beta$ . The next proposition shows that for almost all  $\beta > 0$ , the set  $L_\beta^1$  is actually empty, i.e., the entire region  $L_\beta$  is estimable.

**Proposition 3.** There exists a countable set  $\Xi \subset (0, \infty)$ , such that  $L_\beta^1 = \emptyset$  for  $\beta \notin \Xi$ .

The proof of Proposition 3 is given in Appendix B. However, in the proof we provide an exact enumeration of one countable set  $\Lambda$  satisfying Proposition 3, namely:

$$\Lambda := \{s_{p,q} : q > p \geq 2 \text{ are integers}\} \quad \text{where} \quad s_{p,q} := \frac{1}{p} \left(\frac{p}{q}\right)^{\frac{1-p}{q-p}} \tanh^{-1} \left( \left(\frac{p}{q}\right)^{\frac{1}{q-p}} \right),$$



our analysis did not allow us to dig in any further. In particular, the following question is open:

**Open Problem:** Give an exact enumeration of the minimal countable set  $\Xi_0 = \{\beta > 0 : L_\beta^1 \neq \emptyset\}$  satisfying Proposition 3.

Note that if  $\beta = s_{p,q}$  for some integers  $q > p \geq 2$ , then  $m := (p/q)^{1/(q-p)}$  is a stationary point of both  $H_{\beta,p}$  and  $H_{\beta,q}$ . We can actually say that  $\beta \in \Xi_0$  if this stationary point turns out to be a global maximizer of both these functions. However, checking the last condition will likely involve more intricate analysis, and is left as an open question.

**Funding:** This research was funded by the FoS Tier 1 grant WBS A-8001449-00-00 and the FRC Tier 1 grant WBS A-8002932-00-00.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** No data were used in this work.

**Acknowledgments:** The author thanks Paul Switzer for bringing to notice the problem of estimating  $p$ , and Luc Devroye for several interesting discussions related to this problem.

**Conflicts of Interest:** The authors declare no conflicts of interest.

### Appendix A. Proof of Proposition 1

To show that the sequence  $m_p \rightarrow 1$  as  $p \rightarrow \infty$ , note that with a Taylor expansion, we have the following:

$$H_{\beta,p}(x) \leq \beta x^p - \frac{x^2}{2} \quad \text{for all } \beta > 0.$$

Hence,  $H_{\beta,p}(x) < 0$  for all  $x < (2\beta)^{1/(2-p)}$ , and so, if  $\beta > \beta^*(p)$ , then  $m_*(\beta, p) \geq (2\beta)^{1/(2-p)}$ . Thus, for every  $\beta \in (\beta^*(p), \beta^*(p) + \varepsilon)$ , we have  $m_*(\beta, p) > (2(\beta^*(p) + \varepsilon))^{1/(2-p)}$ . As  $\varepsilon > 0$  is arbitrary, we thus have:

$$m_p \geq (2\beta^*(p))^{1/(2-p)} \tag{A1}$$

It is now easy to see that the right-hand side of (A1) converges to 1 as  $p \rightarrow \infty$ , whereas the left-hand side is bounded above by 1. This proves our claim.

Now, as the sequence  $\{m_p\}_{p \in D}$  is convergent to its supremum, it can be arranged in increasing order as  $m_{p_1} \leq m_{p_2} \leq \dots < 1$ . Also, by (A1) and Lemma A.1 in [10],  $m_{p_1} \geq (\log 4)^{1/(2-p_1)} > 0$ . Suppose that  $(\beta, p) \in \Theta_m$  for some  $m > 0$ . It follows from [9] that  $\beta > \beta^*(p)$ . By Lemma 11 in [28], the function  $m_*(\beta, p)$  is strictly increasing in  $\beta$  on  $(\beta^*(p), \infty)$ ; hence,

$$m = m_*(\beta, p) > m_p. \tag{A2}$$

It follows from (A2), that  $\Theta_m = \emptyset$  for  $m \in (0, m_{p_1}]$ .

Let us now assume that  $m \in (m_{p_k}, m_{p_{k+1}}]$ . If  $(\beta, p) \in \Theta_m$ , then, by (A2), we have  $m_{p_{k+1}} > m_p$ ; hence,  $p \in \{p_1, \dots, p_k\}$ . Conversely, suppose that  $p = p_i$  for some  $1 \leq i \leq k$ . Then,  $m > m_p$ ; hence, there exists  $\beta_0 > \beta^*(p)$ , such that  $m > m_*(\beta_0, p)$ . By Lemma B.3 in [28], we know that  $\lim_{\beta \rightarrow \infty} m_*(\beta, p) = 1$ , which implies the existence of  $\beta_1 > \beta^*(p)$ , such that  $m < m_*(\beta_1, p)$ . Finally, Lemma B.3 in [28] also states that the function  $m_*(\beta, p)$  is continuous in  $\beta$  on  $(\beta^*(p), \infty)$ , which in view of the intermediate value theorem, implies the existence of  $\beta \in (\beta_0, \beta_1)$ , such that  $m = m_*(\beta, p)$ , i.e.,  $(\beta, p) \in \Theta_m$ . This completes the proof of Proposition 1.  $\square$



## Appendix B. Proof of Proposition 3

For every pair of integers  $q > p \geq 2$ , define:

$$s_{p,q} := \frac{1}{p} \left( \frac{p}{q} \right)^{\frac{1-p}{q-p}} \tanh^{-1} \left( \left( \frac{p}{q} \right)^{\frac{1}{q-p}} \right),$$

and let  $\Lambda := \{s_{p,q} : q > p \geq 2 \text{ are integers}\}$ . Clearly,  $\Lambda$  is a countable set. Suppose that  $L_\beta^1 \neq \emptyset$ . Then, there exist  $q > p \geq 2$  such that both the functions  $H_{\beta,p}$  and  $H_{\beta,q}$  are maximized at the same point  $m \in (0, 1)$ . Hence, we have  $H'_{\beta,q}(m) = H'_{\beta,p}(m) = 0$ , i.e.,

$$\beta q m^{q-1} = \beta p m^{p-1} = \tanh^{-1}(m).$$

From the first equality, we have  $m = (p/q)^{1/(q-p)}$ , and from the last equality, we have:

$$\beta = \frac{\tanh^{-1}(m)}{p m^{p-1}} = s_{p,q} \in \Lambda.$$

Hence,  $L_\beta^1 = \emptyset$  if  $\beta \notin \Lambda$ , which completes the proof of Proposition 3.  $\square$

## References

1. Ising, E. Beitrag zur theorie der ferromagnetismus. *Z. für Phys.* **1925**, *31*, 253–258. [\[CrossRef\]](#)
2. Geman, S.; Graffigne, C. Markov random field image models and their applications to computer vision. In Proceedings of the International Congress of Mathematicians, Berkeley, CA, USA, 3–11 August 1986; pp. 1496–1517.
3. Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 2554–2558. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Banerjee, S.; Carlin, B.P.; Gelfand, A.E. *Hierarchical Modeling and Analysis for Spatial Data*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2014.
5. Green, P.J.; Richardson, S. Hidden Markov models and disease mapping. *J. Am. Stat. Assoc.* **2002**, *97*, 1055–1070. [\[CrossRef\]](#)
6. Aslanov, L.A. Crystal symmetry and atomic interactions model. *Comput. Math. Appl.* **1988**, *16*, 443–451. [\[CrossRef\]](#)
7. Barra, A. Notes on ferromagnetic p-spin and REM. *Math. Methods Appl. Sci.* **2009**, *32*, 783–797. [\[CrossRef\]](#)
8. Derrida, B. Random-energy model: Limit of a family of disordered models. *Phys. Lett.* **1980**, *45*, 79. [\[CrossRef\]](#)
9. Mukherjee, S.; Son, J.; Bhattacharya, B.B. Phase transitions of the maximum likelihood estimates in the p-spin Curie-Weiss model. *Bernoulli* **2025**, *31*, 1502–1526. [\[CrossRef\]](#)
10. Mukherjee, S.; Son, J.; Bhattacharya, B.B. Estimation in tensor Ising models. *Inf. Inference J. IMA* **2022**, *11*, 1457–1500. [\[CrossRef\]](#)
11. Ananikyan, L.N. The hexagonal recursive approximation with multisite-interaction Ising model for the solid and fluid  $^3\text{He}$  system. *Int. J. Mod. Phys. B* **2007**, *21*, 755–772. [\[CrossRef\]](#)
12. Bergamaschi, T.R.; Menke, T.; Banner, W.P.; Paolo, A.D.; Weber, S.J.; Hirjibehedin, C.F.; Kerman, A.J.; Oliver, W.D. Distinguishing multi-spin interactions from lower-order effects. *Phys. Rev. Appl.* **2022**, *18*, 044018. [\[CrossRef\]](#)
13. Contucci, P.; Mingione, E.; Osabutey, G. Limit theorems for the cubic mean-field Ising model. *arXiv* **2023**, arXiv:2303.14578. [\[CrossRef\]](#)
14. Eichelsbacher, P. Stein's method and a cubic mean-field model. *arXiv* **2024**, arXiv:2404.07587. [\[CrossRef\]](#)
15. Gardner, E. Spin glasses with p-spin interactions. *Nucl. Phys. B* **1985**, *257*, 747–765. [\[CrossRef\]](#)
16. Mukherjee, S.; Son, J.; Bhattacharya, B. Fluctuations of the magnetization in the p-Spin Curie-Weiss model. *Communications Math. Phys.* **2021**, *387*, 681–728. [\[CrossRef\]](#)
17. Osabutey, G. Phase properties of the mean-field Ising model with three-spin interaction. *J. Math. Phys.* **2024**, *65*, 061901. [\[CrossRef\]](#)
18. Sherrington, D.; de Almeida, J.R.L. Beyond Ising SK. *arXiv* **2022**, arXiv:2208.07309.
19. Bhowal, S.; Mukherjee, S. Limit theorems and phase transitions in the tensor Curie-Weiss Potts model. *arXiv* **2023**, arXiv:2307.01052.
20. Bhowal, S.; Mukherjee, S. Rates of convergence of the magnetization in the tensor Curie-Weiss Potts model. *J. Stat. Phys.* **2025**, *192*, 2. [\[CrossRef\]](#)
21. Wu, F.Y. The Potts model. *Rev. Mod. Phys.* **1982**, *54*, 235–268. [\[CrossRef\]](#)
22. Bhattacharya, B.B.; Mukherjee, S. Inference in Ising models. *Bernoulli* **2018**, *24*, 493–525. [\[CrossRef\]](#)
23. Chatterjee, S. Estimation in spin glasses: A first step. *Ann. Stat.* **2007**, *35*, 1931–1946. [\[CrossRef\]](#)

24. Ghosal, P.; Mukherjee, S. Joint estimation of parameters in Ising model. *Ann. Stat.* **2020**, *18*, 785–810. [[CrossRef](#)]
25. Comets, F.; Gidas, B. Asymptotics of Maximum Likelihood Estimators for the Curie-Weiss Model. *Ann. Stat.* **1991**, *19*, 557–578. [[CrossRef](#)]
26. Mukherjee, S. Limit Theorems for Dependent Combinatorial Data, with Applications in Statistical Inference. *arXiv* **2021**, arXiv:2108.12233v2.
27. Agliar, E.; Barra, A.; Camboni, F. Notes on ferromagnetic diluted p-spin model. *Rep. Math. Phys.* **2011**, *68*, 1–22. [[CrossRef](#)]
28. Mukherjee, S.; Ghosh, S.; Son, J.; Mukherjee, S. Efficient estimation in tensor Curie-Weiss and Erdős-Rényi Ising models. *Electron. J. Stat.* **2024**, *18*, 2405–2449. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.