EFFICIENT ESTIMATION IN TENSOR ISING MODELS

SOMABHA MUKHERJEE, JAESUNG SON, SWARNADIP GHOSH, AND SOURAV MUKHERJEE

Abstract. The tensor Ising model is a discrete exponential family used for modeling binary data on networks with not just pairwise, but higher-order dependencies. A particularly important class of tensor Ising models are the tensor Curie-Weiss models, where all tuples of nodes of a particular order interact with the same intensity. The maximum likelihood estimator (MLE) is not explicit in this model, due to the presence of an intractable normalizing constant in the likelihood, and a computationally efficient alternative is to use the maximum pseudolikelihood estimator (MPLE). In this paper, we show that the MPLE is in fact as efficient as the MLE (in the Bahadur sense) in the 2-spin model, and for all values of the null parameter above $\log 2$ in higher-order tensor models. Even if the null parameter happens to lie within the very small window between the threshold and $\log 2$, they are equally efficient unless the alternative parameter is large. Therefore, not only is the MPLE computationally preferable to the MLE, but also theoretically as efficient as the MLE over most of the parameter space. Our results extend to the more general class of Erdős-Rényi hypergraph Ising models, under slight sparsities too.

Keywords. Tensor Ising model, Curie-Weiss model, Erdős-Rényi model, efficiency, maximum likelihood estimator, maximum pseudolikelihood estimator.

1. Introduction

With the ever increasing demand for modeling dependent network data in modern statistics, there has been a noticeable rise in the necessity for introducing appropriate statistical frameworks for modeling dependent data in the recent past. One such useful and mathematically tractable model which was originally coined by physicists for describing magnetic spins of particles, and later used by statisticians for modeling dependent binary data, is the Ising model [34]. It has found immense applications in diverse places such as image processing [22], neural networks [32], spatial statistics [7], disease mapping in epidemiology [25], structure detection [16] and property testing [42].

The Ising model is a discrete exponential family on the set of all binary tuples of a fixed length, with sufficient statistic given by a quadratic form, designed to capture pairwise dependence between the binary variables, arising from an underlying network structure. However, in most real-life scenarios, pairwise interactions are not enough to capture all the complex dependencies in a network data. For example, the behavior of an individual in a peer group depends not just on pairwise interactions, but is a more complex function of higher order interactions with colleagues. Similarly, in physics, 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 useful framework for capturing such higher order dependencies is the p-tensor Ising model [39], where the quadratic interaction term in the sufficient statistic is replaced by a multilinear polynomial of degree $p \geq 2$. Although constructing consistent estimates of the natural parameter in general p-tensor Ising models is possible [39], more exact inferential tasks such as constructing confidence intervals and hypothesis testing is not possible, unless one imposes additional constraints on the underlying network structure. One such useful structural assumption is that all p-tuples of nodes in the underlying network interact, and that too with the same intensity. The corresponding model is called the p-tensor Curie-Weiss model [38], which is a discrete exponential family on the hypercube $\{-1, 1\}^n$, with probability mass function given by:

$$\mathbb{P}_{\beta,p}(\boldsymbol{x}) := \frac{\exp\left\{\beta n^{1-p} \sum_{1 \leq i_1,\ldots,i_p \leq n} x_{i_1} \ldots x_{i_p}\right\}}{2^n Z_n(\beta, p)} \quad \text{for } \boldsymbol{x} \in \{-1, 1\}^n . \qquad (1.1)$$

Here $Z_n(\beta, p)$ is a normalizing constant required to ensure that $\sum_{\boldsymbol{x} \in \{-1,1\}^n} \mathbb{P}_{\beta,p}(\boldsymbol{x}) = 1$, and $\beta \geq 0$. It is precisely this inexplicit normalizing constant $Z_n(\beta, p)$, that hinders estimation of the parameter β using the maximum likelihood (ML) approach. Although the ML estimator can still be computed in $O(n)$ time in the Curie-Weiss model (since the probability mass function (1.1) is actually a function of the sum $\sum_{i=1}^{n} x_i$ which can take $2n + 1$ values), in even slightly more general models, for example the Erdős-Rényi Ising model (3.1), this is not true, and the ML estimator is computationally infeasible.

An extremely useful approach in the literature to circumvent this issue, is the concept of *maximum pseudolikelihood* (MPL) estimation, which was introduced by Besag in the context of spatial stochastic data with both lattice and non-lattice interactions [10, 11], and is based on computing explicit conditional distributions. This approach was later applied by Chatterjee to parameter estimation in Ising models (for $p = 2$) under general network interactions [18]. To elaborate, the MPL estimate is obtained by maximizing the *pseudolikelihood function*:

$$\hat{\beta}_{\mathrm{MPL}} := \arg \max_{\beta \in \mathbb{R}} \prod_{i=1}^{n} \mathbb{P}_{\beta,p}\left(X_i | (X_j)_{j \neq i}\right) ,$$

where $\boldsymbol{X} = (X_1, \ldots, X_n)$ is simulated from the model (1.1). It should be quite clear that the conditional distributions in the expression of the pseudolikelihood function makes it free of the inexplicit normalizing constant $Z_n(\beta, p)$, thereby making the MPL estimator explicit and computationally feasible. In fact, methods as simple as a grid search can be applied to compute the MPL estimator.

A natural question is thus, to what extent does this computationally feasible MPL approach inherit the desirable theoretical properties of the ML approach? Quite surprisingly, in spite of being just a proxy for the exact ML estimator, the MPL estimator in fact satisfies almost all the theoretical guarantees of the former. We list some of these properties below:

(1) Both the ML and the MPL estimators are $\sqrt{n}$-consistent in the so called *low temperature regime* (high values of the parameter β). This was first established in [18] for the $p = 2$ case, and later extended to tensor Ising models ($p > 2$) in [39]. More precisely, it is shown in [39] and [38] that there exists $\beta^*(p) > 0$, such that for all $\beta > \beta^*(p)$, both $\sqrt{n}(\hat{\beta}_{\mathrm{MPL}} - \beta)$ and $\sqrt{n}(\hat{\beta}_{\mathrm{ML}} - \beta)$ are tight, where $\hat{\beta}_{\mathrm{ML}}$ is the Maximum Likelihood (ML) estimator of β. Further, consistent estimation (and consistent testing) is impossible in the regime $[0, \beta^*(p))$. The initial few values of the threshold $\beta^*(p)$ are given by $\beta^*(2) = 0.5, \beta^*(3) = 0.672$ and $\beta^*(4) = 0.689$.

(2) Above the estimation threshold $\beta^*(p)$, both $\sqrt{n}(\hat{\beta}_{\mathrm{MPL}} - \beta)$ and $\sqrt{n}(\hat{\beta}_{\mathrm{ML}} - \beta)$ converge weakly to the same normal distribution [38]. This asymptotic normality can in fact be used to construct confidence intervals with asymptotically valid coverage probabilities for the parameter β in presence of an external magnetic field term in the model 1.1 (see Section 5 in [38]).

(3) As a consequence of the last point, both the ML and MPL estimates have the same asymptotic variance everywhere above the threshold. In fact, this asymptotic variance equals the limiting inverse Fisher information of the model, so both the estimates saturate the Cramer-Rao information lower bound of the model in this regime [38].

What happens if we wish to perform a hypothesis testing of the natural parameter using these two estimators? In particular, which of the two tests would require a smaller sample size for achieving significance at a given level? The correct way to address this issue is to use the notion of *Bahadur efficiency*. In this paper, we demonstrate that the MPL estimator is in fact as Bahadur efficient as the ML estimator everywhere in the classical 2-spin ($p = 2$) model, and for *most* values of the null and the alternative parameters in higher-order ($p \geq 3$) tensor models. When $p \geq 3$, a two-layer phase transition phenomenon is observed, depending on the magnitudes of the null and the alternative parameters. To elaborate, in this case, the only regime where the MPL estimator is less Bahadur efficient than the ML estimator, is a very small window of variation of the null parameter between $\beta^*(p)$ and $\log 2$, and that too, for large values of the alternative parameter only. Moreover, this small window $(\beta^*(p), \log 2)$ shrinks to the empty set as $p \to \infty$. This shows that the optimal sample size requirements for the tests based on the ML and the MPL estimators to achieve significance, are identical in the 2-spin case and same over most of the parameter space in the higher-order tensor case.

In his seminal paper [6], Bahadur introduced the concept of *slope* of a test statistic to calculate the minimum sample size required to ensure its significance at a given level. The setting considered in [6] involved i.i.d. samples coming from a certain parametric family, and the goal was to detect the minimum sample size $N(\delta)$ required, so that a test T_n (function of the samples) becomes (and remains) significant at level δ for all $n \geq N(\delta)$, i.e. the p-value corresponding to T_n becomes (and remains) bounded by δ for all $n \geq N(\delta)$. If

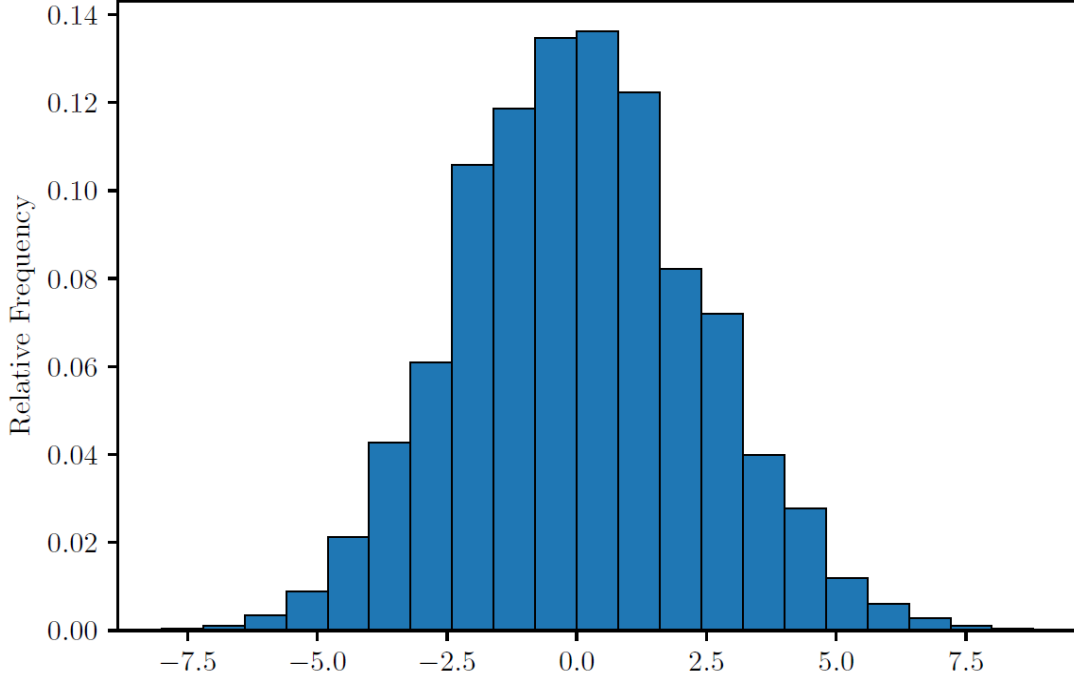FIGURE 1. Histogram of $\sqrt{n}(\hat{\beta}_n(\boldsymbol{X}) - \beta)$, where $\hat{\beta}_n(\boldsymbol{X})$ is the MPL estimator in the 4-tensor Curie-Weiss model at $\beta = 0.75 > \beta^*(4) \approx 0.689$ (above the estimation threshold), $n = 20,000$ [39].

one considers testing a simple null hypothesis $H_0 : \theta = \theta_0$, then the above discussion may be quantified by defining:

$$N(\delta) := \inf \left\{ N \geq 1 : \sup_{n \geq N} L_n \leq \delta \right\} ,$$

where $L_n := 1 - F_{T_n,\theta_0}(T_n)$ and F_{T_n,θ_0} is the cumulative distribution function of T_n under $\mathbb{P}_{\theta_0}$. The p-value L_n typically converges to 0 exponentially fast with probability 1 under alternatives $\mathbb{P}_\theta$ for $\theta > \theta_0$, and this rate is often an indication of the asymptotic efficiency of T_n against θ [1, 2, 3, 4, 5]. In particular, if we have the following $\mathbb{P}_\theta$-almost surely:

$$\frac{1}{n} \log L_n \to -\frac{1}{2}c(\theta) \quad \text{as } n \to \infty , \tag{1.2}$$

then one can easily verify that (see Proposition 8 in [6]) $N(\delta) \sim -2\log(\delta)/c(\theta)$ as $\delta \to 0$. $c(\theta)$ is called the Bahadur slope of T_n at θ. However, as mentioned in [6], it is in general a non-trivial problem to determine the existence of the Bahadur slope in (1.2), and to evaluate it. This issue is addressed in two steps in [6], where it is shown that if T_n satisfies the following two conditions:

(1) For every alternative θ, $n^{-1/2}T_n \to b(\theta)$ as $n \to \infty$ under $\mathbb{P}_\theta$ with probability 1, for some parametric function b defined on the alternative space,

(2) $n^{-1} \log[1 - F_{T_n,\theta_0}(n^{1/2}t)] \to -f(t)$ as $n \to \infty$ for every $t > 0$ in an open interval which includes each value of b, where f is a continuous function on the interval, with $0 < f < \infty$,

then the Bahadur slope exists for every alternative θ, and is given by $2f(b(\theta))$ (see [6]). In this context, let us mention that if the convergence (1.2) holds in probability, then $c(\theta)$ is called the weak Bahadur slope of T_n (see [26]). Finally, if we have two competing estimators $T_{n,1}$ and $T_{n,2}$ estimating the same parameter θ, then the Bahadur asymptotic relative efficiency (ARE) is given by the ratio of their Bahadur slopes (see [26]):

$$\text{eff}(T_{n,1}, T_{n,2}; \theta) = \frac{c_1(\theta)}{c_2(\theta)} .$$

The Bahadur ARE is a well-known tool in the literature for comparing the performance of two estimators in a wide variety of contexts. Gyorfi et al. [28] addressed the problem of comparing the efficiencies of information-divergence-type statistics for testing the goodness of fit. They claim that the Pitman approach is too weak to detect sufficiently sharply the differences in efficiency of these statistics, and instead, focussed their attention on the Bahadur efficiency. Harremoës and Vajda [29] show that in the problem of testing the uniformity of a distribution, the information divergence statistic is more efficient in the Bahadur sense than any power divergence statistic of order $\alpha > 1$. The same authors show in their paper [30] that any two Rényi entropies of different orders $\in (0, 1]$ are equally Bahadur efficient. Huang [33] uses Bahadur efficiency as a measure of performance in the small sample universal hypothesis testing problem, and mentions that in the large sample problem where the number of possible outcomes is at most of the order of the number of samples, the connection between the error exponent and Bahadur efficiency has been studied in [43, 29]. Keziou and Regault [36] compare the performances of independence tests derived by means of dependence thresholding in a semiparametric context, in terms of the Bahadur ARE. Applications of the Bahadur efficiency in the contexts of nonparametric tests for independence and separate hypothesis testing can also be found in [8, 9] and [44], respectively.

In this paper, we compare the estimators $\hat{\beta}_{\text{MPL}}$ and $\hat{\beta}_{\text{ML}}$ in the tensor Curie-Weiss model (1.1) in terms of the Bahadur ARE. This requires deriving the weak Bahadur slopes of both $\hat{\beta}_{\text{MPL}}$ and $\hat{\beta}_{\text{ML}}$ in the model (1.1), which will in turn, enable one to compute the Bahadur ARE of either of these two estimators against some other reference estimator. Similar results have been derived in [17] in the context of Markov random fields on lattices, and in [27] in the context of d-dimensional nearest neighbor isotropic Ising models, but to the best of our knowledge, this is the first such work on tensor Curie-Weiss (and Erdős-Rényi Ising) models. Our basic tools are some recent results on large deviation of the *average magnetization* $\overline{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$ in the Curie-Weiss model, established in [37] and [45]. Throughout the rest of the paper, we will view the entries $X_1, \ldots, X_n$ of the tuple $\boldsymbol{X} \in \{-1, 1\}^n$ as *dependent samples*, and refer to the length n of $\boldsymbol{X}$ as the *sample size*
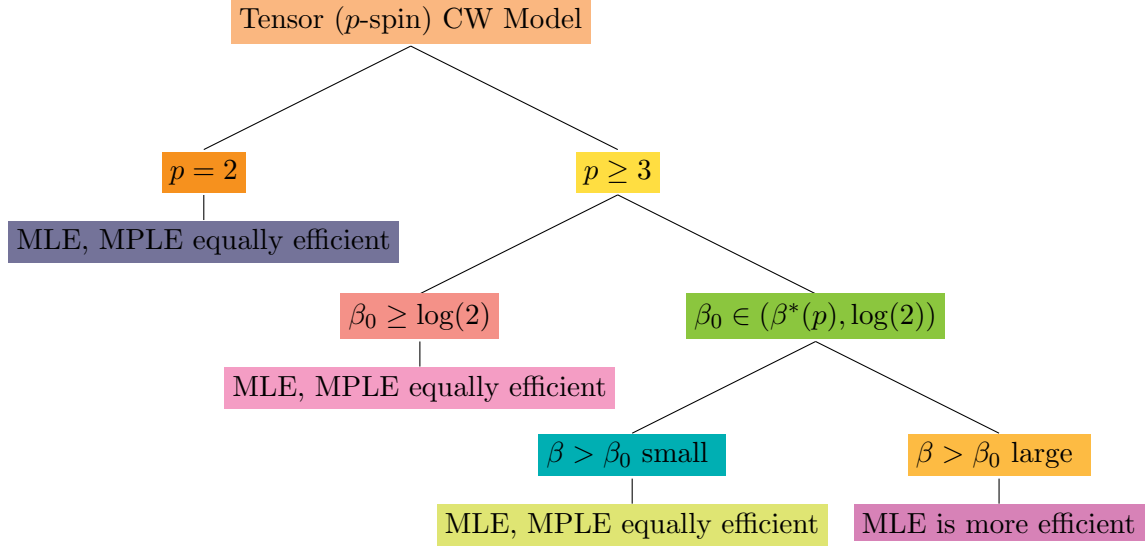
FIGURE 2. Behavior of MLE and MPLE for different values of p

(although technically speaking, we have just one multivariate sample $\boldsymbol{X}$ from the model (1.1)). One of our most interesting findings is that the relative performance of the ML and the MPL estimators based on the notion of Bahadur efficiency depends crucially on whether the model is strictly tensor or not. In other words, in the usual 2-spin ($p = 2$) Curie-Weiss model, the two estimators are indistinguishable from the Bahadur ARE perspective. This is also true if the model is strictly tensor (i.e. $p \geq 3$), but for all values of the null parameter $\beta_0 \geq \log 2$. However, in this case, the MPL estimator is strictly less efficient than the ML estimator for each value of the null parameter β_0 in $\mathcal{R} := (\beta^*(p), \log 2)$, but that too, only if the alternative parameter β is sufficiently large. This loss of Bahadur efficiency for the MPL estimator near threshold in the tensor Curie-Weiss models, can be attributed to its functional form, which derives false signal from a regime where the average magnetization $\overline{X}_n$ is very close to 0.

Our results are a bit more universal, in the sense that they extend beyond the Curie-Weiss model (1.1). They hold verbatim in the tensor Erdős-Rényi Ising model (3.1) too, which is an exponential family, with sufficient statistic given by a tensor form, the tensor being the adjacency of a directed Erdős-Rényi hypergraph with loops. Interestingly, we can even allow for slight sparsities in the underlying Erdős-Rényi hypergraph. We believe that the same results (and techniques) will also extend to Ising models on dense stochastic block model hypergraphs, and leave it open for future research.

The rest of the paper is organized as follows. In Section 2, we compare the ML and the MPL estimators in terms of their Bahadur ARE. The corresponding Bahadur slope and optimal sample size calculations necessary for the tests based on these two estimators are also provided in Section 2. Section 3 shows that these results for the tensor Curie-Weiss

model possess a somewhat universality property, i.e. they are true even for the tensor Erdős-Rényi Ising model. In Section 4, we provide numerical illustrations of our theoretical findings in various settings. In Section 5, we prove the main results in Sections 2 and 3. In Section 6, we summarize some main and interesting aspects of our results, and talk about possible directions for future research in this area. Finally, proofs of some technical results needed for showing the main theorems are given in the appendix.

1.1. **Our Contributions.** Before delving in the main results, let us summarize our contributions once again in a pointwise fashion. We enlist the results we derived below.

(1) In the classical 2-spin ($p = 2$) Curie-Weiss/Erdős-Rényi model, the ML and MPL estimators are equally Bahadur efficient.

(2) In the higher-order tensor ($p \geq 3$) Curie-Weiss/Erdős-Rényi model, the MPL estimator is equally Bahadur efficient as the ML estimator, if the null parameter is greater than or equal to $\log 2$. Even if the null parameter lies strictly between the model threshold and $\log 2$, they are equally efficient unless the alternative parameter is very large. In the last case, the MPL estimator is less Bahadur efficient than the ML estimator.

(3) For any arbitrary value of the null parameter above the threshold and any arbitrary value of the alternative parameter, a necessary and sufficient condition for the tests based on the ML and MPL estimators to be equally Bahadur efficient, is provided.

(4) The exact Bahadur slopes of the ML and MPL estimators, and the optimal sample sizes required by the tests based on these estimators to achieve significance, are derived for both the tensor Curie-Weiss and the Erdős-Rényi Ising model.

(5) Even if the the null parameter lies in the $(\beta^*(p), \log 2)$ window and the alternative parameter is very large, the MPL estimator is always at least a positive fraction as efficient as the ML estimator, an observation similar to Hodges and Lehmann's (1956) remarkable result that the Pitman ARE of Wilcoxon's test with respect to Student's T-test, under location alternatives, never falls below 0.864, despite the former being non-parametric and exactly distribution-free for all sample sizes. The same conclusion holds for fixed values of the alternative above threshold, provided the null parameter is bounded away from the threshold.

(6) The Bahadur slopes and the optimal sample size requirements for both the ML and MPL estimators remain unchanged, if one moves from the tensor Curie-Weiss model to the tensor Erdős-Rényi model. This indicates a possible universality of our results to Ising models on dense, random block hypergraphs.

Although the MPL estimator shows almost all the desirable efficiency properties of the ML estimator, in order to provide a completely unbiased and clear picture, we must mention two aspects in which the latter beats the former:

(1) For $p \geq 3$ and every fixed value of the alternative parameter, the Bahadur ARE of the MPL estimator with respect to the ML estimator approaches 0 as the null parameter approaches the threshold from the right. This is because, the expression for asymptotic optimal sample size (2.2) for significance of the test based on the MPL estimator approaches ∞, but the corresponding asymptotic optimal sample size for the test based on the ML estimator remains bounded.

(2) In the $p \geq 3$ case, as long as the null parameter lies strictly between the threshold and $\log 2$, the asymptotic optimal sample size required by the MPL estimator stabilizes at a fixed value after the alternative parameter exceeds a certain finite value, unlike that of the ML estimator. This is an undesirable property of the MPL estimator, since the sample size requirement does not decrease with increase in the separation between the null and alternative parameters, above a certain limit. However, as already mentioned above, the asymptotic optimal sample size for the MPL estimator exceeds that of the ML estimator by a bounded fraction only, as $\beta \to \infty$.

2. THEORETICAL RESULTS FOR THE TENSOR CURIE-WEISS MODEL

In this section we compare the MPL and the ML estimates in the tensor Curie-Weiss model (1.1) in terms of their Bahadur ARE. The ML estimator $\hat{\beta}_{\mathrm{ML}}$ does not have an explicit form, but it is shown in [39] that the MPL estimator is given by:

$$\hat{\beta}_{\mathrm{MPL}} = \begin{cases} p^{-1}\overline{X}_n^{1-p}\tanh^{-1}(\overline{X}_n) & \text{if } \overline{X}_n \neq 0, \\ 0 & \text{if } \overline{X}_n = 0. \end{cases}$$

Furthermore, it is shown in [38] and [39] that both the ML and MPL estimators have the same asymptotic normal distribution:

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{D} N\left(0, -\frac{H''_{\beta,p}(m_*(\beta,p))}{p^2 m_*(\beta,p)^{2p-2}}\right),$$

for all $\beta > \beta^*(p)$, where $\hat{\beta}$ is either $\hat{\beta}_{\mathrm{ML}}$ or $\hat{\beta}_{\mathrm{MPL}}$,

$$H_{\beta,p}(x) := \beta x^p - \frac{1}{2}\left\{(1+x)\log(1+x) + (1-x)\log(1-x)\right\} \quad \text{for } x \in [-1,1], \qquad (2.1)$$

$m_*(\beta,p)$ is the unique positive global maximizer of $H_{\beta,p}$, and

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

A few initial values of the threshold $\beta^*(p)$ are $\beta^*(2) = 0.5, \beta^*(3) \approx 0.672$ and $\beta^*(4) \approx 0.689$. The exact value of $\beta^*(p)$ is in general inexplicit, but $\beta^*(p) \uparrow \log 2$ as $p \to \infty$ (see Lemma A.1 in [39]).

In this paper, we will consider testing the hypothesis

$$H_0 : \beta = \beta_0 \qquad \text{vs} \qquad H_1 : \beta > \beta_0$$

for some known $\beta_0 > \beta^*(p)$. The most powerful test for this hypothesis in model (1.1) is based on the sufficient statistic $\overline{X}_n$, and its asymptotic power is derived in [13]. Clearly, one can think of using the statistic $T_n := \sqrt{n}(\hat{\beta} - \beta_0)$ for testing the above hypotheses, where $\hat{\beta}$ is either $\hat{\beta}_{\mathrm{ML}}$ or $\hat{\beta}_{\mathrm{MPL}}$, and large values of T_n will denote significance.

We now state the main result in this paper about the Bahadur slopes of the tests based on the MPL and ML estimators, and the minimum sample size required to ensure their significance. Towards this, we define a function $\eta_p : [-1, 1] \mapsto \mathbb{R}$ as:

$$\eta_p(t) = \begin{cases} p^{-1} t^{1-p} \tanh^{-1}(t) & \text{if } t \neq 0, \\ 0 & \text{if } t = 0. \end{cases}$$

Theorem 1. The Bahadur slopes of $\hat{\beta}_{\mathrm{MPL}}$ and $\hat{\beta}_{\mathrm{ML}}$ for the model (1.1) at an alternative β are respectively given by:

$$c_{\hat{\beta}_{\mathrm{MPL}}}(\beta_0, \beta, p) = 2 \left(\sup_{x \in [-1,1]} H_{\beta_0,p}(x) - \sup_{x \in \eta_p^{-1}((\beta,\infty))} H_{\beta_0,p}(x) \right),$$

$$c_{\hat{\beta}_{\mathrm{ML}}}(\beta_0, \beta, p) = 2 \left(\sup_{x \in [-1,1]} H_{\beta_0,p}(x) - \sup_{x \, > \, m_*(\beta,p)} H_{\beta_0,p}(x) \right)$$

Consequently, the minimum sample sizes required, so that the tests $\sqrt{n}(\hat{\beta}_{\mathrm{MPL}} - \beta_0)$ and $\sqrt{n}(\hat{\beta}_{\mathrm{ML}} - \beta_0)$ become (and remain) significant at level $\delta \to 0$, are respectively given by:

$$N_{\hat{\beta}_{\mathrm{MPL}}}(\beta_0, \beta, \delta, p) \ \sim \ \frac{\log(\delta)}{\sup_{x \in \eta_p^{-1}((\beta,\infty))} H_{\beta_0,p}(x) - \sup_{x \in [-1,1]} H_{\beta_0,p}(x)},$$

$$N_{\hat{\beta}_{\mathrm{ML}}}(\beta_0, \beta, \delta, p) \ \sim \ \frac{\log(\delta)}{\sup_{x \, > \, m_*(\beta,p)} H_{\beta_0,p}(x) - \sup_{x \in [-1,1]} H_{\beta_0,p}(x)}.$$

Theorem 1 is proved in Section 5.1. Let us introduce the following notation, which will be used throughout the rest of the paper:

$$N_{\hat{\beta}_{\mathrm{MPL}}}^*(\beta_0, \beta, \delta, p) = -\frac{2\log(\delta)}{c_{\hat{\beta}_{\mathrm{MPL}}}(\beta_0, \beta, p)} \quad \text{and} \quad N_{\hat{\beta}_{\mathrm{ML}}}^*(\beta_0, \beta, \delta, p) = -\frac{2\log(\delta)}{c_{\hat{\beta}_{\mathrm{ML}}}(\beta_0, \beta, p)} . \quad (2.2)$$

Note that $N_{\hat{\beta}_{\mathrm{MPL}}}(\beta_0, \beta, \delta, p) \sim N_{\hat{\beta}_{\mathrm{MPL}}}^*(\beta_0, \beta, \delta, p)$ and $N_{\hat{\beta}_{\mathrm{ML}}}(\beta_0, \beta, \delta, p) \sim N_{\hat{\beta}_{\mathrm{ML}}}^*(\beta_0, \beta, \delta, p)$ as $\delta \to 0$, so the latter quantities can be called the *asymptotic* optimal sample sizes. The natural question at this stage, is when do the expressions for the Bahadur slope and the asymptotic optimal sample size for the MPL and the ML estimators differ? It turns out that we have two different scenarios depending on whether $p = 2$ or $p \geq 3$.

2.1. **The $p = 2$ Case:** In this case, it turns out that the Bahadur slopes and the asymptotic optimal sample sizes for the MPL and the ML estimators agree, and consequently, they are equally Bahadur efficient.

Theorem 2. Consider the model (1.1) for $p = 2$. For every $\beta > \beta_0 > \beta^*(2)$ and $\delta \in (0,1)$, we have

$$c_{\hat{\beta}_{\mathrm{MPL}}}(\beta_0, \beta, 2) = c_{\hat{\beta}_{\mathrm{ML}}}(\beta_0, \beta, 2) \quad \text{and} \quad N^*_{\hat{\beta}_{\mathrm{MPL}}}(\beta_0, \beta, \delta, 2) = N^*_{\hat{\beta}_{\mathrm{ML}}}(\beta_0, \beta, \delta, 2) .$$

Consequently, the Bahadur ARE $\mathrm{eff}(\hat{\beta}_{\mathrm{ML}}, \hat{\beta}_{\mathrm{MPL}}; \beta_0, \beta) = 1$.

Theorem 2 is proved in Section 5.2. It says that in the classical 2-spin Curie-Weiss model, one cannot distinguish the estimators $\hat{\beta}_{\mathrm{ML}}$ and $\hat{\beta}_{\mathrm{MPL}}$ based on even the Bahadur efficiency.

2.2. **The Strictly Tensor $p \geq 3$ Case:** All the interesting phenomena occur when the model (1.1) goes beyond the classical 2-spin system to the higher-order tensor ($p \geq 3$) system. In this case, a two layer phase transition is observed with respect to both the null and alternative parameters. To be precise, there exists a small window around the estimation threshold $\beta^*(p)$, such that for all values of the null parameter in this window, $\hat{\beta}_{\mathrm{MPL}}$ is strictly less Bahadur efficient than $\hat{\beta}_{\mathrm{ML}}$ provided the alternative parameter β is greater than a second threshold (which is different from $\beta^*(p)$).

Theorem 3. Consider the model (1.1) for $p \geq 3$. Two different situations arise depending on whether $\beta_0 \geq \log 2$ or $\beta_0 \in (\beta^*(p), \log 2)$.

(1) For every $\beta > \beta_0 \geq \log 2$ and $\delta \in (0,1)$, we have

$$c_{\hat{\beta}_{\mathrm{MPL}}}(\beta_0, \beta, p) = c_{\hat{\beta}_{\mathrm{ML}}}(\beta_0, \beta, p) \quad \text{and} \quad N^*_{\hat{\beta}_{\mathrm{MPL}}}(\beta_0, \beta, \delta, p) = N^*_{\hat{\beta}_{\mathrm{ML}}}(\beta_0, \beta, \delta, p) .$$

Consequently, the Bahadur ARE $\mathrm{eff}(\hat{\beta}_{\mathrm{ML}}, \hat{\beta}_{\mathrm{MPL}}; \beta_0, \beta) = 1$ in this regime.

(2) For every $\beta_0 \in (\beta^*(p), \log 2)$ and $\delta \in (0,1)$, the Bahadur slopes and asymptotic optimal sample sizes for $\hat{\beta}_{\mathrm{MPL}}$ and $\hat{\beta}_{\mathrm{ML}}$ do not agree, and $\mathrm{eff}(\hat{\beta}_{\mathrm{ML}}, \hat{\beta}_{\mathrm{MPL}}; \beta_0, \beta) > 1$, for all $\beta > \beta_0$ large enough.

Theorem 3 is proved in Section 5.3. Some more insight is obtained if one fixes the alternative $\beta > \beta^*(p)$, and looks at the behavior of the asymptotic optimal sample sizes and Bahadur ARE of the MPL and ML estimators by varying the null parameter in the small window near $\beta^*(p)$.

Theorem 4 (The fixed alternative scenario). For $p \geq 3$ and fixed $\beta > \beta^*(p)$, the set of all $\beta_0 > \beta^*(p)$ satisfying $\mathrm{eff}(\hat{\beta}_{\mathrm{ML}}, \hat{\beta}_{\mathrm{MPL}}; \beta_0, \beta) > 1$ is given by:

$$\left(\beta^*(p), \frac{I(m_*(\beta, p))}{m_*(\beta, p)^p} \right)$$

where $I(x) = \frac{1}{2}\{(1+x)\log(1+x) + (1-x)\log(1-x)\}$. Further, for every $p \geq 3$ and every fixed $\beta > \beta^*(p)$, we have:

$$\lim_{\beta_0 \to \beta^*(p)^+} N^*_{\hat{\beta}_{\mathrm{MPL}}}(\beta_0, \beta, \delta, p) = \infty \quad \text{and} \quad \lim_{\beta_0 \to \beta^*(p)^+} N^*_{\hat{\beta}_{\mathrm{ML}}}(\beta_0, \beta, \delta, p) = \frac{\log(\delta)}{H_{\beta^*(p), p}(m_*(\beta, p))} < \infty.$$

$$(2.3)$$

and hence,

$$\lim_{\beta_0 \to \beta^*(p)^+} \mathrm{eff}(\hat{\beta}_{\mathrm{ML}}, \hat{\beta}_{\mathrm{MPL}}; \beta_0, \beta) = \infty \ .$$

Theorem 4 is proved in Section 5.4.

Remark 2.1. Theorem 4 says that as the null parameter approaches the estimation threshold, the ML estimator becomes infinitely more Bahadur efficient than the MPL estimator. The reason behind the discrepancy between the efficiencies of the ML and MPL estimators near the threshold, is the functional form of the latter. For $p \geq 3$, unlike the ML estimator, the MPL estimator takes very high values if the average magnetization $\overline{X}_n$ is close to 0. This false signal coming from the average magnetization lying in a region very close to 0, leads to an increase in the null probability of the MPL estimator exceeding the observed MPL estimate, thereby inflating its p-value. This inflation occurs only in a close neighborhood of the threshold, because for lower values of the parameter β, there is a higher probability that the average magnetization $\overline{X}_n$ is small.

Remark 2.2. It follows from Lemma C.12 in [38] that

$$\sup_{x \in [-1,1]} H_{\beta_0,p}(x) = H_{\beta_0,p}(m_*(\beta_0, p)) = \Theta(\beta_0 - \beta^*(p)) \ .$$

Since $\sup_{x \in \eta_p^{-1}((\beta,\infty))} H_{\beta_0,p}(x)$ eventually becomes 0 as β_0 approaches $\beta^*(p)$ from the right, the rate at which $N^*_{\hat{\beta}_{\mathrm{MPL}}}(\beta_0, \beta, \delta, p)$ approaches ∞ as $\beta_0 \to \beta^*(p)^+$ for $p \geq 3$, is determined just by the $\sup_{x \in [-1,1]} H_{\beta_0,p}(x)$ term in the denominator of the formula for $N^*_{\hat{\beta}_{\mathrm{MPL}}}(\beta_0, \beta, \delta, p)$, and is given by $(\beta_0 - \beta^*(p))^{-1}$.

Another undesirable property of $\hat{\beta}_{\mathrm{MPL}}$ is that for every fixed value of the null parameter in the window $(\beta^*(p), \log 2)$, its asymptotic optimal sample size does not decrease further when the alternative parameter exceeds a certain value. So, no matter how large the separation between the null and the alternative are, one does not have any concession in the asymptotic sample size requirement after a certain value of the separation. This is formalized in the theorem below:

Theorem 5. For $p \geq 3$ and fixed $\beta_0 \in (\beta^*(p), \log 2)$, there exists $\underline{\beta} > 0$ such that

$$N^*_{\hat{\beta}_{\mathrm{MPL}}}(\beta_0, \beta, \delta, p) = -\frac{\log(\delta)}{H_{\beta_0,p}(m_*(\beta_0, p))} > 0 \quad \text{for all } \beta > \underline{\beta}.$$

Theorem 5 is proved in Section 5.5. It says that the asymptotic optimal sample size requirement for the MPL estimator stabilizes at a certain value once the separation between the null and the alternative is large enough, provided the null lies in the window $(\beta^*(p), \log 2)$. In this case, note that for β large enough, we have:

$$N^*_{\hat{\beta}_{\mathrm{ML}}}(\beta_0, \beta, \delta, p) = -\frac{\log(\delta)}{H_{\beta_0,p}(m_*(\beta_0, p)) - H_{\beta_0,p}(m_*(\beta, p))} < N^*_{\hat{\beta}_{\mathrm{MPL}}}(\beta_0, \beta, \delta, p).$$

It also shows that unlike the $\hat{\beta}_{\mathrm{MPL}}$, the asymptotic optimal sample size for $\hat{\beta}_{\mathrm{ML}}$ is strictly decreasing with the alternative β, for all sufficiently large values of β. This is precisely due to the presence of the $H_{\beta_0,p}(m_*(\beta,p))$ term in the denominator of $N^*_{\hat{\beta}_{\mathrm{ML}}}(\beta_0,\beta,\delta,p)$, which is strictly decreasing in β for all β large enough.

Remark 2.3. It follows from Theorem 1 that for every fixed $\beta_0 > \beta^*(p)$,

$$\lim_{\beta\to\infty} N^*_{\hat{\beta}_{\mathrm{ML}}}(\beta_0,\beta,\delta,p) = \frac{\log(\delta)}{\beta_0 - \log 2 - H_{\beta_0,p}(m_*(\beta_0,p))}\ .$$

On the other hand, Theorem 5 implies that as long as $\beta_0 \in (\beta^*(p), \log 2)$, we have:

$$\lim_{\beta\to\infty} N^*_{\hat{\beta}_{\mathrm{MPL}}}(\beta_0,\beta,\delta,p) = \frac{\log(\delta)}{-H_{\beta_0,p}(m_*(\beta_0,p))}\ .$$

Hence, for every $\beta_0 \in (\beta^*(p), \log 2)$,

$$\lim_{\beta\to\infty} \mathrm{eff}(\hat{\beta}_{\mathrm{MPL}}, \hat{\beta}_{\mathrm{ML}}; \beta_0, \beta) = \frac{H_{\beta_0,p}(m_*(\beta_0,p))}{H_{\beta_0,p}(m_*(\beta_0,p)) + \log 2 - \beta_0} > 0\ .$$

Hence, the Bahadur ARE of the MPL estimator with respect to the ML estimator is bounded away from 0 even when the null parameter is in the interval $(\beta^*(p), \log 2)$ and the alternative parameter is large. This is analogous to the remarkable result of Hodges and Lehmann (1956), which shows that the Pitman ARE of Wilcoxon's test with respect to Student's T-test, under location alternatives, never falls below 0.864, despite the former being non-parametric and exactly distribution-free for all sample sizes (see [19]).

3. Theoretical Results for the Hypergraph Erdős-Rényi Ising Model

In this section, we are going to see that all our results for the tensor Curie-Weiss model hold verbatim for the more general class of hypergraph Erdős-Rényi Ising model, where we can even introduce some sparsity. A classical survey on these and more general models of disordered ferromagnets in the physics literature can be found in [21]. This model for $p = 2$ was introduced and studied in [12], and analyzed further in [35], where a quenched central limit theorem has been proved for the magnetization in the high temperature (small β) regime, allowing for some sparsity in the underlying random graph. Of course, we will be concerned with the general p case in this section.

Let $\boldsymbol{A} := \{A_{i_1\ldots i_p}\}_{1\le i_1,\ldots,i_p\le n}$ be a collection of i.i.d. Bernoulli random variables with mean α_n. Note that $\boldsymbol{A}$ can be viewed as the adjacency tensor of a directed Erdős-Rényi hypergraph with loops. The p-tensor Erdős-Rényi Ising model in this context, is a discrete exponential family on $\{-1,1\}^n$ with probability mass function given by:

$$\mathbb{P}^*_{\beta,p}(\boldsymbol{x}) = \frac{\exp\{\beta H_n(\boldsymbol{x})\}}{2^n Z^*_n(\beta,p)} \quad (\text{for } \boldsymbol{x} \in \{-1,1\}^n)\ , \tag{3.1}$$

where

$$H_n(\boldsymbol{x}) := \alpha_n^{-1} n^{1-p} \sum_{1\le i_1,\ldots,i_p\le n} A_{i_1\ldots i_p} x_{i_1}\ldots x_{i_p}$$

denotes the Hamiltonian of the model, and $Z_n^*(\beta, p)$ is the normalizing constant. Note that we will use a $*$ superscript to denote probabilities and moments corresponding to the model (3.1). Below, we state the main result of this section:

Theorem 6. The Bahadur slopes and asymptotic minimum sample sizes of $\hat{\beta}_{\mathrm{MPL}}$ and $\hat{\beta}_{\mathrm{ML}}$ for the model (3.1) at an alternative β are respectively equal to the Bahadur slopes and asymptotic minimum sample sizes of $\hat{\beta}_{\mathrm{MPL}}$ and $\hat{\beta}_{\mathrm{ML}}$ for the model (1.1) at β.

Consequently, all the results in Section 2 for the tensor Curie-Weiss model also hold for the tensor Erdős-Rényi Ising model. Theorem 6 is proved in Section 5.6. The main approach, in a nutshell, is an approximation of the Hamiltonian and the local fields of the tensor Erdős-Rényi model by the corresponding quantities for the tensor Curie-Weiss model. In order to show that the Hamiltonian $H_n(\boldsymbol{x})$ of the hypergraph Erdős-Rényi Ising model is very close to that of the tensor Curie-Weiss model, it is enough to establish a uniform (in $\boldsymbol{x}$) concentration of $H_n(\boldsymbol{x})$ around its mean (with respect to the Erdős-Rényi measure) $\mathbb{E}H_n(\boldsymbol{x})$. Define

$$\gamma_n := 3(\alpha_n n^{p-1})^{-\frac{1}{2}} .$$

Lemma 1. Let H_n denote the Hamiltonian of the p-tensor Erdős-Rényi Ising model. Then,

$$\mathbb{P}\left(\frac{1}{n} \sup_{\boldsymbol{x}\in\{-1,1\}^n} |H_n(\boldsymbol{x}) - \mathbb{E}H_n(\boldsymbol{x})| \le 3\gamma_n \text{ for all but finitely many } n\right) = 1 .$$

Lemma 1 says that as long as the Erdős-Rényi hyperedge probability $\alpha_n \gg n^{1-p}$, the Hamiltonian $H_n(\boldsymbol{x})$ concentrates around its mean (when scaled by a factor of $1/n$) uniformly in $\boldsymbol{x}$. The proof of Lemma 1 is given in Appendix C. The following result is a corollary of the proof of Lemma 1, which says that not only are the Hamiltonians of the two models close, but their local fields $m_i^{(n)}(\boldsymbol{x})$ (defined below) are also close, uniformly over all $1 \le i \le n$ and all $\boldsymbol{x} \in \{-1, 1\}^n$. It will be useful in deriving the Bahadur slope of the MPL estimator.

Corollary 1. For each $1 \le i \le n$, define $m_i^{(n)}(\boldsymbol{x}) := \alpha_n^{-1} n^{1-p} \sum_{(i_2,...,i_p)\in[n]^{p-1}} A_{ii_2...i_p} x_{i_2} \ldots x_{i_p}$. Then, we have:

$$\mathbb{P}\left(\sup_{1\le i\le n} \sup_{\boldsymbol{x}\in\{-1,1\}^n} \left|m_i^{(n)}(\boldsymbol{x}) - \bar{x}_n^{p-1}\right| \le 3\gamma_n \text{ for all but finitely many } n\right) = 1 .$$

The proof of Corollary 1 is given in Appendix C. We will henceforth assume the slightly stronger condition $\alpha_n = \Omega(n^{1-p} \log n)$, which in particular implies that $\gamma_n \ll 1$. Note that for $p = 2$, this condition is satisfied if the Erdős-Rényi graph is almost-surely connected. This technical condition is required for establishing the consistency of the MPL estimator (see Lemma 13). See Section 5.6 for a complete proof of Theorem 6.

3.1. **Sketch of Proofs of the Main Results.** We now provide a brief sketch of some of the main ideas involved in the proofs of the main results stated in Sections 2 and 3. The main result in Section 2 is Theorem 1, and its proof starts by establishing the probability limit of $\hat{\beta} - \beta_0$ under the alternative, where $\hat{\beta}$ denotes either the ML or the MPL estimator (see Lemma 2). This verifies Condition (1). In order to verify Condition (2), we use a large deviation principle of $\overline{X}_n$ to pinpoint the asymptotic behavior of the p-values (see Lemma 3). The proofs of the other results in Section 2 follow from Theorem 1 and an analysis of the behavior of the function $H_{\beta_0,p}$ and its maximum value over certain regions.

The proof of Theorem 6 relies on comparing the Erdős-Rényi Ising model with the Curie-Weiss model, and showing that the two models are close in the exponential scale (see Lemma 7). This follows from an exponential concentration of the Hamiltonian of the Erdős-Rényi Ising model around its expected value, the latter being the Hamiltonian of the Curie-Weiss model. This keeps the asymptotics of the parameter estimates and the large deviation behavior of the sample mean unchanged in the Erdős-Rényi Ising model.

4. Numerical Results

In this section, we provide a graphical presentation of the numerical values of the asymptotic optimal sample sizes for the tests based on the ML and MPL estimators in the model (1.1), using the theoretical formula given in Theorem 1. Note that the Curie-Weiss model can be rewritten as

$$\mathbb{P}_{\beta,h,p}(\overline{X}_N = m) = \frac{1}{2^n Z_n(\beta,p)} \binom{n}{n(1+m)/2} e^{n\beta m^p}$$

for $m \in \left\{ -1, -1 + \frac{2}{n}, \cdots, 1 - \frac{2}{n}, 1 \right\}$, whence the partition function can be computed as

$$Z_n(\beta,p) = \sum_{m \in \mathcal{M}} \frac{1}{2^n} \binom{n}{n(1+m)/2} e^{n\beta m^p}.$$

Consequently, the mass function can be computed exactly for all moderately large n (say n up to 1000) and the Curie-Weiss model can be sampled directly from this mass function without the use of any MCMC approach.

In Figures 3–7, we fix the level $\delta = 0.05$. In Figure 3, we fix $\beta_0 = 0.7$, a value slightly larger than $\log 2$, and plot the asymptotic optimal sample size (for both the MPL and the ML tests, which must be same in this regime) for $p = 2, 3$ and 4, across $\beta > \beta_0$. We see that the asymptotic optimal sample size decreases as the alternative parameter β increases, which is expected, since the detection capability of the tests should increase as the alternative β moves far apart from the null β_0. Another important observation is that with increase in the interaction complexity p, the asymptotic optimal sample size requirement also increases at every alternative. One possible explanation of this phenomenon is that with increase in p, the threshold $\beta^*(p)$ also increases, and hence, the null β_0 (which is fixed at 0.7) gets closer to $\beta^*(p)$, which causes a slight increase in the difficulty of the testing problem.
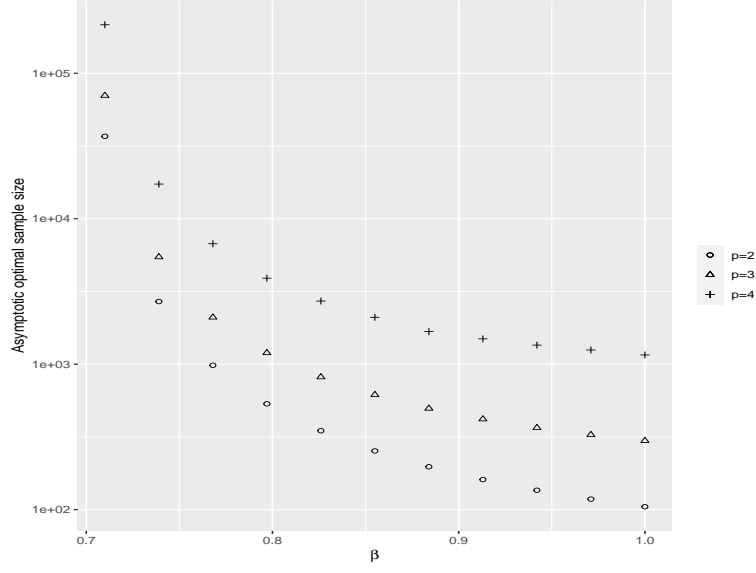
FIGURE 3. Asymptotic optimal sample size for the tests based on MLE and MPLE with varying β; $\beta_0 = 0.7 > \log 2, \delta = 0.05, p \in \{2, 3, 4\}$ (with logarithmic vertical scale).
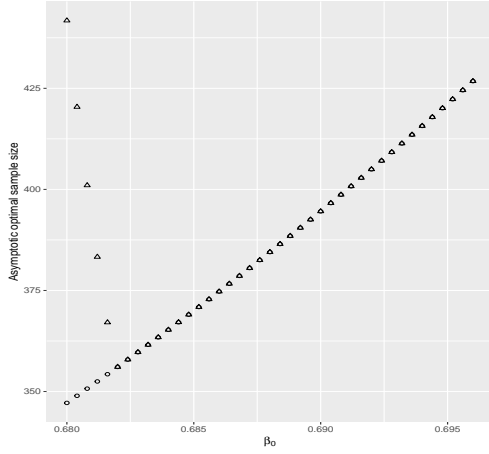

FIGURE 4. Asymptotic optimal sample size for the tests based on MLE and MPLE with varying β_0; $p = 3, \beta = 0.90, \delta = 0.05$.
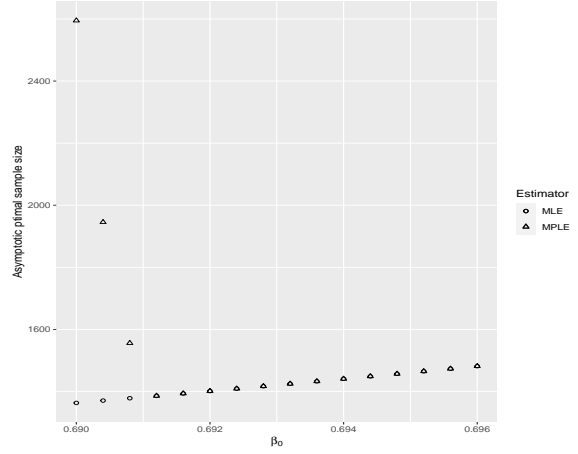
FIGURE 5. Asymptotic optimal sample size for the tests based on MLE and MPLE with varying β_0; $p = 4, \beta = 0.90, \delta = 0.05$.

In Figures 4 and 5, we fix the alternative $\beta = 0.9$, and demonstrate graphically (for the cases $p = 3$ and 4 respectively), that the asymptotic optimal sample sizes for the MPL and the ML tests differ for all β_0 in a small right neighborhood of $\beta^*(p)$ below $\log 2$, and agree above that neighborhood. The figures also demonstrate that the asymptotic optimal
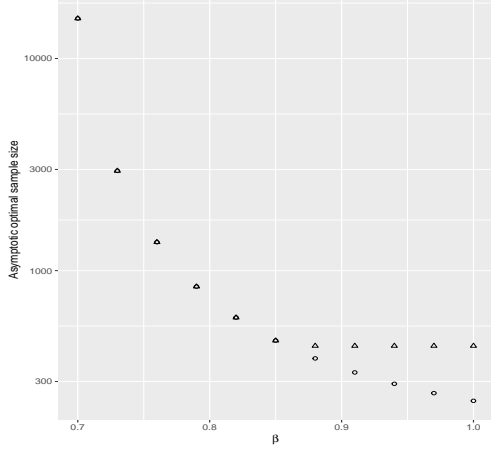
Figure 6. Asymptotic optimal sample size for the tests based on MLE and MPLE with varying β; $p = 3, \beta_0 = 0.68 < \log 2, \delta = 0.05$ (with logarithmic vertical scale).
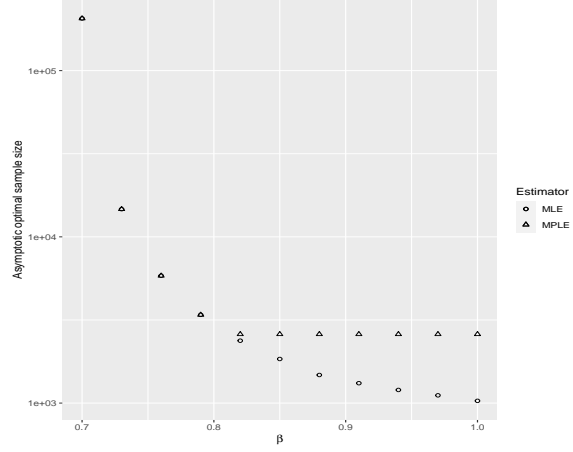
Figure 7. Asymptotic optimal sample size for the tests based on MLE and MPLE with varying β; $p = 4, \beta_0 = 0.69 < \log 2, \delta = 0.05$ (with logarithmic vertical scale).

sample size for the MPL test approaches ∞ as the null β_0 approaches the threshold $\beta^*(p)$. In Figures 6 and 7, we fix the null β_0 to values slightly smaller than $\log 2$, and demonstrate (for the cases $p = 3$ and 4 respectively) that although the asymptotic optimal sample sizes for the MPL and the ML tests coincide for all small values of the alternative $\beta > \beta_0$, they disagree for all β large enough. All these results reflect the contents of Theorems 3 and 4.

Now, we illustrate our theoretical results with numericals obtained from simulated data. In Figure 8, we plot the average p-value of the MPL test obtained from $10,000$ tuples generated from the 2-spin Curie-Weiss model at $\beta = 0.9$ against the null $\beta_0 = 0.7 > \beta^*(2)$, for each sample size ranging from around 175 to 375. We see that from around sample size 266, the average p-value goes down (and remains) below $\delta := 0.01$. This matches very closely with the theoretical sample-size value of 270 that one will obtain in this setting, from Theorem 1. Figure 9 illustrates the average p-value of the MPL test obtained from $10,000$ tuples generated from the 3-tensor Curie-Weiss model at $\beta = 0.9$ against the null $\beta_0 = 0.68 \in (\beta^*(3), \log 2)$, for each sample size ranging from around 575 to 775. We see that from around sample size 625, the average p-value goes down (and remains) below $\delta := 0.01$. Once again, this matches somewhat closely with the theoretical asymptotic sample-size value of 679 for the MPL test that one will obtain in this setting, from Theorem 1. In this case, the theoretical asymptotic sample size for the ML test turns out to be 533. This is smaller than the theoretical and empirically obtained sample sizes of 679 and 625 (respectively) for the MPL test, thereby demonstrating our theoretical finding that the
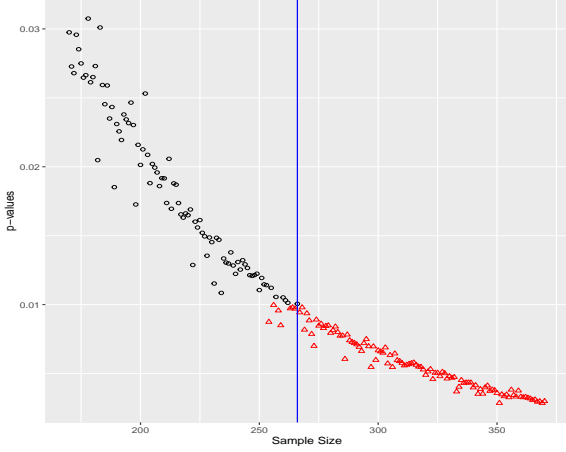
FIGURE 8. p-values for different sample sizes in the 2-spin Curie-Weiss model ($\beta_0 = 0.7 > \beta^*(2)$ and $\beta = 0.9$).

FIGURE 9. p-values for different sample sizes in the 3-tensor Curie-Weiss model ($\beta_0 = 0.68 \in (\beta^*(3), \log 2)$, $\beta = 0.9$).

MPL test becomes much less efficient than the ML test for $\beta_0 < \log 2$ and sufficiently high β.

Next, we consider the Erdős-Rényi Ising model in our numerical studies. The strategy is to generate a random matrix $\boldsymbol{A}$ with independent Bernoulli(α_n) entries, and simulate 1000 samples under both the null and alternative Erdős-Rényi Ising distributions using Glauber dynamics. We vary n in steps of size 5 in a region around the asymptotic optimal sample size given by Theorem 6. Given $\boldsymbol{A}$, we generate each sample using 10^6 iterations of the Glauber dynamics repeating the simulation independently 1000 times each under the null distribution β_0 and the alternative distribution $\beta \neq \beta_0$. In Figure 10, we plot the average p-values of the MPL tests obtained from 1000 tuples generated from the 2-spin Erdős-Rényi model at $\beta_0 = 0.7 > \beta^*(2)$ against the alternative $\beta = 0.9$, for sample size ranging from around 150 to 350. The simulation results match with the theoretical results indicating that the chains in the Glauber dynamics have mixed properly.

5. Proofs of the Main Results

In this section, we prove the main results stated in Sections 2 and 3. We start with the proof of Theorem 1.

5.1. **Proof of Theorem 1.** A main ingredient in the proof of Theorem 1 is the verification of conditions (1) and (2) in [6] . We begin with the verification of condition (1) with almost sure convergence replaced by convergence in probability.

Lemma 2. Under every $\beta > \beta^*(p)$, we have:

$$n^{-1/2}T_n \xrightarrow{P} \beta - \beta_0 \quad \text{as } n \to \infty \ ,$$
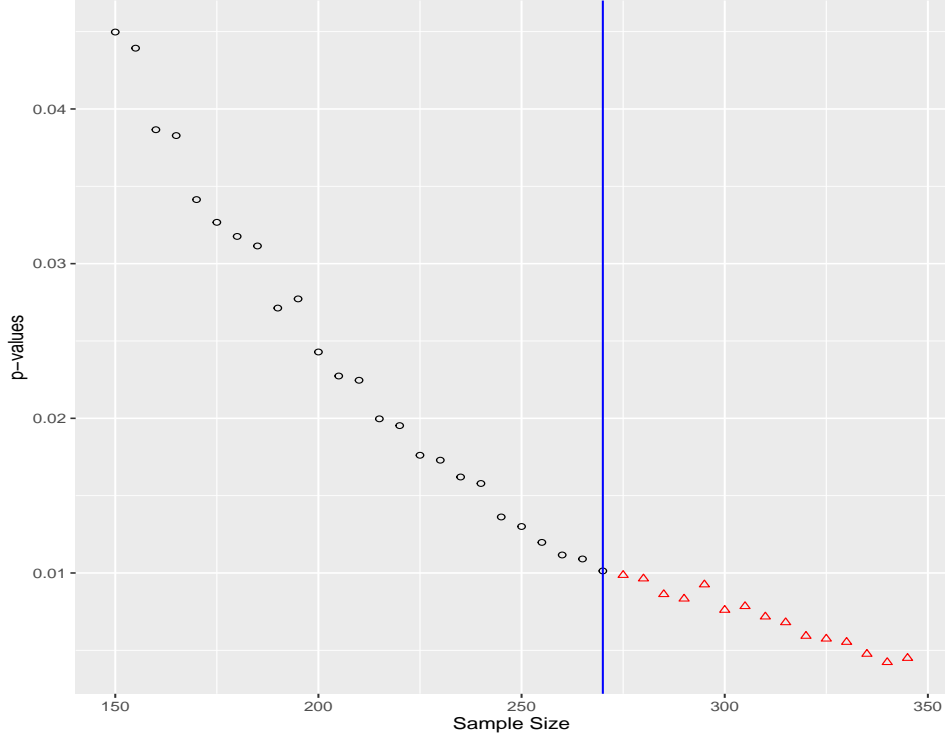
FIGURE 10. p-values for different sample sizes in the 2-spin Erdős-Rényi model $(\beta_0 = 0.7 > \beta^*(2)$ and $\beta = 0.9)$.

where T_n is either $\sqrt{n}(\hat{\beta}_{\mathrm{MPL}} - \beta_0)$ or $\sqrt{n}(\hat{\beta}_{\mathrm{ML}} - \beta_0)$.

We will now verify condition 2. For this, we will need the following lemma on the large deviation of $\overline{X}_n$, which follows from [45].

Lemma 3. For every subset $A \subseteq [-1, 1]$ such that A^o is dense in $\overline{A}$, we have:

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}_{\beta,p}(\overline{X}_n \in A) = \sup_{x \in A} H_{\beta,p}(x) - \sup_{x \in [-1,1]} H_{\beta,p}(x) \ .$$

where $H_{\beta,p}$ is as defined in (2.1).

Lemmas 2 and 3 are proved in Appendix A. We are now ready to prove Theorem 1.

Proof of Theorem 1. We deal with the test based on the MPL estimator first. To begin with, note that $\hat{\beta}_{\mathrm{MPL}} = \eta_p(\overline{X}_n)$. Fix $t > 0$, whence we have by Lemma 3:

$$
\begin{aligned}
n^{-1} \log[1 - F_{T_n,\beta_0}(n^{1/2}t)] &= n^{-1} \log \mathbb{P}_{\beta_0,p}\left(n^{-1/2}T_n > t\right) \\
&= n^{-1} \log \mathbb{P}_{\beta_0,p}\left(\hat{\beta}_{\mathrm{MPL}} - \beta_0 > t\right) \\
&= n^{-1} \log \mathbb{P}_{\beta_0,p}\left(\eta_p(\overline{X}_n) > \beta_0 + t\right) \\
&= n^{-1} \log \mathbb{P}_{\beta_0,p}\left(\overline{X}_n \in \eta_p^{-1}((\beta_0 + t, \infty))\right)
\end{aligned}
$$

$$= \sup_{x \in \eta_p^{-1}((\beta_0+t,\infty))} H_{\beta_0,p}(x) - \sup_{x \in [-1,1]} H_{\beta_0,p}(x) + o(1).$$

where the last step follows from Lemma 3, since it follows from the proof of Lemma 4, that the set $\eta_p^{-1}((\beta_0 + t, \infty))$ is a union of finitely many disjoint, non-degenerate intervals, and hence, its interior is dense in its closure.

In view of the above discussion, we conclude that the function f in condition (2) is given by:

$$f(t) = \sup_{x \in [-1,1]} H_{\beta_0,p}(x) - \sup_{x \in \eta_p^{-1}((\beta_0+t,\infty))} H_{\beta_0,p}(x) .$$

Since $x \mapsto H_{\beta_0,p}(x)$ and $\beta \mapsto m_*(\beta, p)$ are continuous functions (by Lemma 11) on $[-1, 1]$ and $(\beta^*(p), \infty)$ respectively, we conclude (in view of Lemma 4) that f is continuous on an open neighborhood of $\beta - \beta_0$. Also, in view of Lemma 4, the argument given below for the ML estimator, and the fact that $\beta_0 > \beta^*(p)$, it will follow that $f > 0$ on a non-empty open neighborhood of $\beta - \beta_0$.

The Bahadur slope of $\hat{\beta}_{\mathrm{MPL}}$ at an alternative β is then given by $2f(\beta - \beta_0)$. This completes the proof of Theorem 1 for the test based on the MPL estimator.

For the test based on the ML estimator, note that for every $t > 0$, we have:

$$
\begin{aligned}
n^{-1} \log[1 - F_{T_n,\beta_0}(n^{1/2}t)] &= n^{-1} \log \mathbb{P}_{\beta_0,p}\left(n^{-1/2}T_n > t\right) \\
&= n^{-1} \log \mathbb{P}_{\beta_0,p}\left(\hat{\beta}_{\mathrm{ML}} - \beta_0 > t\right) \\
&= n^{-1} \log \mathbb{P}_{\beta_0,p}\left(\overline{X}_n^p > \mathbb{E}_{\beta_0+t,p}(\overline{X}_n^p)\right) .
\end{aligned}
$$

The last step follows from the following facts:

(1) The function $F_n(\beta, p) := \log Z_n(\beta, p)$ is strictly convex in β (Lemma C.5 in [38]) and hence, $\frac{\partial F_N(\beta,p)}{\partial \beta}$ is strictly increasing in β;

(2) The ML equation is given by $\left.\frac{\partial F_N(\beta,p)}{\partial \beta}\right|_{\beta=\hat{\beta}_{\mathrm{ML}}} = N\overline{X}_n^p$;

(3) $\frac{\partial F_N(\beta,p)}{\partial \beta} = \mathbb{E}_{\beta,p}(N\overline{X}_n^p)$.

Now, it follows from [38] and the dominated convergence theorem, that

$$\mathbb{E}_{\beta_0+t,p}(\overline{X}_n^p) \to m_*(\beta_0 + t, p)^p .$$

Fix $\varepsilon \in (0, m_*(\beta_0 + t, p))$, to begin with. Then, there exists $N \geq 1$, such that

$$(m_*(\beta_0 + t, p) - \varepsilon)^p \leq \mathbb{E}_{\beta_0+t,p}(\overline{X}_n^p) \leq (m_*(\beta_0 + t, p) + \varepsilon)^p$$

for all $n \geq N$. Let us first consider the case p is odd. We then have the following by Lemma 3:

$$
\begin{aligned}
\limsup_{n\to\infty} n^{-1} \log[1 - F_{T_n,\beta_0}(n^{1/2}t)] &= \limsup_{n\to\infty} n^{-1} \log \mathbb{P}_{\beta_0,p}\left(\overline{X}_n^p > \mathbb{E}_{\beta_0+t,p}(\overline{X}_n^p)\right) \\
&\leq \limsup_{n\to\infty} n^{-1} \log \mathbb{P}_{\beta_0,p}\left(\overline{X}_n^p > (m_*(\beta_0 + t, p) - \varepsilon)^p\right)
\end{aligned}
$$

$$\begin{aligned}
&= \limsup_{n\to\infty} n^{-1} \log \mathbb{P}_{\beta_0,p}\left(\overline{X}_n > m_*(\beta_0 + t, p) - \varepsilon\right) \\
&= \sup_{x \, > \, m_*(\beta_0+t,p)-\varepsilon} H_{\beta_0,p}(x) - \sup_{x\in[-1,1]} H_{\beta_0,p}(x) \, .
\end{aligned}$$

Similarly, we also have:

$$\begin{aligned}
\liminf_{n\to\infty} n^{-1} \log[1 - F_{T_n,\beta_0}(n^{1/2}t)] &= \liminf_{n\to\infty} n^{-1} \log \mathbb{P}_{\beta_0,p}\left(\overline{X}_n^{\,p} > \mathbb{E}_{\beta_0+t,p}(\overline{X}_n^{\,p})\right) \\
&\geq \liminf_{n\to\infty} n^{-1} \log \mathbb{P}_{\beta_0,p}\left(\overline{X}_n^{\,p} > (m_*(\beta_0 + t, p) + \varepsilon)^p\right) \\
&= \liminf_{n\to\infty} n^{-1} \log \mathbb{P}_{\beta_0,p}\left(\overline{X}_n > m_*(\beta_0 + t, p) + \varepsilon\right) \\
&= \sup_{x \, > \, m_*(\beta_0+t,p)+\varepsilon} H_{\beta_0,p}(x) - \sup_{x\in[-1,1]} H_{\beta_0,p}(x) \, .
\end{aligned}$$

Since $\varepsilon > 0$ can be arbitrarily small, and $H_{\beta,p}$ is continuous, we must have for all odd p:

$$\lim_{n\to\infty} n^{-1} \log[1 - F_{T_n,\beta_0}(n^{1/2}t)] = \sup_{x \, > \, m_*(\beta_0+t,p)} H_{\beta_0,p}(x) - \sup_{x\in[-1,1]} H_{\beta_0,p}(x) \, . \qquad (5.1)$$

Next, suppose that p is even. In this case, $\boldsymbol{X}$ and $-\boldsymbol{X}$ have the same distribution, and hence, so do $\overline{X}_n$ and $-\overline{X}_n$. Hence, for every positive real number α, we have:

$$\begin{aligned}
n^{-1} \log \mathbb{P}_{\beta_0,p}\left(\overline{X}_n^{\,p} > \alpha^p\right) &= n^{-1} \log\left[2\, \mathbb{P}_{\beta_0,p}\left(\overline{X}_n > \alpha\right)\right] \\
&= n^{-1} \log \mathbb{P}_{\beta_0,p}(\overline{X}_n > \alpha) + o(1) \, .
\end{aligned}$$

Hence, the same argument as for the case of odd p also works here, showing that (5.1) holds when p is even, too.

In view of the above discussion, we conclude that the function f in condition (2) is given by:

$$f(t) = \sup_{x\in[-1,1]} H_{\beta_0,p}(x) - \sup_{x \, > \, m_*(\beta_0+t,p)} H_{\beta_0,p}(x) \, .$$

Since $x \mapsto H_{\beta_0,p}(x)$ and $\beta \mapsto m_*(\beta, p)$ are continuous functions (by Lemma 11) on $[-1, 1]$ and $(\beta^*(p), \infty)$ respectively, we conclude that f is continuous on an open neighborhood of $\beta - \beta_0$. Also, $f(\beta - \beta_0) > 0$ (and hence, $f(t) > 0$ on a non-empty open neighborhood of $\beta - \beta_0$), since $H_{\beta_0,p}$ is strictly decreasing on $m_*(\beta, p)$, and $m_*(\beta, p) > m_*(\beta_0, p)$ (by Lemma 11).

The Bahadur slope of $\hat{\beta}_{\mathrm{ML}}$ at an alternative β is thus given by $2f(\beta - \beta_0)$. This completes the proof of Theorem 1 for the test based on the ML estimator. The proof of Theorem 1 is now complete. $\qquad \square$

5.2. **Proof of Theorem 2.** The result for $p = 2$ follows directly from Theorem 1 and the following lemma, which is proved in Appendix A.

Lemma 4. For every $\beta > \beta_0 > \beta^*(p)$, we have:

$$\sup_{x\in\eta_p^{-1}((\beta,\infty))} H_{\beta_0,p}(x) = \begin{cases} \sup_{x \, > \, m_*(\beta,p)} H_{\beta_0,p}(x) & \text{if } p = 2, \\ \max\left\{\sup_{x \, > \, m_*(\beta,p)} H_{\beta_0,p}(x) \, , \, 0\right\} & \text{if } p \geq 3. \end{cases}$$

5.3. **Proof of Theorem 3.** The proof of Theorem 3 depends crucially on the following important lemma:

Lemma 5. For $p \geq 3$, $\beta > \beta_0 > \beta^*(p)$ and $\delta \in (0,1)$, a necessary and sufficient condition for $c_{\hat{\beta}_{\mathrm{MPL}}}(\beta_0, \beta, p) = c_{\hat{\beta}_{\mathrm{ML}}}(\beta_0, \beta, p) \iff N^*_{\hat{\beta}_{\mathrm{MPL}}}(\beta_0, \beta, \delta, p) = N^*_{\hat{\beta}_{\mathrm{ML}}}(\beta_0, \beta, \delta, p)$ is $H_{\beta_0,p}(m_*(\beta, p)) \geq 0$.

Proof of Lemma 5. For $p \geq 3$, it follows from Lemma 10 that $H'_{\beta,p} \leq 0$ on $(m_*(\beta, p), 1)$, and hence, $H'_{\beta_0,p} \leq 0$ on $(m_*(\beta, p), 1)$. Consequently,

$$\sup_{x \, > \, m_*(\beta,p)} H_{\beta_0,p}(x) = H_{\beta_0,p}(m_*(\beta, p)) \, . \tag{5.2}$$

Lemma 5 now follows from Lemma 4. $\square$

Returning to the proof of Theorem 3, note that it follows from (5.2) that

$$H_{\beta_0,p}(m_*(\beta, p)) \geq H_{\beta_0,p}(1) = \beta_0 - \log 2 \, .$$

This shows that the condition $\beta_0 \geq \log 2$ is sufficient to ensure equality of the Bahadur slopes and the asymptotic optimal sample sizes. On the other hand, if $\beta_0 < \log 2$, then $\lim_{x\to 1} H_{\beta_0,p}(x) < 0$. Since $\lim_{\beta\to\infty} m_*(\beta, p) = 1$ (by Lemma 11), we must have for all $\beta > \beta_0$ large enough, $H_{\beta_0,p}(m_*(\beta, p)) < 0$, which shows, in view of Lemma 5, that the Bahadur slopes and asymptotic optimal sample sizes for the tests based on the MPL and ML estimators do not agree in this case.

5.4. **Proof of Theorem 4.** By Lemma 5, $\beta_0 > \beta^*(p)$ satisfies $\mathrm{eff}(\hat{\beta}_{\mathrm{ML}}, \hat{\beta}_{\mathrm{MPL}}, \beta_0, \beta) > 1$ if and only if $H_{\beta_0,p}(m_*(\beta, p)) < 0$, i.e. $\beta_0 m_*(\beta, p)^p < I(m_*(\beta, p))$. This establishes the first part of Theorem 4. Next, note that

$$\lim_{\beta_0\to\beta^*(p)^+} \sup_{x\in[-1,1]} H_{\beta_0,p}(x) = 0 \, . \tag{5.3}$$

On the other hand, we also have:

$$\lim_{\beta_0\to\beta^*(p)^+} \sup_{x \, > \, m_*(\beta,p)} H_{\beta_0,p}(x) = H_{\beta^*(p),p}(m_*(\beta, p)) < 0 \, .$$

Hence, it follows from Lemma 4 that:

$$\lim_{\beta_0\to\beta^*(p)^+} \sup_{x\in\eta_p^{-1}((\beta,\infty))} H_{\beta_0,p}(x) = \lim_{\beta_0\to\beta^*(p)^+} \max\left\{0 \, , \, \sup_{x \, > \, m_*(\beta,p)} H_{\beta_0,p}(x)\right\} = 0 \, . \tag{5.4}$$

It thus follows from (5.3), (5.4) and Theorem 1, that

$$\lim_{\beta_0\to\beta^*(p)^+} N^*_{\hat{\beta}_{\mathrm{MPL}}}(\beta_0, \beta, \delta, p) = \infty \quad \text{and} \quad \lim_{\beta_0\to\beta^*(p)^+} N^*_{\hat{\beta}_{\mathrm{ML}}}(\beta_0, \beta, \delta, p) = \frac{\log(\delta)}{H_{\beta^*(p),p}(m_*(\beta, p))} < \infty. \tag{5.5}$$

Finally, the limiting Bahadur efficiency part follows directly from (5.5). This completes the proof of Theorem 4.

5.5. **Proof of Theorem 5.** Since $m_*(\beta, p) \to 1$ as $\beta \to \infty$ (by Lemma 11), we have:

$$\lim_{\beta \to \infty} H_{\beta_0, p}(m_*(\beta, p)) = \beta_0 - \log 2 < 0 .$$

Hence, $H_{\beta_0, p}(m_*(\beta, p)) < 0$ for all β large enough, which shows, in view of Lemma 4, that:

$$\sup_{x \in \eta_p^{-1}((\beta, \infty))} H_{\beta_0, p}(x) = 0 .$$

Theorem 5 now follows from Theorem 1.

5.6. **Proof of Theorem 6.** As in the Curie-Weiss model, a main ingredient in the proof of Theorem 6 is the verification of conditions (1) and (2) in [6]. Once again, we begin with the verification of condition (1) with almost sure convergence replaced by convergence in probability. The following result is an analogous version of Lemma 2 for the model (3.1).

Lemma 6. Under the model (3.1), for every $\beta > \beta^*(p)$, we have:

$$n^{-1/2} T_n \xrightarrow{P} \beta - \beta_0 \quad \text{as } n \to \infty ,$$

where T_n is either $\sqrt{n}(\hat{\beta}_{\mathrm{MPL}} - \beta_0)$ or $\sqrt{n}(\hat{\beta}_{\mathrm{ML}} - \beta_0)$.

We will now verify condition 2. Towards this, we will use Lemma 1 to compare the probability models (1.1) and (3.1). In fact, we prove a slightly more general result below, which solves our purpose, but may be of independent interest for more general objectives.

Lemma 7. Let $\{\beta_n\}_{n \geq 1}$ be a bounded sequence of positive real numbers. Then, with probability 1, we have:

$$\sup_{A, B \subseteq \{-1, 1\}^n} \left| \log \mathbb{P}^*_{\beta_n, p}(A | B) - \log \mathbb{P}_{\beta_n, p}(A | B) \right| = O(n \gamma_n) .$$

Remark 5.1. One can compare the logarithms of the two (unconditional) measures $\mathbb{P}^*$ (3.1) and $\mathbb{P}$ (1.1) by taking $B := \{-1, 1\}^n$ in Lemma 7.

One can now use Lemmas 1 and 7 to compare the log-normalizing constants and asymptotics of the sample mean in the two models (1.1) and (3.1).

Lemma 8. We have the following with probability 1.
   1. If $Z_n^*(\beta, p)$ denotes the normalizing constant of the model (3.1), then

$$|\log Z_n(\beta, p) - \log Z_n^*(\beta, p)| = O(n \gamma_n) .$$
   2. If $\boldsymbol{X}$ is generated from the model (3.1), then for every $\beta > \beta^*(p)$ and fixed $\varepsilon > 0$, we have:

$$\mathbb{P}^*_{\beta, p} \left(|\overline{X}_n^p - m_*(\beta, p)^p| \geq \varepsilon \right) \leq e^{-n \Omega(1)} . \tag{5.6}$$

   In particular, $\overline{X}_n \xrightarrow{P} m_*(\beta, p)$ under the model (3.1).

Lemmas 7 and 8 are proved in Appendix A. In view of all that we have above, we are now ready to prove Theorem 6.

Proof of Theorem 6. We begin with the ML estimator first. Note that by Lemma 1, we have:

$$
\begin{aligned}
\mathbb{P}^*_{\beta_0,p}(\overline{X}_n^p > \mathbb{E}^*_{\beta_0+t,p}\overline{X}_n^p + 6\gamma_n) &\leq \mathbb{P}^*_{\beta_0,p}(H_n(\boldsymbol{X}) > \mathbb{E}^*_{\beta_0+t,p}H_n(\boldsymbol{X})) \\
&\leq \mathbb{P}^*_{\beta_0,p}(\overline{X}_n^p > \mathbb{E}^*_{\beta_0+t,p}\overline{X}_n^p - 6\gamma_n).
\end{aligned}
$$

Now, note that by part (2) of Lemma 8 and the dominated convergence theorem, $\mathbb{E}^*_{\beta_0+t,p}\overline{X}_n^p \to m_*(\beta_0+t,p)^p$. Hence, we have:

$$
\begin{aligned}
\mathbb{P}^*_{\beta_0,p}(\overline{X}_n^p > m_*(\beta_0+t,p)^p + o(1)) &\leq \mathbb{P}^*_{\beta_0,p}(H_n(\boldsymbol{X}) > \mathbb{E}^*_{\beta_0+t,p}H_n(\boldsymbol{X})) \\
&\leq \mathbb{P}^*_{\beta_0,p}(\overline{X}_n^p > m_*(\beta_0+t,p)^p + \tilde{o}(1))
\end{aligned}
$$

where $o(1)$ and $\tilde{o}(1)$ denote two real sequences converging to 0. Hence, for every fixed $\varepsilon > 0$ sufficiently small, one has the following for all large n:

$$
\begin{aligned}
\mathbb{P}^*_{\beta_0,p}(\overline{X}_n^p > (m_*(\beta_0+t,p)+\varepsilon)^p) &\leq \mathbb{P}^*_{\beta_0,p}(H_n(\boldsymbol{X}) > \mathbb{E}^*_{\beta_0+t,p}H_n(\boldsymbol{X})) \\
&\leq \mathbb{P}^*_{\beta_0,p}(\overline{X}_n^p > (m_*(\beta_0+t,p)-\varepsilon)^p).
\end{aligned}
$$

It now follows from Lemma 7, that:

$$
\begin{aligned}
\frac{1}{n}\log\mathbb{P}_{\beta_0,p}(\overline{X}_n^p > (m_*(\beta_0+t,p)+\varepsilon)^p) + o(1) &\leq \frac{1}{n}\log\mathbb{P}^*_{\beta_0,p}(H_n(\boldsymbol{X}) > \mathbb{E}^*_{\beta_0+t,p}H_n(\boldsymbol{X})) \\
&\leq \frac{1}{n}\log\mathbb{P}_{\beta_0,p}(\overline{X}_n^p > (m_*(\beta_0+t,p)-\varepsilon)^p) + o(1).
\end{aligned}
$$

Theorem 6 for the ML estimator now follows from the proof of Theorem 1 and Lemma 6.

Next, we consider the MPL estimator. It follows from (2.3) in [39], that $\hat{\beta}_{\mathrm{MPL}}$ is the least solution of the equation (in β):

$$
H_n(\boldsymbol{X}) = \sum_{i=1}^n m_i(\boldsymbol{X})\tanh(p\beta m_i(\boldsymbol{X})) \ .
$$

Define $\psi_n(\beta) := n^{-1}\sum_{i=1}^n m_i(\boldsymbol{X})\tanh(p\beta m_i(\boldsymbol{X}))$. Since with probability 1, we have the following for all large n

$$
\psi'_n(\beta) = \frac{p}{n}\sum_{i=1}^n m_i^2(\boldsymbol{X})\mathrm{sech}^2(p\beta m_i(\boldsymbol{X})) > 0,
$$

the function ψ_n is strictly increasing for all large n, with probability 1. Hence, we have (with probability 1 for all large n) the following for all $t > 0$:

$$
\begin{aligned}
\frac{1}{n}\log\mathbb{P}^*_{\beta_0,p}(\hat{\beta}_{\mathrm{MPL}} > \beta_0+t) &= \frac{1}{n}\log\mathbb{P}^*_{\beta_0,p}(\psi_n(\hat{\beta}_{\mathrm{MPL}}) > \psi_n(\beta_0+t)) \\
&= \frac{1}{n}\log\mathbb{P}^*_{\beta_0,p}\left(\frac{1}{n}H_n(\boldsymbol{X}) > \frac{1}{n}\sum_{i=1}^n m_i(\boldsymbol{X})\tanh(p(\beta_0+t)m_i(\boldsymbol{X}))\right)
\end{aligned}
$$

$$= \frac{1}{n} \log \mathbb{P}^*_{\beta_0,p} \left(\overline{X}_n^{\,p} > \frac{1}{n} \sum_{i=1}^n m_i(\boldsymbol{X}) \tanh(p(\beta_0 + t) m_i(\boldsymbol{X})) + o(1) \right)$$

Therefore, in view of Lemma 1, Corollary 1 and Lemma 7, we have the following for every fixed $\varepsilon > 0$ sufficiently small:

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}^*_{\beta_0,p}(\hat{\beta}_{\mathrm{MPL}} > \beta_0 + t)$$

$$\leq \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}^*_{\beta_0,p} \left(\overline{X}_n^{\,p} > \frac{1}{n} \sum_{i=1}^n \overline{X}_n^{\,p-1} \tanh(p(\beta_0 + t)\overline{X}_n^{\,p-1}) + o(1) \right)$$

$$= \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}^*_{\beta_0,p} \left(\overline{X}_n^{\,p} > \overline{X}_n^{\,p-1} \tanh(p(\beta_0 + t)\overline{X}_n^{\,p-1}) + o(1) \right)$$

$$\leq \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}^*_{\beta_0,p} \left(\eta_p(\overline{X}_n) > \beta_0 + t - \varepsilon \right)$$

$$= \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}_{\beta_0,p} \left(\eta_p(\overline{X}_n) > \beta_0 + t - \varepsilon \right)$$

$$= \sup_{x \in \eta_p^{-1}((\beta_0+t-\varepsilon,\infty))} H_{\beta_0,p}(x) - \sup_{x \in [-1,1]} H_{\beta_0,p}(x).$$

We can now take $\varepsilon \downarrow 0$ to conclude that:

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}^*_{\beta_0,p}(\hat{\beta}_{\mathrm{MPL}} > \beta_0 + t) \leq \sup_{x \in \eta_p^{-1}((\beta_0+t,\infty))} H_{\beta_0,p}(x) - \sup_{x \in [-1,1]} H_{\beta_0,p}(x) . \quad (5.7)$$

By an exactly similar approach, we can show that:

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}^*_{\beta_0,p}(\hat{\beta}_{\mathrm{MPL}} > \beta_0 + t) \geq \sup_{x \in \eta_p^{-1}((\beta_0+t,\infty))} H_{\beta_0,p}(x) - \sup_{x \in [-1,1]} H_{\beta_0,p}(x) . \quad (5.8)$$

Theorem 6 now follows from (5.7) and (5.8). $\square$

6. Discussion

In this paper, we derived the Bahadur slopes and optimal sample sizes required for significance of the tests based on the maximum likelihood (ML) and the maximum pseudolikelihood (MPL) estimators for the tensor Curie-Weiss model. One of our interesting findings is that although the MPL estimator is just an approximation of the ML estimator, the former is as Bahadur efficient as the latter everywhere in the parameter space for $p = 2$, and throughout most of the parameter space for $p \geq 3$. More precisely, the MPL estimator is equally Bahadur efficient as the ML estimator for $p = 2$, and for $p \geq 3$ this is true for all values of the alternative, if and only if the null parameter is greater than or equal to $\log 2$. For $p \geq 3$, if the null parameter lies strictly between the estimation threshold and $\log 2$, then the MPL estimator is strictly less efficient than the ML estimator for all sufficiently large values of the alternative parameter. However, even in this regime, the Bahadur ARE of the MPL estimator with respect to the ML estimator is lower bounded by a positive fraction. We also showed that our results hold verbatim for the more general class of tensor

Erdős-Rényi Ising models, where we can even allow for some sparsity in the underlying hypergraph.

We conjecture that similar results are also true for Ising models on dense deterministic and stochastic block model hypergraphs (with the Hamiltonian being suitably scaled) too, and believe that this can be shown by slight (routine) modifications of the methods used in Section 3. A potentially interesting direction for future research in this area, would be to consider Ising models on more general hypergraphs, for example arbitrary regular hypergraphs. Probability limits and fluctuations of the average magnetization for 2-spin Ising models on d-regular graphs have been recently derived in [20], where the authors show that the fluctuations are universal and same as that of the 2-spin Curie-Weiss model in the entire ferromagnetic parameter regime as long as $d \gg \sqrt{n}$. The next natural step would thus, be to derive a large deviation principle for the Hamiltonians of such models, not just in the 2-spin case, but also for the tensor case. These will in turn enable one to derive the Bahadur slopes and the optimal sample sizes for the ML and the MPL estimators in Ising models on regular hypergraphs.

Since the behavior of the MPL estimator in terms of its Bahadur ARE with respect to the ML estimator is dependent on whether $p = 2$ or $p > 2$, one may be interested in asking if it is possible to test the hypothesis $p = 2$ versus $p > 2$, at least in the Curie-Weiss model, based on a single observation $\boldsymbol{X}$ from the model. This is an interesting problem in its own right, and one plausible approach may be to estimate β assuming $p = 2$ (by either the ML or the MPL approach), simulate a large number B of observations $\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \ldots, \boldsymbol{X}^{(B)}$ from the 2-spin Curie-Weiss model with parameter $\hat{\beta}$, and determine whether $\overline{X}$ lies within the 2.5% and 97.5% quantiles of the histogram formed by $\overline{X^{(1)}}, \overline{X^{(2)}}, \ldots, \overline{X^{(B)}}$. If not, then that should be a reasonable evidence against the null hypothesis $p = 2$. This approach has some flaws though, one being that we are never sure if the actual observation $\boldsymbol{X}$ is coming from a Curie-Weiss model at low temperature (β above the threshold), because otherwise, $\hat{\beta}$ will be inconsistent. The second drawback is that one can never say surely that the observation is coming from a 2-spin Curie-Weiss model indeed, even if $\overline{X}$ does lie within the 2.5% and 97.5% quantiles of the empirical histogram. In other words, the power guarantee of this testing approach is dubious. We leave this problem as an interesting direction for future research.

Finally, we would like to mention that although this paper is concerned with efficient estimation and testing of the coupling strength in tensor Ising models and the model considered here does not have any external magnetic field term, there is a significant literature on testing for external fields as well (see for example, [40, 41]), at least in the classical 2-spin setting. In the Curie-Weiss model with external field, the sufficient statistic is still the sample mean $\overline{X}_n$, and both the ML and the MPL estimates are once again functions of $\overline{X}_n$. Hence, we expect similar techniques involving asymptotics and large deviations of

$\overline{X}_n$ to apply in deriving the Bahadur slopes and optimal sample sizes corresponding to the estimates of the external fields, as well.

7. Acknowledgment

References

[1] T.W. Anderson and L.A. Goodman, Statistical Inference about Markov Chains *The Annals of Mathematical Statistics*, Vol. 28, No. 1, 89–109, 1957.

[2] R. R. Bahadur, Asymptotic Efficiency of Tests and Estimates *Sankhya*, Vol. 22, 229–252, 1960.

[3] R. R. Bahadur, Simultaneous Comparison of the Optimum and Sign Tests of a Normal Mean *Contributions to Probability and Statistics – Essays in Honor of Harold Hotelling*, Stanford Univ. Press. 79–88, 1960.

[4] R. R. Bahadur, Stochastic Comparison of Tests *The Annals of Mathematical Statistics*, Vol. 31, No. 2, 276–295, 1960.

[5] R. R. Bahadur, An Optimal Property of the Likelihood Ratio Statistic *Proc. Fifth Berkeley Symp. Math. Statist. Prob. 1*, Univ. of California Press, 1965.

[6] R. R. Bahadur, Rates of Convergence of Estimates and Test Statistics *The Annals of Mathematical Statistics*, Vol. 38, No. 2, 303–324, 1967.

[7] S. Banerjee, B. P. Carlin, and A. E. Gelfand, Hierarchical modeling and analysis for spatial data, *Chapman and Hall/CRC, 2014.*

[8] L. Beghin and Y.Y. Nikitin, Approximate asymptotic Bahadur efficiency of independence tests with random sample size *Journal of the Italian Statistical Society*, Vol. 8, No. 1, 1–23, 1999.

[9] N. Berrahou and L. Douge, Bahadur efficiency of nonparametric test for independence based on L_1-error, *arXiv*:1211.1725, 2012.

[10] J. Besag, Spatial interaction and the statistical analysis of lattice systems, *J. Roy. Stat. Soc. B*, Vol. 36, 192–236, 1974.

[11] J. Besag, Statistical analysis of non-lattice data, *The Statistician*, Vol. 24, No. 3, 179–195, 1975.

[12] A. Bovier and V. Gayrard, The Thermodynamics of the Curie-Weiss Model with Random Couplings *Journal of Statistical Physics*, Vol. 72, Nos. 3/4, 1993.

[13] B. B. Bhattacharya and S. Mukherjee, Inference in ising models, *Bernoulli*, Vol. 24, No. 1, 493–525, 2018.

[14] S. Bhattacharya, R. Mukherjee and G. Ray, Sharp Signal Detection under Ferromagnetic Ising Models, *arXiv*:2110.02949, 2021.

[15] G. Bresler, Efficiently learning Ising models on arbitrary graphs, *Proceedings Symposium on Theory of Computing (STOC)*, 771–782, 2015.

[16] Y. Cao, M. Neykov and H. Liu, High-Temperature Structure Detection in Ferromagnets, *Information and Inference: A Journal of the IMA, iaaa032*, https://doi.org/10.1093/imaiai/iaaa032, 2020.

[17] F. Comets, On Consistency of a Class of Estimators for Exponential Families of Markov Random Fields on the Lattice, *The Annals of Statistics*, Vol. 20, No. 1, 455–468, 1992.

[18] S. Chatterjee, Estimation in spin glasses: A first step, *The Annals of Statistics*, Vol. 35, No. 5, 1931–1946, 2007.

[19] N. Deb, B.B. Bhattacharya and B. Sen, Efficiency Lower Bounds for Distribution-Free Hotelling-Type Two-Sample Tests Based on Optimal Transport, *arXiv*:2104.01986, 2021.

[20] N. Deb and S. Mukherjee, Fluctuations in mean-field Ising models, *arXiv*:2005.00710, 2020.

[21] J. Fröhlich, Mathematical aspects of the physics of disordered systems, *Phénomènes critiques, systèmes aléatoires, théories de jauge, Part I, II (Les Houches, 1984)*, 725–893, 1986.

[22] S. Geman and C. Graffigne, Markov random field image models and their applications to computer vision, *Proceedings of the International Congress of Mathematicians*, 1496–1517, 1986.

[23] P. Ghosal and S. Mukherjee, Joint estimation of parameters in Ising model, *arXiv*:1801.06570, 2019.

[24] M. Goemans, Chernoff bounds, and some applications, *Lecture Notes: 18.310*, Feb 21, 2015. http://math.mit.edu/~goemans/18310S15/chernoff-notes.pdf

[25] P. J. Green and S. Richardson, Hidden markov models and disease mapping, *Journal of the American Statistical Association*, Vol. 97, No. 460, 1055–1070, 2002.

[26] P. Groeneboom and J. Oosterhoff, Bahadur Efficiency and Small-Sample Efficiency, *International Statistical Review / Revue Internationale De Statistique*, Vol. 49, No. 2, 127-141, 1981.

[27] X. Guyon and H.R. Künsh, Asymptotic Comparison of Estimators in the Ising Model, *Stochastic Models, Statistical Methods, and Algorithms in Image Analysis. Lecture Notes in Statistics*, Vol. 74 Springer, New York, NY. https://doi.org/10.1007/978-1-4612-2920-9_12

[28] L. Gyorfi, G. Morvai and I. Vajda, Information-theoretic methods in testing the goodness of fit. *2000 IEEE International Symposium on Information Theory (Cat. No.00CH37060)*, 28–, 2000.

[29] P. Harremoës and I. Vajda, On the Bahadur-Efficient Testing of Uniformity by means of the Entropy. *IEEE Transactions on Information Theory*, Vol. 9, No. 1, 321–331, 2008.

[30] P. Harremoes and I. Vajda, Efficiency of entropy testing. *2008 IEEE International Symposium on Information Theory*, 2639-2643, 2008.

[31] W. Hoeffding, Probability Inequalities for Sums of Bounded Random Variables *Journal of the American Statistical Association*, Vol. 58, No. 301, 13–30, 1963.

[32] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. USA*, Vol. 79, No. 8, 2554–2558, 1982.

[33] D. Huang and S. Meyn, Generalized Error Exponents for Small Sample Universal Hypothesis Testing. *IEEE Transactions on Information Theory*, Vol. 59, No. 12, 8157–8181, 2013.

[34] E. Ising, Beitrag zur theorie der ferromagnetismus, *Zeitschrift für Physik*, Vol. 31, 253–258, 1925.

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

[36] A. Keziou and P. Regnault, Semiparametric Estimation of Mutual Information and Related Criteria: Optimal Test of Independence. *IEEE Transactions on Information Theory*, Vol. 63, No. 1, 57–71, 2017.

[37] M. Löwe, R. Meiners and F. Torres, Large Deviations Principle for Curie-Weiss Models with Random Fields, *Journal of Physics A: Mathematical and Theoretical*, Vol. 46, No. 12, 2013.

[38] S. Mukherjee, J. Son and B. B. Bhattacharya, Phase transitions of the maximum likelihood estimates in the p-spin Curie-Weiss model, *arXiv*:2005.03631, 2020.

[39] S. Mukherjee, J. Son and B. B. Bhattacharya, Estimation in Tensor Ising Models, *arXiv*:2008.12882, 2020.

[40] R. Mukherjee, S. Mukherjee and M. Yuan, Global testing against sparse alternatives under Ising models, *The Annals of Statistics*, Vol. 46 (5), 2062–2093, 2018.

[41] R. Mukherjee and G. Ray, On testing for parameters in Ising models, *arXiv*:1906.00456, 2019.

[42] M. Neykov and H. Liu, Property testing in high-dimensional Ising models, *The Annals of Statistics*, 2019.

[43] M. P. Quine and J. Robinson, Efficiencies of chi-square and likelihood ratio goodness-of-fit tests, *The Annals of Statistics*, Vol. 13, No. 2, 727 – 742, 1985.

[44] A.L. Rukhin, Bahadur Efficiency of Tests of Separate Hypotheses and Adaptive Test Statistics, *Journal of the American Statistical Association* , Vol. 88, No. 421, 161–165, 1993.

[45] S. Sumedha and S.K. Singh, Effect of Random Field Disorder on the First Order Transition in p-spin Interaction model, *Article in Physica A: Statistical Mechanics and its Applications*, 2015.

Appendix A. Proofs of Technical Lemmas

In this section, we prove the technical lemmas mentioned in Sections 2 and 3.

A.1. **Proof of Lemma 2.** Note that $n^{-1/2}T_n = \hat{\beta} - \beta_0$, where $\hat{\beta}$ is either $\hat{\beta}_{\mathrm{ML}}$ or $\hat{\beta}_{\mathrm{MPL}}$. It follows from [38] and [39], that under every $\beta > \beta^*(p)$, $\hat{\beta} \xrightarrow{P} \beta$. This proves Lemma 2.

A.2. **Proof of Lemma 3.** It follows from display (18) in [45] that $\overline{X}_n$ satisfies a large deviation principle (LDP) with rate function:

$$I(x) := -\beta x^p + \frac{x}{2}\sinh^{-1}\left(\frac{2x}{1-x^2}\right) + \frac{1}{2}\log\left(1-x^2\right) - \inf_{y\in\mathbb{R}}\left\{\sup_{z\in\mathbb{R}}\{yz - \log\cosh(z)\} - \beta y^p\right\}.$$

Using the identity

$$\sinh^{-1}(x) = \log\left(x + \sqrt{x^2+1}\right),$$

we have:

$$-\beta x^p + \frac{x}{2}\sinh^{-1}\left(\frac{2x}{1-x^2}\right) + \frac{1}{2}\log\left(1-x^2\right)$$

$$= -\beta x^p + \frac{x}{2}\log\left(\frac{2x}{1-x^2} + \sqrt{\frac{4x^2}{(1-x^2)^2} + 1}\right) + \frac{1}{2}\log\left(1-x^2\right)$$

$$= -\beta x^p + \frac{x}{2}\log\left(\frac{(1+x)^2}{1-x^2}\right) + \frac{1}{2}\log\left(1-x^2\right)$$

$$= -\beta x^p + x\log(1+x) + \frac{1-x}{2}\log(1-x^2)$$

$$= -\beta x^p + x\log(1+x) + \frac{1-x}{2}\log(1+x) + \frac{1-x}{2}\log(1-x)$$

$$= -\beta x^p + \frac{1}{2}\left\{(1+x)\log(1+x) + (1-x)\log(1-x)\right\} = -H_{\beta,p}(x).$$

It now follows from Lemma 9 that

$$I(x) = -H_{\beta,p}(x) + \sup_{y\in[-1,1]} H_{\beta,p}(y). \tag{A.1}$$

Lemma 3 now follows from (A.1) and the fact that I is a continuous function on $[-1,1]$.

A.3. **Proof of Lemma 4.** First, suppose that p is even. Then, η_p is an even function, and hence, the set $\eta_p^{-1}((\beta,\infty))$ is symmetric around 0. This, together with the fact that $H_{\beta_0,p}$ is an even function, implies that

$$\sup_{x\in\eta_p^{-1}((\beta,\infty))} H_{\beta_0,p}(x) = \sup_{x\in\eta_p^{-1}((\beta,\infty))\bigcap(0,1]} H_{\beta_0,p}(x). \tag{A.2}$$

Now, note that if p is odd, then $\eta_p^{-1}((\beta, \infty)) \bigcap [-1, 0] = \varnothing$, since $\eta_p(x) \leq 0$ for all $x \in [-1, 0]$. Hence, (A.2) is valid for odd p, too.

Now, $x \in \eta_p^{-1}((\beta, \infty)) \bigcap (0, 1]$ if and only if $x \in (0, 1]$ satisfies $p^{-1} x^{1-p} \tanh^{-1}(x) > \beta$, if and only if $x \in (0, 1]$ satisfies

$$H'_{\beta, p}(x) = \beta p x^{p-1} - \tanh^{-1}(x) < 0 .$$

If $p \geq 3$, then by Lemma 10, this region is precisely equal to $\{(0, \underline{m}(\beta, p)) \bigcup (m_*(\beta, p), 1]\} \setminus F$ for some finite set F (which is either singleton or empty). Hence, for $p \geq 3$, we have by continuity of $H_{\beta_0, p}$, that:

$$\sup_{x \in \eta_p^{-1}((\beta, \infty))} H_{\beta_0, p}(x) = \max \left\{ \sup_{x \in (0, \underline{m}(\beta, p))} H_{\beta_0, p}(x) , \sup_{x \in (m_*(\beta, p), 1]} H_{\beta_0, p}(x) \right\} . \tag{A.3}$$

Since $H_{\beta, p}(0) = 0$ and $H_{\beta, p}$ is decreasing on $(0, \underline{m}(\beta, p))$, we must have:

$$\sup_{x \in (0, \underline{m}(\beta, p))} H_{\beta, p}(x) = 0 .$$

Further, since $H_{\beta_0, p}(0) = 0$ and $H_{\beta_0, p} \leq H_{\beta, p}$ on $[0, 1]$, we must have:

$$\sup_{x \in (0, \underline{m}(\beta, p))} H_{\beta_0, p}(x) = 0 .$$

Lemma 4 for $p \geq 3$ now follows from (A.3). Now, let $p = 2$. Then,

$$\eta_p^{-1}((\beta, \infty)) \bigcap (0, 1] = (m_*(\beta, p), 1] .$$

Hence,

$$\sup_{x \in \eta_p^{-1}((\beta, \infty))} H_{\beta_0, p}(x) = \sup_{x > m_*(\beta, p)} H_{\beta_0, p}(x) .$$

This completes the proof of Lemma 4.

A.4. **Proof of Lemma 6.** Note that $n^{-1/2} T_n = \hat{\beta} - \beta_0$, where $\hat{\beta}$ is either $\hat{\beta}_{\mathrm{ML}}$ or $\hat{\beta}_{\mathrm{MPL}}$. All the following arguments are on the following event, which has probability 1 in view of Lemma 1:

$$\mathcal{E} := \left\{ \sup_{\boldsymbol{x} \in \{-1, 1\}^n} |H_n(\boldsymbol{x}) - n\overline{x}_n^p| \leq 3n\gamma_n \text{ for all but finitely many } n \right\} .$$

Let us first consider the case $\hat{\beta} = \hat{\beta}_{\mathrm{ML}}$. Then, for every fixed $t > 0$, we have:

$$\mathbb{P}^*_{\beta, p}(\hat{\beta}_{\mathrm{ML}} > \beta + t) = \mathbb{P}^*_{\beta, p}(H_n(\boldsymbol{X}) > \mathbb{E}^*_{\beta+t, p} H_n(\boldsymbol{X})) \leq \mathbb{P}^*_{\beta, p} \left(\overline{X}_n^p > \mathbb{E}^*_{\beta+t, p}(\overline{X}_n^p) - 6\gamma_n \right) .$$

Now, by part (2) of Lemma 8 and the dominated convergence theorem, we have:

$$\mathbb{E}^*_{\beta+t, p}(\overline{X}_n^p) \to m_*(\beta + t, p)^p .$$

Hence, once again by part (2) of Lemma 8, we have:

$$\mathbb{P}^*_{\beta, p}(\hat{\beta}_{\mathrm{ML}} > \beta + t) \leq \mathbb{P}^*_{\beta, p}(\overline{X}_n^p > m_*(\beta + t, p)^p - o(1)) \leq \mathbb{P}^*_{\beta, p} \left(|\overline{X}_n^p - m_*(\beta, p)^p| > \Omega(1) \right) = o(1).$$

Similarly, we can show that for every $t \in (0, \beta - \beta^*(p))$,

$$\mathbb{P}^*_{\beta,p}(\hat{\beta}_{\mathrm{ML}} < \beta - t) \leq \mathbb{P}^*_{\beta,p}(\overline{X}_n^p < m_*(\beta - t, p)^p + o(1)) \leq \mathbb{P}^*_{\beta,p}\left(|\overline{X}_n^p - m_*(\beta, p)^p| > \Omega(1)\right) = o(1).$$

Hence, we conclude that $\hat{\beta}_{\mathrm{ML}} \xrightarrow{P} \beta$ under the model $\mathbb{P}^*_{\beta,p}$. This proves Lemma 6 when $\hat{\beta} = \hat{\beta}_{\mathrm{ML}}$.

Now, suppose that $\hat{\beta} = \hat{\beta}_{\mathrm{MPL}}$. By part 1 of Lemma 8, we have:

$$\frac{1}{n} \log Z_n^*(\beta, p) = \frac{1}{n} \log Z_n(\beta, p) + o(1) \ .$$

By Theorem 2.3 in [39] and Lemma 13, we conclude that $\hat{\beta}_{\mathrm{MPL}}$ is a consistent estimator of β under the model $\mathbb{P}^*_{\beta,p}$. This completes the proof of Lemma 6.

A.5. **Proof of Lemma 7.** For any two sets $A, B \subseteq \{-1, 1\}^n$, note that:

$$\begin{aligned} &\mathbb{P}^*_{\beta_n,p}(A|B) \\ &= \frac{\sum_{\boldsymbol{x} \in A \cap B} \exp\{\beta_n H_n(\boldsymbol{x})\}}{\sum_{\boldsymbol{x} \in B} \exp\{\beta_n H_n(\boldsymbol{x})\}} \\ &= \frac{\sum_{\boldsymbol{x} \in A \cap B} \exp\{\beta_n \mathbb{E}H_n(\boldsymbol{x})\} e^{\beta_n(H_n(\boldsymbol{x}) - \mathbb{E}H_n(\boldsymbol{x}))}}{\sum_{\boldsymbol{x} \in B} \exp\{\beta_n \mathbb{E}H_n(\boldsymbol{x})\} e^{\beta_n(H_n(\boldsymbol{x}) - \mathbb{E}H_n(\boldsymbol{x}))}} \ . \end{aligned} \tag{A.4}$$

It follows from (A.4) and Lemma 1, that with probability 1, we have the following for all large n,

$$e^{-6n\beta_n\gamma_n} \mathbb{P}_{\beta_n,p}(A|B) \leq \mathbb{P}^*_{\beta_n,p}(A|B) \leq e^{6n\beta_n\gamma_n} \mathbb{P}_{\beta_n,p}(A|B) \ . \tag{A.5}$$

Lemma 7 now follows on taking logarithm on both sides of (A.4), and recalling that the sequence $\{\beta_n\}_{n \geq 1}$ is bounded.

A.6. **Proof of Lemma 8.**

Proof of 1. Using Lemma 1, we have the following for all large n, with probability 1:

$$e^{-3\beta n\gamma_n} Z_n(\beta, p) \leq Z_n^*(\beta, p) \leq e^{3\beta n\gamma_n} Z_n(\beta, p) \ . \tag{A.6}$$

Part 1 follows on taking logarithm on both sides of (A.6).

Proof of 2. To begin with, define:

$$M_\varepsilon := \begin{cases} (m_*(\beta, p) - \varepsilon, m_*(\beta, p) + \varepsilon) & \text{if } p \text{ is odd} \\ (m_*(\beta, p) - \varepsilon, m_*(\beta, p) + \varepsilon) \bigcup (-m_*(\beta, p) - \varepsilon, -m_*(\beta, p) + \varepsilon) & \text{if } p \text{ is even} \end{cases}$$

It follows from the arguments used in the proof of Lemma 3.1 in [38], that

$$\mathbb{P}_{\beta,p}(\overline{X}_n \in M_\varepsilon^c) = O(n^{\frac{3}{2}}) \exp\left\{n\left(\sup_{t \in M_\varepsilon^c} H_{\beta,p}(t) - H_{\beta,p}(m_*(\beta, p))\right)\right\} = e^{-n\Omega(1)}. \tag{A.7}$$

It follows from (A.7) and Lemma 7, that

$$\mathbb{P}^*_{\beta,p}(\overline{X}_n \in M_\varepsilon^c) \leq \mathbb{P}_{\beta,p}(\overline{X}_n \in M_\varepsilon^c) e^{O(n\gamma_n)} = e^{-n\Omega(1)}. \tag{A.8}$$

Hence, we have:

$$\mathbb{P}^*_{\beta,p}\left(|\overline{X}_n{}^p - m_*(\beta,p)^p| \geq \varepsilon\right) \leq \mathbb{P}^*_{\beta,p}(\overline{X}_n \in M^c_{\varepsilon/p}) = e^{-n\Omega(1)} \ .$$

This completes the proof of Lemma 8.

Appendix B. Technical Results for the Tensor Curie-Weiss Model

In this section, we prove some technical results, which will be useful in the proofs of the main results concerning the tensor Curie-Weiss model.

Lemma 9. We have the following:

$$\sup_{z\in\mathbb{R}}\{yz - \log\cosh(z)\} = \begin{cases} \frac{1}{2}\left\{(1+y)\log(1+y) + (1-y)\log(1-y)\right\} & \text{if } y \in [-1,1], \\ \infty & \text{otherwise .} \end{cases}$$

Proof. Fix $y \in \mathbb{R}$ and define $g(z) := yz - \log\cosh(z)$. Let us begin with the case $y \in (-1,1)$. In this case, $g''(z) = -\text{sech}^2(z) < 0$ and hence, g is a strictly concave function. Consequently, any stationary point of g is the unique global maximum of g. Since $g'(z) = y - \tanh(z)$, it follows that the only stationary point of g is $\tanh^{-1}(y)$, and hence,

$$\sup_{z\in\mathbb{R}}\{yz - \log\cosh(z)\} = y\tanh^{-1}y - \log\cosh(\tanh^{-1}(y)) = y\tanh^{-1}(y) + \frac{1}{2}\log(1-y^2) \ ,$$

where in the last step, we used the identity:

$$\cosh(\tanh^{-1}(y)) = \frac{1}{\sqrt{1-y^2}} \qquad \text{for } y \in (-1,1).$$

The proof for the case $y \in (-1,1)$ now follows from the observation that

$$y\tanh^{-1}(y) + \frac{1}{2}\log(1-y^2) = \frac{1}{2}\left\{(1+y)\log(1+y) + (1-y)\log(1-y)\right\} .$$

Now, suppose that $y \geq 1$. Then, $g'(z) > 0$ for all $z \in \mathbb{R}$, and hence,

$$\sup_{z\in\mathbb{R}} g(z) = \lim_{z\to\infty} g(z) \ .$$

Now, note that

$$\lim_{z\to\infty} e^{g(z)} = \lim_{z\to\infty} \frac{2e^{yz}}{e^z + e^{-z}} = \lim_{z\to\infty} \frac{2e^{(y-1)z}}{1 + e^{-2z}} = \begin{cases} 2 & \text{if } y = 1, \\ \infty & \text{if } y > 1. \end{cases}$$

Hence,

$$\lim_{z\to\infty} g(z) = \begin{cases} \log 2 & \text{if } y = 1, \\ \infty & \text{if } y > 1. \end{cases}$$

This completes the case $y \geq 1$. Finally, suppose that $y \leq -1$. Then, $g'(z) < 0$ for all $z \in \mathbb{R}$, and hence,

$$\sup_{z\in\mathbb{R}} g(z) = \lim_{z\to-\infty} g(z) \ .$$

Now, note that

$$\lim_{z \to -\infty} e^{g(z)} = \lim_{z \to -\infty} \frac{2e^{yz}}{e^z + e^{-z}} = \lim_{z \to -\infty} \frac{2e^{(y+1)z}}{1 + e^{2z}} = \begin{cases} 2 & \text{if } y = -1, \\ \infty & \text{if } y < -1. \end{cases}$$

Hence,

$$\lim_{z \to \infty} g(z) = \begin{cases} \log 2 & \text{if } y = -1, \\ \infty & \text{if } y < -1. \end{cases}$$

This completes the case $y \le -1$, and the proof of Lemma 9. $\qquad\square$

The following lemma describes the behavior of the function $H_{\beta,p}$.

Lemma 10. Suppose that $\beta > \beta^*(p)$. Then, the following are true.

(1) $H'_{\beta,2} > 0$ on $(0, m_*(\beta, 2))$ and $H'_{\beta,2} < 0$ on $(m_*(\beta, 2), 1)$.
(2) If $p \ge 3$, then $H_{\beta,p}$ can have at most 3 positive stationary points. Further, there exists $\underline{m}(\beta, p) \in (0, m_*(\beta, p))$ such that $H'_{\beta,p} \le 0$ on $(0, \underline{m}(\beta, p))$, $H'_{\beta,p} \ge 0$ on $(\underline{m}(\beta, p), m_*(\beta, p))$ and $H'_{\beta,p} \le 0$ on $(m_*(\beta, p), 1)$.

Proof. To begin with, define $N_{\beta,p}(x) := (1 - x^2)H''_{\beta_p}(x)$. Then,

$$N'_{\beta,p}(x) = \beta p(p - 1)x^{p-3}(p - 2 - px^2) .$$

Let us first consider the case $p \ge 3$. Since $N'_{\beta,p}$ has exactly 1 root in $(0, 1)$, it follows by repeated applications of Rolle's theorem, that $H'_{\beta,p}$ can have at most 3 roots in $(0, 1)$. Define:

$$\underline{m}(\beta, p) := \sup\{t \in (0, 1] : H'_{\beta,p} \le 0 \text{ on } (0, t]\} .$$

Since $H'_{\beta,p}(x) = \beta p x^{p-1} - \tanh^{-1}(x)$ and $\lim_{x \to 0} \tanh^{-1}(x)/x = 1$, we have $\underline{m}(\beta, p) > 0$. Clearly, $H'_{\beta,p} \le 0$ on $(0, \underline{m}(\beta, p)]$. On the other hand, since $m_*(\beta, p)$ is a global maximizer of $H_{\beta,p}$, and since $H_{\beta,p}$ can have at most finitely many stationary points, we must have $H'_{\beta,p}(x) > 0$ for some $x < m_*(\beta, p)$. This shows that $\underline{m}(\beta, p) < m_*(\beta, p)$. Now, by definition of $\underline{m}(\beta, p)$, there must exist a sequence $x_n \downarrow \underline{m}(\beta, p)$, such that $H'_{\beta,p}(x_n) > 0$ and $x_n > \underline{m}(\beta, p)$ for all n. Continuity of $H'_{\beta,p}$ now implies that $\underline{m}(\beta, p)$ is a stationary point of $H_{\beta,p}$.

We will now show that $H'_{\beta,p} \ge 0$ on $(\underline{m}(\beta, p), m_*(\beta, p))$. Suppose towards a contradiction, that $H'_{\beta,p}(y) < 0$ for some $y \in (\underline{m}(\beta, p), m_*(\beta, p))$. Then, there exist $y_1 \in (\underline{m}(\beta, p), y)$ and $y_2 \in (y, m_*(\beta, p))$, such that $H'_{\beta,p}(y_1) > 0$ and $H'_{\beta,p}(y_2) > 0$. This creates two extra stationary points of $H_{\beta,p}$, one within (y_1, y) and the other within (y, y_2), giving a total of at least 4 positive stationary points of $H_{\beta,p}$, a contradiction! Hence, $H'_{\beta,p} \ge 0$ on $(\underline{m}(\beta, p), m_*(\beta, p))$.

Finally, we show that $H'_{\beta,p} \le 0$ on $(m_*(\beta, p), 1)$. Once again, suppose towards a contradiction, that $H'_{\beta,p}(y) > 0$ for some $y \in (m_*(\beta, p), 1)$. Since $m_*(\beta, p)$ is a global maximizer of $H_{\beta,p}$ and $\lim_{x \to 1} H'_{\beta,p}(x) = -\infty$, there exist $y_1 \in (m_*(\beta, p), y)$ and $y_2 \in (y, 1)$, such that $H'_{\beta,p}(y_1) < 0$ and $H'_{\beta,p}(y_2) < 0$. This creates two extra stationary points of $H_{\beta,p}$, one within (y_1, y) and the other within (y, y_2), giving a total of at least 4 positive stationary points of

$H_{\beta,p}$, a contradiction! Hence, $H'_{\beta,p} \le 0$ on $(m_*(\beta, p), 1)$. This completes the proof of part (2) of Lemma 10.

Now, suppose that $p = 2$. Since $N'_{\beta,2}$ has exactly one root in $(-1, 1)$, it follows by repeated applications of Rolle's theorem, that $H'_{\beta,2}$ can have at most 3 roots in $(-1, 1)$. Since $H'_{\beta,2}$ is an odd function, it follows that it can have at most 1 positive root, which must be $m_*(\beta, 2)$. Hence, $H'_{\beta,2}$ must be non-zero and cannot change sign on each of the intervals $(0, m_*(\beta, 2))$ and $(m_*(\beta, 2), 1)$. Since $m_*(\beta, 2)$ is a global maximizer of $H_{\beta,2}$, we must have $H'_{\beta,2}(x) > 0$ for some $x \in (0, m_*(\beta, 2))$, and since $\lim_{x \to 1} H'_{\beta,2}(x) = -\infty$, we must have $H'_{\beta,2}(x) < 0$ for some $x \in (m_*(\beta, 2), 1)$. Hence, we must have $H'_{\beta,2}(x) > 0$ for all $x \in (0, m_*(\beta, 2))$ and $H'_{\beta,2}(x) < 0$ for all $x \in (m_*(\beta, 2), 1)$. This proves (1), and completes the proof of Lemma 10. $\qquad\square$

Lemma 11. The function $\xi_p(\beta) := m_*(\beta, p)$ is continuous and strictly increasing on $(\beta^*(p), \infty)$. Further,

$$\lim_{\beta \to \infty} \xi_p(\beta) = 1 .$$

Proof. Fix $\beta \in (\beta^*(p), \infty)$ and take a sequence $\beta_n \to \beta$. Then, $\beta_n \in (\beta^*(p), \infty)$ for all large n, and hence, $H_{\beta_n, p}$ will have a unique global maximizer $m_*(\beta_n, p) \in (0, 1)$ for all large n. Take a subsequence $\{n_k\}_{k \ge 1}$ of the positive integers. This subsequence must have a further subsequence $\{n_{k_\ell}\}_{\ell \ge 1}$ such that $m_*(\beta_{n_{k_\ell}}, p) \to m'$ for some $m' \in [0, 1]$. Clearly, $H_{\beta_{n_{k_\ell}}, \, p}(m_*(\beta_{n_{k_\ell}}, p)) \to H_{\beta,p}(m')$. Since $H_{\beta_{n_{k_\ell}}, \, p}(m_*(\beta_{n_{k_\ell}}, p)) \ge H_{\beta_{n_{k_\ell}}, \, p}(x)$ for all $x \in [0, 1]$ and for all large ℓ, we must have $H_{\beta,p}(m') \ge H_{\beta,p}(x)$ for all $x \in [0, 1]$ (taking $\lim_{\ell \to \infty}$ on both sides). This means that m' is a non-negative global maximizer of $H_{\beta,p}$. Since $m_*(\beta, p)$ is the only non-negative global maximizer of $H_{\beta,p}$, it follows that $m' = m_*(\beta, p)$. Hence, $m_*(\beta_{n_{k_\ell}}, p) \to m_*(\beta, p)$, showing that $\xi_p(\beta_n) \to \xi_p(\beta)$, and thereby establishing continuity of ξ_p.

Next, take any $t \in (0, 1)$, whence $H'_{\beta,p}(t) = \beta p t^{p-1} - \tanh^{-1}(t) > 0$ for all β large enough. On the other hand, it follows from Lemma 10, that $H'_{\beta,p} \le 0$ on $[m_*(\beta, p), 1]$. This shows that $m_*(\beta, p) > t$ for all β large enough, showing that $\lim_{\beta \to \infty} \xi_p(\beta) = 1$.

Finally, to show that ξ_p is increasing on $(\beta^*(p), \infty)$, take $\beta_2 > \beta_1 > \beta^*(p)$. Then, by Lemma 10, $H'_{\beta_2,p} \le 0$ on $[m_*(\beta_2, p), 1]$. Since $H'_{\beta_1,p} < H'_{\beta_2,p}$ on $(0, 1]$, we must have $H'_{\beta_1,p} < 0$ on $[m_*(\beta_2, p), 1]$. However, since $m_*(\beta_1, p)$ is a global maximizer of $H_{\beta_1,p}$, and since $H_{\beta_1,p}$ can have at most finitely many stationary points, there must exist $\varepsilon > 0$, such that $H'_{\beta_1,p} > 0$ on $(m_*(\beta_1, p) - \varepsilon, m_*(\beta_1, p))$. Continuity of $H'_{\beta_1,p}$ now implies that $m_*(\beta_2, p) > m_*(\beta_1, p)$, proving that ξ_p is strictly increasing. This completes the proof of Lemma 11. $\qquad\square$

APPENDIX C. TECHNICAL RESULTS FOR THE HYPERGRAPH ERDŐS-RÉNYI ISING MODEL

In this section, we prove some technical results related to the hypergraph Erdős-Rényi Ising model. We start with the proof of Lemma 1.

C.1. **Proof of Lemma 1.** To begin with, for every $\boldsymbol{x} \in \{-1, 1\}^n$, let us define the set:

$$\Lambda_n(\boldsymbol{x}) := \{(i_1, \ldots, i_p) \in [n]^p : x_{i_1} \ldots x_{i_p} = 1\} .$$

Also, let $\mathscr{L}_n(\boldsymbol{x}) := \sum_{(i_1,\ldots,i_p) \in \Lambda_n(\boldsymbol{x})} A_{i_1 \ldots i_p}$. In these notations, we have:

$$H_n(\boldsymbol{x}) = \alpha_n^{-1} n^{1-p} \left(2\mathscr{L}_n(\boldsymbol{x}) - \sum_{(i_1,\ldots,i_p) \in [n]^p} A_{i_1 \ldots i_p} \right) .$$

For each $\gamma > 0$, define an event:

$$\Omega_n(\gamma) := \left\{ \sup_{\boldsymbol{x} \in \{-1,1\}^n} \left| \frac{\mathscr{L}_n(\boldsymbol{x})}{\mathbb{E}\mathscr{L}_n(\boldsymbol{x})} - 1 \right| \leq \gamma \right\} .$$

Since $\sup_{\boldsymbol{x} \in \{-1,1\}^n} |\mathbb{E}\mathscr{L}_n(\boldsymbol{x})| \leq \alpha_n n^p$, we have the following on the event $\Omega_n(\gamma_n)$:

$$\frac{1}{n} \sup_{\boldsymbol{x} \in \{-1,1\}^n} |H_n(\boldsymbol{x}) - \mathbb{E}H_n(\boldsymbol{x})|$$

$$\leq 2\alpha_n^{-1} n^{-p} \sup_{\boldsymbol{x} \in \{-1,1\}^n} |\mathscr{L}_n(\boldsymbol{x}) - \mathbb{E}\mathscr{L}_n(\boldsymbol{x})| + \left| \alpha_n^{-1} n^{-p} \sum_{(i_1,\ldots,i_p) \in [n]^p} A_{i_1 \ldots i_p} - 1 \right|$$

$$\leq 2\gamma_n + \left| \alpha_n^{-1} n^{-p} \sum_{(i_1,\ldots,i_p) \in [n]^p} A_{i_1 \ldots i_p} - 1 \right| . \tag{C.1}$$

It follows from Theorem 4 in [24], that

$$\mathbb{P}\left(\left| \alpha_n^{-1} n^{-p} \sum_{(i_1,\ldots,i_p) \in [n]^p} A_{i_1 \ldots i_p} - 1 \right| > \gamma_n \right) \leq 2e^{-\frac{1}{3}\gamma_n^2 \alpha_n n^p} = 2e^{-3n} \tag{C.2}$$

In view of (C.1), (C.2) and the Borel-Cantelli Lemma, it thus suffices to show that

$$\mathbb{P}\left(\Omega_n(\gamma_n) \text{ occurs for all but finitely many } n \right) = 1 , \tag{C.3}$$

in order to complete the proof of Lemma 1. Towards this, note that by a union bound,

$$\mathbb{P}(\Omega_n(\gamma)^c) \leq \sum_{\boldsymbol{x} \in \{-1,1\}^n} \mathbb{P}\left(\mathscr{L}_n(\boldsymbol{x}) > (1+\gamma)\mathbb{E}\mathscr{L}_n(\boldsymbol{x}) \right) + \sum_{\boldsymbol{x} \in \{-1,1\}^n} \mathbb{P}\left(\mathscr{L}_n(\boldsymbol{x}) < (1-\gamma)\mathbb{E}\mathscr{L}_n(\boldsymbol{x}) \right) . \tag{C.4}$$

It follows from Theorem 1 in [31], that

$$\mathbb{P}\left(\mathscr{L}_n(\boldsymbol{x}) > (1+\gamma)\mathbb{E}\mathscr{L}_n(\boldsymbol{x}) \right) = \mathbb{P}\left(\frac{\mathscr{L}_n(\boldsymbol{x})}{|\Lambda_n(\boldsymbol{x})|} > (1+\gamma)\alpha_n \right) \leq e^{-|\Lambda_n(\boldsymbol{x})|D((1+\gamma)\alpha_n \| \alpha_n)} , \tag{C.5}$$

and

$$\mathbb{P}\left(\mathscr{L}_n(\boldsymbol{x}) < (1-\gamma)\mathbb{E}\mathscr{L}_n(\boldsymbol{x}) \right) = \mathbb{P}\left(\frac{\mathscr{L}_n(\boldsymbol{x})}{|\Lambda_n(\boldsymbol{x})|} < (1-\gamma)\alpha_n \right) \leq e^{-|\Lambda_n(\boldsymbol{x})|D((1-\gamma)\alpha_n \| \alpha_n)} , \tag{C.6}$$

where $D(x\|y) := x \log \frac{x}{y} + (1-x) \log\left(\frac{1-x}{1-y}\right)$. Also, let

$$\mathcal{M}_n := \left\{ -1, -1 + \frac{2}{n}, \ldots, 1 - \frac{2}{n}, 1 \right\}$$

denote the set of all values $\overline{x}_n := n^{-1} \sum_{i=1}^n x_i$ can take, for some $\boldsymbol{x} \in \{-1, 1\}^n$. Combining (C.4), (C.5) and (C.6), we have by Lemma 12 and Equation (2.17) in [12],

$$
\begin{aligned}
&\mathbb{P}(\Omega_n(\gamma)^c) \\
&\leq \sum_{\boldsymbol{x} \in \{-1,1\}^n} \left\{ e^{-|\Lambda_n(\boldsymbol{x})|D((1+\gamma)\alpha_n\|\alpha_n)} + e^{-|\Lambda_n(\boldsymbol{x})|D((1-\gamma)\alpha_n\|\alpha_n)} \right\} \\
&= \sum_{m \in \mathcal{M}_n} \binom{n}{n(1+m)/2} \left\{ e^{-\frac{1}{2}n^p(1+m^p)D((1+\gamma)\alpha_n\|\alpha_n)} + e^{-\frac{1}{2}n^p(1+m^p)D((1-\gamma)\alpha_n\|\alpha_n)} \right\} \\
&= e^{-n\left[\frac{n^{p-1}}{2}D((1+\gamma)\alpha_n\|\alpha_n) - \log 2\right] - \frac{\log n}{2} + O(n^{-1})} \sum_{m \in \mathcal{M}_n} e^{-\frac{n^p m^p}{2}D((1+\gamma)\alpha_n\|\alpha_n) - nI(m)} \\
&\quad + e^{-n\left[\frac{n^{p-1}}{2}D((1-\gamma)\alpha_n\|\alpha_n) - \log 2\right] - \frac{\log n}{2} + O(n^{-1})} \sum_{m \in \mathcal{M}_n} e^{-\frac{n^p m^p}{2}D((1-\gamma)\alpha_n\|\alpha_n) - nI(m)} , \quad \text{(C.7)}
\end{aligned}
$$

where $I(t) := \frac{1}{2}(1 + t)\log(1 + t) + \frac{1}{2}(1 - t)\log(1 - t)$. Since the functions D and I are non-negative, we have

$$
\sum_{m \in \mathcal{M}_n} e^{-\frac{n^p m^p}{2}D((1\pm\gamma)\alpha_n\|\alpha_n) - nI(m)} \leq |\mathcal{M}_n| = n + 1 .
$$

Hence, we have from (C.7) and Equation (2.30) in [12],

$$
\begin{aligned}
\mathbb{P}(\Omega_n(\gamma_n)^c) &\leq O(\sqrt{n}) \left(e^{-n\left[\frac{n^{p-1}}{2}D((1+\gamma_n)\alpha_n\|\alpha_n) - \log 2\right]} + e^{-n\left[\frac{n^{p-1}}{2}D((1-\gamma_n)\alpha_n\|\alpha_n) - \log 2\right]} \right) \\
&\leq O(\sqrt{n}) \left(\exp\left\{ -n\left[\alpha_n n^{p-1}\frac{\gamma_n^2}{6} - \log 2 \right] \right\} + \exp\left\{ -n\left[\alpha_n n^{p-1}\frac{\gamma_n^2}{4} - \log 2 \right] \right\} \right) \\
&\leq O(\sqrt{n}) \exp(-0.8n) . \quad \text{(C.8)}
\end{aligned}
$$

Since $\sum_{n=1}^\infty \mathbb{P}(\Omega_n(\gamma_n)^c) < \infty$, (C.3) follows from (C.8) and the Borel-Cantelli lemma, completing the proof of Lemma 1. □

C.2. **Proof of Corollary 1.** For each $1 \leq i \leq n$, define $A_{i_2 \dots i_p}^{(i)} := A_{i i_2 \dots i_p}$. Then, for each $1 \leq i \leq n$, one can view $m_i(\boldsymbol{X})$ as the Hamiltonian (scaled by n^{-1}) of the $(p-1)$-spin Erdős-Rényi Ising model with adjacency tensor $\boldsymbol{A}^{(i)} := ((A_{i_2 \dots i_p}^{(i)}))$. The rest of the proof will follow exactly as the proof of Lemma 1.

C.3. **Other Technical Lemmas.** In this section, we prove some other technical lemmas required for the proof of the main results in Section 3. We begin with deriving the cardinality of $\Lambda_n(\boldsymbol{x})$ that was required in the proof of Lemma 1.

Lemma 12. For every $\boldsymbol{x} \in \{-1, 1\}^n$, we have:

$$
|\Lambda_n(\boldsymbol{x})| = \frac{1}{2}n^p(1 + \overline{x}_n{}^p) ,
$$

where $\overline{x}_n := n^{-1} \sum_{i=1}^n x_i$.

Proof. First, note that $(i_1, \ldots, i_p) \in \Lambda_n(\boldsymbol{x})$ if and only if $x_{i_\ell} = -1$ for an even number of $\ell \in [p] := \{1, \ldots, p\}$. Now, it is easy to see that the number of indices $i \in [n]$ for which $x_i = -1$, is given by $n(1 - \overline{x}_n)/2$. To form an $(i_1, \ldots, i_p) \in \Lambda_n(\boldsymbol{x})$, we must thus choose an even number of these p indices from the total number of $n(1 - \overline{x}_n)/2$ possible indices where we have -1, and the rest of these p indices from the remaining $n(1 + \overline{x}_n)/2$ number of possible indices where we have $+1$. We thus have:

$$
\begin{aligned}
|\Lambda_n(\boldsymbol{x})| &= \sum_{k \in [p] \bigcup \{0\}: \ k \text{ is even}} \binom{p}{k} \left(\frac{n(1 - \overline{x}_n)}{2} \right)^k \left(\frac{n(1 + \overline{x}_n)}{2} \right)^{p-k} \\
&= \frac{1}{2} \left(\frac{n(1 + \overline{x}_n)}{2} + \frac{n(1 - \overline{x}_n)}{2} \right)^p + \frac{1}{2} \left(\frac{n(1 + \overline{x}_n)}{2} - \frac{n(1 - \overline{x}_n)}{2} \right)^p . \quad \text{(C.9)}
\end{aligned}
$$

Lemma 12 follows from (C.9). $\qquad \square$

The following lemma is crucial in showing consistency of the MPL estimator in the hypergraph Erdős-Rényi Ising model.

Lemma 13. With probability 1, we have the following:

$$
\max_{1 \leq i_1 \leq n} \sum_{(i_2, \ldots, i_p) \in [n]^{p-1}} A_{i_1 \ldots i_p} = O\left(\alpha_n n^{p-1} \right) .
$$

Proof. Note that $\sum_{(i_2, \ldots, i_p) \in [n]^{p-1}} A_{i_1 \ldots i_p} \sim \text{Bin}(n^{p-1}, \alpha_n)$. So, by Theorem 4 in [24], we have:

$$
\mathbb{P}\left(\alpha_n^{-1} n^{1-p} \sum_{(i_2, \ldots, i_p) \in [n]^{p-1}} A_{i_1 \ldots i_p} \geq 1 + \delta \right) \leq e^{-\frac{\delta^2}{2+\delta} \alpha_n n^{p-1}}
$$

for every $\delta > 0$. Hence,

$$
\mathbb{P}\left(\max_{1 \leq i_1 \leq n} \sum_{(i_2, \ldots, i_p) \in [n]^{p-1}} A_{i_1 \ldots i_p} \geq 2(1 + \delta) \alpha_n n^{p-1} \right) \leq n e^{-\frac{\delta^2}{2+\delta} \alpha_n n^{p-1}} = e^{\log n - \frac{\delta^2}{2+\delta} \alpha_n n^{p-1}} .
$$

Since $\alpha_n = \Omega(n^{1-p} \log n)$, we can choose $\delta > 0$ large enough, so that $\log n - \frac{\delta^2}{2+\delta} \alpha_n n^{p-1} \leq -2 \log n$ thereby ensuring that

$$
\mathbb{P}\left(\max_{1 \leq i_1 \leq n} \sum_{(i_2, \ldots, i_p) \in [n]^{p-1}} A_{i_1 \ldots i_p} \geq 2(1 + \delta) \alpha_n n^{p-1} \right) \leq n^{-2} .
$$

It now follows by an application of the Borel-Cantelli lemma, that:

$$
\mathbb{P}\left(\max_{1 \leq i_1 \leq n} \sum_{(i_2, \ldots, i_p) \in [n]^{p-1}} A_{i_1 \ldots i_p} \leq 2(1 + \delta) \alpha_n n^{p-1} \text{ for all large n} \right) = 1 ,
$$

which completes the proof of Lemma 13. $\qquad \square$

DEPARTMENT OF STATISTICS AND DATA SCIENCE, NATIONAL UNIVERSITY OF SINGAPORE `somabha@nus.edu.sg`

DEPARTMENT OF STATISTICS, COLUMBIA UNIVERSITY `js4638@columbia.edu`

DEPARTMENT OF STATISTICS, STANFORD UNIVERSITY `raswa281@stanford.edu`

DEPARTMENT OF STATISTICS, UNIVERSITY OF FLORIDA `souravmukherjee@ufl.edu`