BAHADUR EFFICIENCY IN SOME TENSOR ISING MODELS

SOMABHA MUKHERJEE, SWARNADIP GHOSH, JAESUNG SON, 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. In this exponential family, the sufficient statistic is a multi-linear form of degree \( p \geq 2 \), designed to capture \( p \)-fold interactions between the binary variables sitting on the nodes of a network. A particularly useful class of tensor Ising models are the tensor Curie-Weiss models, where one assumes that all \( p \)-tuples of nodes interact with the same intensity. Computing the maximum likelihood estimator (MLE) is computationally cumbersome in this model, due to the presence of an inexplicit normalizing constant in the likelihood, for which the standard alternative is to use the maximum pseudolikelihood estimator (MPLE). Both the MLE and the MPLE are consistent estimators of the natural parameter, provided the latter lies strictly above a certain threshold, which is slightly below \( \log 2 \), and approaches \( \log 2 \) as \( p \) increases. In this paper, we compute the Bahadur efficiencies of the MLE and the MPLE above the threshold, and derive the optimal sample size (number of nodes) needed for either of these tests to achieve significance. We show that the optimal sample size for the MPLE and the MLE agree if either \( p = 2 \) or the null parameter is greater than or equal to \( \log 2 \). On the other hand, if \( p \geq 3 \) and the null parameter lies strictly between the threshold and \( \log 2 \), then the two differ for sufficiently large values of the alternative. In particular, for every fixed alternative above the threshold, the Bahadur asymptotic relative efficiency of the MLE with respect to the MPLE goes to \( \infty \) as the null parameter approaches the threshold. We also provide graphical presentations of the exact numerical values of the theoretical optimal sample sizes in different settings. Finally, we show a universality phenomenon, which says that these results extend beyond the tensor Curie-Weiss model, and hold for the more general class of Erdős-Rényi hypergraph Ising models, where we can even allow for some sparsity in the underlying Erdős-Rényi hypergraph.

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 [24]. It has found immense applications in diverse places such as image processing [16], neural networks [23], spatial statistics [7], and disease mapping in epidemiology [19].

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 [28], 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 \cite{28}, 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 \cite{27}, which is a discrete exponential family on the hypercube \( \{-1,1\}^n \), with probability mass function given by:

\[
P_{\beta,p}(x) := \frac{\exp \left\{ \beta n^{1-p} \sum_{1 \leq i_1, \ldots, i_p \leq n} x_{i_1} \cdots x_{i_p} \right\}}{2^n Z_n(\beta,p)} \quad \text{for } x \in \{-1,1\}^n.
\]  

(1.1)

Here \( Z_n(\beta,p) \) is a normalizing constant required to ensure that \( \sum_{x \in \{-1,1\}^n} P_{\beta,p}(x) = 1 \), and \( \beta \geq 0 \). It is precisely this inexplicit normalizing constant \( Z_n(\beta,p) \), that hinders estimation of the parameter \( \beta \) using the maximum likelihood approach. In this context, and also for more general Ising models, an alternative, computationally efficient algorithm was suggested by Chatterjee \cite{14} for the case \( p = 2 \), which goes by the name of maximum pseudolikelihood (MPL) estimation \cite{8, 9}, and is based on computing explicit conditional distributions. To elaborate, the MPL estimate is obtained by maximizing the pseudolikelihood function:

\[
\hat{\beta}_{\text{MPL}} := \arg \max_{\beta \in \mathbb{R}} \prod_{i=1}^{n} P_{\beta,p}(X_i | (X_j)_{j \neq i}) ,
\]

where \( X = (X_1, \ldots, X_n) \) is simulated from the model (1.1). The \( \sqrt{n} \)-consistency of the MPL estimator in the so called low temperature regime (high values of the parameter \( \beta \)) was established in \cite{14}, and later extended to tensor Ising models for \( p > 2 \) in \cite{28}. More precisely, it is shown in \cite{28} and \cite{27} that there exists \( \beta^*(p) > 0 \), such that for all \( \beta > \beta^*(p) \), both \( \sqrt{n}(\hat{\beta}_{\text{MPL}} - \beta) \) and \( \sqrt{n}(\hat{\beta}_{\text{ML}} - \beta) \) are tight, where \( \hat{\beta}_{\text{ML}} \) is the Maximum Likelihood (ML) estimator of \( \beta \). Further, consistent estimation (and consistent testing) is impossible in the regime \([0, \beta^*(p)]\). The exact asymptotics of both \( \sqrt{n}(\hat{\beta}_{\text{MPL}} - \beta) \) and \( \sqrt{n}(\hat{\beta}_{\text{ML}} - \beta) \) for the model (1.1) for \( \beta \) above the threshold \( \beta^*(p) \), was worked out in \cite{28}, where it is shown that both these statistics converge in distribution to the same normal distribution (see Figure 1). Consequently, both the ML and MPL estimates have the same asymptotic variance everywhere above the threshold, and in fact, both saturate the Cramer-Rao information lower bound in this regime. One of the main goals of this paper is to calculate and compare another notion of efficiency of estimators, called Bahadur efficiency for the ML and MPL estimators in the model (1.1).

In his seminal paper \cite{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 \cite{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 \( \delta \) for all \( n \geq N(\delta) \), i.e. the \( p \)-value corresponding to \( T_n \) becomes (and remains) bounded by \( \delta \) for all \( n \geq N(\delta) \). If 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 \( F_{T_n, \theta_0} \) 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 \( \theta \) [1, 2, 3, 4, 5]. In particular, if under \( \mathbb{P}_\theta \) we have the following with probability 1:

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

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 \( \theta \). 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 \( \theta \), \( n^{-1/2} T_n \to b(\theta) \) 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 \( \theta \), 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 [20]). Finally, if we have two competing estimators \( T_{n,1} \) and \( T_{n,2} \) estimating the same parameter \( \theta \), then the Bahadur asymptotic relative efficiency (ARE) is given
by the ratio of their Bahadur slopes (see [20]):

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

In this paper, we derive the weak Bahadur slopes of both \( \hat{\beta}_{\text{MPL}} \) and \( \hat{\beta}_{\text{ML}} \) in the tensor Curie-Weiss model (1.1). This will in turn, enable one to compute the Bahadur relative efficiency between any of these two estimators with some other reference estimator. Similar results have been derived in [13] in the context of Markov random fields on lattices, and in [21] 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 models. Our basic tool will be some recent results on large deviation of the average magnetization \( X_n := \frac{1}{n} \sum_{i=1}^{n} X_i \) in the Curie-Weiss model, established in [26] and [31]. Also, throughout the rest of the paper, we will view the entries \( x_1, \ldots, x_n \) of the tuple \( x \in \{-1, 1\}^n \) as dependent samples, and refer to the length \( n \) of \( x \) as the sample size (although technically speaking, we have just one multivariate sample \( x \) from the model (1.1)). One of our most interesting findings is that for \( p \geq 3 \), the Bahadur slopes and optimal sample sizes for the tests based on the MPL and ML estimators do not agree for each value of the null parameter \( \beta_0 \) in \( \mathcal{R} := (\beta^*(p), \log 2) \), provided the alternative parameter \( \beta \) is sufficiently large. This loss of Bahadur efficiency for the MPL estimator near threshold, can be attributed to its functional form, which derives false signal from a regime where the average magnetization \( X_n \) is very close to 0. Our results actually exhibit a certain universality phenomenon beyond the Curie-Weiss model (1.1). They hold ditto in tensor Erdős-Rényi Ising models (3.1) too, which is an exponential family, with sufficient statistic given by a tensor form, the tensor being the adjacency tensor 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 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 derive the Bahadur slopes and the optimal sample sizes for the tests based on the ML and the MPL estimators. Section 3 deals with extensions of these results for the tensor Curie-Weiss model to the tensor Erdős-Rényi Ising model. In Section 4, we provide numerical illustrations of our theoretical findings in various settings. In Section 5, we summarize some main and interesting aspects of our results, and talk about directions for future research in this area. Finally, proofs of some technical results needed for showing the main theorems (Theorem 1 in Section 2 and Theorem 2 in Section 3) are given in the appendix.

2. Theoretical Results

In this section, we derive the weak Bahadur slopes of the MPL and ML estimators of \( \beta \) in the model (1.1). The ML estimator \( \hat{\beta}_{\text{ML}} \) does not have an explicit form, but it is shown in [28] that the MPL estimator is given by:

\[
\hat{\beta}_{\text{MPL}} = \begin{cases} 
p^{-1}X_n^{-1-p} \tanh^{-1}(X_n) & \text{if } X_n \neq 0, \\
0 & \text{if } X_n = 0.
\end{cases}
\]

Furthermore, it is shown in [27] and [28] that both the ML and MPL estimators have the same asymptotic normal distribution:

\[
\sqrt{N} (\hat{\beta} - \beta) \overset{D}{\rightarrow} N \left( 0, \frac{H_{\beta,p}(m_*(\beta,p))}{p^2 m_*(\beta,p)^{2p-2}} \right),
\]
for all $\beta > \beta^*(p)$, where $\hat{\beta}$ is either $\hat{\beta}_{ML}$ or $\hat{\beta}_{MPL}$,

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

(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 [28]).

In this paper, we will consider testing the hypothesis

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

for some known $\beta_0 > \beta^*(p)$. The most powerful test is based on the sufficient statistic $X_n$, and its asymptotic power is derived in [11]. 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 $\hat{\beta}_{ML}$ or $\hat{\beta}_{MPL}$, and large values of $T_n$ will denote significance. The main goal of this section is to prove conditions (1) and (2) in [6] for deriving the exact Bahadur slope of $T_n$. We begin with the proof of condition (1) with almost sure convergence replaced by convergence in probability.

**Lemma 1.** Under every $\beta > \beta^*(p)$, we have:

$$n^{-1/2} T_n \xrightarrow{P} \beta - \beta_0 \quad \text{as } n \to \infty ,$$
Figure 3. Optimal sample size for the tests based on MLE and MPLE with varying \( \beta_0; \ p = 3, \beta = 0.90, \delta = 0.05 \).

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

Proof. Note that \( n^{-1/2}T_n = \hat{\beta} - \beta_0 \), where \( \hat{\beta} \) is either \( \hat{\beta}_{ML} \) or \( \hat{\beta}_{MPL} \). It follows from [27] and [28], that under every \( \beta > \beta^*(p) \), \( \hat{\beta} \xrightarrow{P} \beta \). This proves Lemma 1.

We will now prove condition 2. For this, we will need the following lemma on the large deviation of \( X_n \), which follows from [31].

**Lemma 2.** For every subset \( A \subseteq [-1,1] \) such that \( A^c \) is dense in \( \overline{A} \), we have:

\[
\lim_{n \to \infty} \frac{1}{n} \log P_{\beta,p}(X_n \in A) = \sup_{x \in A} H_{\beta,p}(x) - \sup_{x \in [-1,1]} H_{\beta,p}(x).
\]

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

Proof. It follows from display (18) in [31] that \( X_n \) satisfies a large deviation principle (LDP) with rate function:

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

Using the identity \( \sinh^{-1}(x) = \log \left( x + \sqrt{x^2 + 1} \right) \).
we have:

\[
-\beta x^p + \frac{x}{2} \sinh^{-1} \left( \frac{2x}{1-x^2} \right) + \frac{1}{2} \log (1 - x^2) \\
= -\beta x^p + \frac{x}{2} \log \left( \frac{2x}{1-x^2} + \sqrt{\frac{4x^2}{(1-x^2)^2} + 1} \right) + \frac{1}{2} \log (1 - x^2) \\
= -\beta x^p + \frac{x}{2} \log \left( \frac{(1+x)^2}{1-x^2} \right) + \frac{1}{2} \log (1 - x^2) \\
= -\beta x^p + x \log(1+x) + \frac{1-x}{2} \log(1-x) \\
= -\beta x^p + x \log(1+x) + \frac{1-x}{2} \log(1+x) + \frac{1-x}{2} \log(1-x) \\
= -\beta x^p \{ (1 + x) \log (1 + x) + (1 - x) \log (1 - x) \} = -H_{\beta,p}(x) 
\]

It now follows from Lemma 7 that

\[
I(x) = -H_{\beta,p}(x) + \sup_{y \in [-1,1]} H_{\beta,p}(y) .
\]  
(2.2)

Lemma 2 now follows from (2.2) and the fact that \( I \) is a continuous function on \([-1,1]\).

We now state and prove the main result in this paper about the Bahadur slope of the test based on the MPL and ML estimators, and the minimum sample size required to ensure its significance.
Towards this, we define a function \( \eta_p : [-1, 1] \rightarrow \mathbb{R} \) 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 \( \hat{\beta}_{\text{MPL}} \) and \( \hat{\beta}_{\text{ML}} \) for the model (1.1) at an alternative \( \beta \) are respectively given by:

\[
c_{\beta_{\text{MPL}}} (\beta_0, \beta, p) = 2 \left( \sup_{x \in [-1, 1]} H_{\hat{\beta}_{0}, p}(x) - \sup_{x \in \eta_p^{-1}((\beta, \infty))} H_{\hat{\beta}_{0}, p}(x) \right),
\]

\[
c_{\beta_{\text{ML}}} (\beta_0, \beta, p) = 2 \left( \sup_{x \in [-1, 1]} H_{\hat{\beta}_{0}, p}(x) - \sup_{x > m_*(\beta, p)} H_{\hat{\beta}_{0}, p}(x) \right)
\]

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

\[
N_{\beta_{\text{MPL}}} (\beta_0, \beta, \delta, p) = \log(\delta) \left( \sup_{x \in \eta_p^{-1}((\beta, \infty))} H_{\hat{\beta}_{0}, p}(x) - \sup_{x \in [-1, 1]} H_{\hat{\beta}_{0}, p}(x) \right),
\]

\[
N_{\beta_{\text{ML}}} (\beta_0, \beta, \delta, p) = \log(\delta) \left( \sup_{x > m_*(\beta, p)} H_{\hat{\beta}_{0}, p}(x) - \sup_{x \in [-1, 1]} H_{\hat{\beta}_{0}, p}(x) \right).
\]

**Proof.** We deal with the test based on the MPL estimator first. To begin with, note that \( \hat{\beta}_{\text{MPL}} = \eta_p(X_n) \). Fix \( t > 0 \), whence we have by Lemma 2:

\[
n^{-1} \log (1 - F_{T_n, \hat{\beta}_0}(n^{1/2}t)) = n^{-1} \log p_{\beta_0, p} \left( n^{-1/2} T_n > t \right)
\]

\[
= n^{-1} \log p_{\beta_0, p} \left( \hat{\beta}_{\text{MPL}} - \beta_0 > t \right)
\]

\[
= n^{-1} \log p_{\beta_0, p} \left( \eta_p(X_n) > \beta_0 + t \right)
\]

\[
= n^{-1} \log p_{\beta_0, p} \left( X_n \in \eta_p^{-1}((\beta_0 + t, \infty)) \right)
\]

\[
= \sup_{x \in \eta_p^{-1}((\beta_0 + t, \infty))} H_{\hat{\beta}_0, p}(x) - \sup_{x \in [-1, 1]} H_{\hat{\beta}_0, p}(x) + o(1).
\]

where the last step follows from Lemma 2, since it follows from the proof of Lemma 9, 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:

\[
f(t) = \sup_{x \in [-1, 1]} H_{\hat{\beta}_0, p}(x) - \sup_{x \in \eta_p^{-1}((\beta_0 + t, \infty))} H_{\hat{\beta}_0, p}(x).
\]

Since \( x \mapsto H_{\hat{\beta}_0, p}(x) \) and \( \beta \mapsto m_*(\beta, p) \) are continuous functions (by Lemma 10) on \([-1, 1]\) and \((\beta^*(p), \infty)\) respectively, we conclude (in view of Lemma 9) that \( f \) is continuous on an open neighborhood of \( \beta - \beta_0 \). Also, in view of Lemma 9, 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 \( \hat{\beta}_{\text{MPL}} \) at an alternative \( \beta \) 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:

$$n^{-1} \log[1 - F_{T_n, \beta_0}(n^{1/2}t)] = n^{-1} \log P_{\beta_0, p_0}(n^{-1/2}T_n > t)$$

$$= n^{-1} \log P_{\beta_0, p_0}(\hat{\beta}_{ML} - \beta_0 > t)$$

$$= n^{-1} \log P_{\beta_0, p_0}(\overline{X}_n^p > E_{\beta_0 + t, p}(\overline{X}_n^p)).$$

The last step follows from the following facts:

1. The function $F_n(\beta, p) := \log Z_n(\beta, p)$ is strictly convex in $\beta$ (Lemma C.5 in [27]) and hence, $\frac{\partial F_n(\beta, p)}{\partial \beta}$ is strictly increasing in $\beta$;
2. The ML equation is given by $\frac{\partial F_n(\beta, p)}{\partial \beta}\big|_{\beta = \beta_{ML}} = N\overline{X}_n^p$;
3. $\frac{\partial F_n(\beta, p)}{\partial \beta} = E_{\beta, p}(N\overline{X}_n^p)$.

Now, it follows from [27] and the dominated convergence theorem, that

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

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

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

for all $n \geq N$. Let us first consider the case $p$ is odd. We then have the following by Lemma 2:

$$\limsup_{n \to \infty} n^{-1} \log[1 - F_{T_n, \beta_0}(n^{1/2}t)] = \limsup_{n \to \infty} n^{-1} \log P_{\beta_0, p_0}(\overline{X}_n^p > E_{\beta_0 + t, p}(\overline{X}_n^p))$$

$$\leq \limsup_{n \to \infty} n^{-1} \log P_{\beta_0, p_0}(\overline{X}_n^p > (m_*(\beta_0 + t, p) - \varepsilon)^p)$$

$$\leq \limsup_{n \to \infty} n^{-1} \log P_{\beta_0, p_0}(\overline{X}_n^p > m_*(\beta_0 + t, p) - \varepsilon)$$

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

Similarly, we also have:

$$\liminf_{n \to \infty} n^{-1} \log[1 - F_{T_n, \beta_0}(n^{1/2}t)] = \liminf_{n \to \infty} n^{-1} \log P_{\beta_0, p_0}(\overline{X}_n^p > E_{\beta_0 + t, p}(\overline{X}_n^p))$$

$$\geq \liminf_{n \to \infty} n^{-1} \log P_{\beta_0, p_0}(\overline{X}_n^p > (m_*(\beta_0 + t, p) + \varepsilon)^p)$$

$$\geq \liminf_{n \to \infty} n^{-1} \log P_{\beta_0, p_0}(\overline{X}_n^p > m_*(\beta_0 + t, p) + \varepsilon)$$

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

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

$$\lim_{n \to \infty} n^{-1} \log[1 - F_{T_n, \beta_0}(n^{1/2}t)] = \sup_{x > m_*(\beta_0 + t, p)} H_{\beta_0, p}(x) - \sup_{x \in [-1, 1]} H_{\beta_0, p}(x). \quad (2.3)$$

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

$$n^{-1} \log P_{\beta_0, p_0}(\overline{X}_n^p > \alpha^p) = n^{-1} \log [2P_{\beta_0, p_0}(\overline{X}_n > \alpha)]$$

$$= n^{-1} \log P_{\beta_0, p_0}(\overline{X}_n > \alpha) + o(1).$$

Hence, the same argument as for the case of odd $p$ also works here, showing that (2.3) 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 10) 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 10).

The Bahadur slope of $\hat{\beta}_{\text{ML}}$ at an alternative $\beta$ 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. □

The following result compares that Bahadur slopes and the optimal sample sizes for the tests based on the MPL and the ML estimators.

**Corollary 1.** For every $\beta > \beta_0 > \beta^*(2)$ and $\delta \in (0,1)$, we have

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

For $p \geq 3$, $\beta > \beta_0 > \beta^*(p)$ and $\delta \in (0,1)$, we have the following:

$$c_{\hat{\beta}_{\text{MPL}}} (\beta_0, \beta, p) = c_{\hat{\beta}_{\text{ML}}} (\beta_0, \beta, p) \iff N_{\hat{\beta}_{\text{MPL}}} (\beta_0, \beta, \delta, p) = N_{\hat{\beta}_{\text{ML}}} (\beta_0, \beta, \delta, p) \iff H_{\beta_0,p}(m_*(\beta,p)) \geq 0. \quad (2.4)$$
A sufficient condition for (2.4) to hold for all $\beta > \beta_0$, is $\beta_0 \geq \log 2$. On the other hand, for every $\beta^*(p) < \beta_0 < \log 2$, $H_{\beta_0,p}(m_*(\beta,p)) < 0$ for all $\beta > \beta_0$ large enough, in which case, the Bahadur slopes and optimal sample sizes for the tests based on the MPL and ML estimators do not agree. Further, for every $p \geq 3$ and every fixed $\beta > \beta^*(p)$, we have:

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

(2.5)

**Proof.** The result for $p = 2$ follows directly from Lemma 9. For $p \geq 3$, it follows from Lemma 8 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,

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

(2.6)

(2.4) now follows from Lemma 9. Now, it follows from (2.6) 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 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 10), we must have $H_{\beta_0,p}(m_*(\beta,p)) < 0$ for all $\beta > \beta_0$ large enough, which shows, in view of (2.4), that the Bahadur slopes and optimal sample sizes for the tests based on the MPL and ML estimators do not agree in this case.

Finally, towards proving (2.5), note that

$$\lim_{\beta_0 \to \beta^*(p)^+} \sup_{x \in [-1,1]} H_{\beta_0,p}(x) = 0 .$$
Figure 7. Optimal sample size for the tests based on MLE and MPLE with varying $\delta$; $p \in \{2,3,4\}, \beta_0 = 0.7, \beta = 0.71, \delta = 0.05$ (with logarithmic vertical scale).

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,

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

Corollary 1 now follows from Theorem 1 and Lemma 9. □

Remark 2.1. Note that for $p \geq 3$ and fixed $\beta \in (\beta^*(p), \infty)$, the set of all $\beta_0 > \beta^*(p)$ satisfying equality of the Bahadur slopes and the optimal sample sizes for the ML and MPL estimates, is given by:

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

where $I(x) = \frac{1}{2} \{(1 + x) \log(1 + x) + (1 - x) \log(1 - x)\}$. 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 $\bar{X}_n$ 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 $\beta$, there is a higher probability that the average magnetization $\bar{X}_n$ is small.
Remark 2.2. It follows from Lemma C.12 in \cite{27} 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 \(\sup_{x \in [-1,1]} H_{\beta_0,p}(x)\) eventually becomes 0 as \(\beta_0\) approaches \(\beta^*(p)\) from the right, the rate at which \(N_{\beta_{\text{MPL}}} (\beta_0, \beta, \delta, p)\) approaches \(\infty\) 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 \(N_{\beta_{\text{MPL}}} (\beta_0, \beta, \delta, p)\), and is given by (\(\beta_0 - \beta^*(p)\))\(^{-1}\).

Remark 2.3. It follows from the proof of Corollary 1 that for a fixed \(\beta > \beta^*(p)\),
\[
\lim_{\beta_0 \to \beta^*(p)^+} N_{\beta_{\text{MPL}}} (\beta_0, \beta, \delta, p) = \frac{\log(\delta)}{H_{\beta^*(p),p}(m_*(\beta, p))} < \infty .
\]
Another interesting phenomenon is that for fixed \(\beta_0 \in (\beta^*(p), \log 2)\), \(N_{\beta_{\text{MPL}}} (\beta_0, \beta, \delta, p)\) is a non-zero constant for \(\beta \in [\beta_0, \infty)\) for some \(\beta > 0\). This implies that as long as \(\beta_0 \in (\beta^*(p), \log 2)\), once the separation between \(\beta_0\) and \(\beta\) exceeds a certain finite value, the optimal sample size requirement for the MPL estimator does not decrease with further increase in the separation, unlike \(\hat{\beta}_{\text{ML}}\), which is an undesirable property of \(\hat{\beta}_{\text{MPL}}\). To see this, note that it follows from the proof of Corollary 1, that \(H_{\beta_0,p}(m_*(\beta, p)) < 0\) for all \(\beta\) large enough. Hence,
\[
N_{\beta_{\text{MPL}}} (\beta_0, \beta, \delta, p) = -\frac{\log(\delta)}{H_{\beta_0,p}(m_*(\beta_0,p))} > 0
\]
for all \(\beta\) large enough and \(\delta < 1\). Hence,
\[
N_{\beta_{\text{ML}}} (\beta_0, \beta, \delta, p) = -\frac{\log(\delta)}{H_{\beta_0,p}(m_*(\beta_0,p)) - H_{\beta_0,p}(m_*(\beta, p))} < N_{\beta_{\text{MPL}}} (\beta_0, \beta, \delta, p)
\]
whenever \(H_{\beta_0,p}(m_*(\beta, p)) < 0\), showing that \(N_{\beta_{\text{ML}}} (\beta_0, \beta, \delta, p)\) is a strictly decreasing function of \(\beta\) for all \(\beta\) large enough, and is strictly less than \(N_{\beta_{\text{MPL}}} (\beta_0, \beta, \delta, p)\) in this regime.

3. The Hypergraph Erdős-Rényi Ising Model

In this section, we are going to extend the results obtained for the hypergraph Curie-Weiss model to the hypergraph Erdős-Rényi Ising Model. Let \(A := \{A_{i_1 \ldots i_p}\}_{1 \leq i_1, \ldots, i_p \leq n}\) be a collection of i.i.d. Bernoulli random variables with mean \(\alpha_n\). Note that \(A\) 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:
\[
P^*_\beta,p(x) = \frac{\exp\{\beta H_n(x)\}}{2^n Z^*_n(\beta, p)} \quad (\text{for } x \in \{-1,1\}^n) ,
\]
where
\[
H_n(x) := \alpha_n^{-1} n^{1-p} \sum_{1 \leq i_1, \ldots, i_p \leq n} A_{i_1 \ldots i_p} x_{i_1} \ldots x_{i_p}
\]
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 2. The Bahadur slopes and minimum sample sizes of \(\hat{\beta}_{\text{MPL}}\) and \(\hat{\beta}_{\text{ML}}\) for the model (3.1) at an alternative \(\beta\) are respectively equal to the Bahadur slopes and minimum sample sizes of \(\hat{\beta}_{\text{MPL}}\) and \(\hat{\beta}_{\text{ML}}\) for the model (1.1) at \(\beta\).
Figure 8. p-values for different sample sizes in the 2-spin Curie-Weiss model ($\beta_0 = 0.7 > \beta^*(2)$ and $\beta = 0.9$).

In order to prove Theorem 2, we will need some preparation. The main approach, in a nutshell, is going to be an approximation of the model (3.1) in terms of the hypergraph Curie-Weiss model (1.1). The first step towards this, is to show that the Hamiltonian $H_n(x)$ of the hypergraph Erdős-Rényi Ising model is very close to that of the hypergraph Curie-Weiss model, and this is done by establishing a uniform (in $x$) concentration of $H_n(x)$ around its mean (with respect to the Erdős-Rényi measure) $\mathbb{E}H_n(x)$. Define

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

Lemma 3. Let $H_n$ denote the Hamiltonian of the $p$-tensor Erdős-Rényi Ising model. Then,

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

Lemma 3 says that as long as the Erdős-Rényi hyperedge probability $\alpha_n \gg n^{1-p}$, the Hamiltonian $H_n(x)$ concentrates around its mean (when scaled by a factor of $1/n$) uniformly in $x$. The proof of Lemma 3 is given in Appendix B. The following result is a corollary of the proof of Lemma 3, that will be useful in deriving the Bahadur slope of the MPL estimator.

Corollary 2. For each $1 \leq i \leq n$, define $m_i^{(n)}(x) := \alpha_n n^{1-p} \sum_{(i_2, \ldots, i_p) \in [n]^{p-1}} A_{i_2 \ldots i_p} x_{i_2} \ldots x_{i_p}$. Then, we have:

$$\mathbb{P}\left(\sup_{1 \leq i \leq n} \sup_{x \in \{-1,1\}^n} \left|m_i^{(n)}(x) - \mathbb{E}n^{p-1}\right| \leq 3\gamma_n \text{ for all but finitely many } n\right) = 1.$$
Figure 9. p-values for different sample sizes in the 3-tensor Curie-Weiss model ($\beta_0 = 0.68 \in (\beta^*(3), \log 2)$, $\beta = 0.9$).

**Proof.** For each $1 \leq i \leq n$, define $A_{i_{2\ldots p}}^{(i)} := A_{i_{2\ldots p}}$. Then, for each $1 \leq i \leq n$, one can view $m_i(X)$ as the Hamiltonian (scaled by $n^{-1}$) of the $(p-1)$-spin Erdős-Rényi Ising model with adjacency tensor $A^{(i)} := ((A_{i_{2\ldots p}}^{(i)}))$. The rest of the proof will follow exactly as the proof of Lemma 3. □

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. We can now use Lemma 3 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 4.** Let $\{\beta_n\}_{n \geq 1}$ be a bounded sequence of positive real numbers. Then, with probability 1, we have:

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

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

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

(3.2)
It follows from (3.2) and Lemma 3, that with probability 1, we have the following for all large \( n \),
\[
e^{-6n\beta_n \gamma_n} P_{\beta_n, p}(A|B) \leq P^\ast_{\beta_n, p}(A|B) \leq e^{6n\beta_n \gamma_n} P_{\beta_n, p}(A|B).
\] (3.3)

Lemma 4 now follows on taking logarithm on both sides of (3.2), and recalling that the sequence \( \{\beta_n\}_{n \geq 1} \) is bounded. \( \square \)

**Remark 3.1.** One can compare the logarithms of the two (unconditional) measures \( P^\ast \) (3.1) and \( P \) (1.1) by taking \( B := \{-1, 1\}^n \) in Lemma 4.

One can now use Lemmas 3 and 4 to compare the log-normalizing constants and asymptotics of the sample mean in the two models (1.1) and (3.1).

**Lemma 5.** We have the following with probability 1.

1. If \( Z_n^\ast(\beta, p) \) denotes the normalizing constant of the model (3.1),
\[
|\log Z_n(\beta, p) - \log Z_n^\ast(\beta, p)| = O(n\gamma_n).
\]

2. If \( X \) is generated from the model (3.1), then for every \( \beta > \beta^\ast(p) \) and fixed \( \epsilon > 0 \), we have:
\[
P^\ast_{\beta, p}(|X_n^p - m_\ast(\beta, p)| \geq \epsilon) \leq e^{-n\Omega(1)}. \tag{3.4}
\]

In particular, \( X_n \overset{P}{\rightarrow} m_\ast(\beta, p) \) under the model (3.1).

**Proof of 1.** Using Lemma 3, we have the following for all large \( n \), with probability 1:
\[
e^{-3\beta_n \gamma_n} Z_n(\beta, p) \leq Z_n^\ast(\beta, p) \leq e^{3\beta_n \gamma_n} Z_n(\beta, p).
\] (3.5)

Part 1 follows on taking logarithm on both sides of (3.5).

**Proof of 2.** To begin with, define:
\[
M_\epsilon := \begin{cases} 
(m_\ast(\beta, p) - \epsilon, m_\ast(\beta, p) + \epsilon) & \text{if } p \text{ is odd} \\
(m_\ast(\beta, p) - \epsilon, m_\ast(\beta, p) + \epsilon) \cup (-m_\ast(\beta, p) - \epsilon, -m_\ast(\beta, p) + \epsilon) & \text{if } p \text{ is even}
\end{cases}
\]

It follows from the arguments used in the proof of Lemma 3.1 in [27], that
\[
P_{\beta, p}(X_n \in M_\epsilon^c) = O(n^{3/2}) \exp \left\{ n \left( \sup_{t \in M_\epsilon} H_{\beta, p}(t) - H_{\beta, p}(m_\ast(\beta, p)) \right) \right\} = e^{-n\Omega(1)}. \tag{3.6}
\]

It follows from (3.6) and Lemma 4, that
\[
P^\ast_{\beta, p}(X_n \in M_\epsilon^c) \leq P_{\beta, p}(X_n \in M_\epsilon^c) e^{O(n\gamma_n)} = e^{-n\Omega(1)}. \tag{3.7}
\]

Hence, we have:
\[
P^\ast_{\beta, p}(|X_n^p - m_\ast(\beta, p)| \geq \epsilon) \leq P^\ast_{\beta, p}(X_n \in M_\epsilon^c) = e^{-n\Omega(1)}.
\]

This completes the proof of Lemma 5.

We are now ready to prove the analogous version of Lemma 1 for the model (3.1).

**Lemma 6.** Under the model (3.1), for every \( \beta > \beta^\ast(p) \), we have:
\[
n^{-1/2} T_n \overset{P}{\rightarrow} \beta - \beta_0 \quad \text{as } n \to \infty,
\]
where \( T_n \) is either \( \sqrt{n}(\hat{\beta}_{\text{MPL}} - \beta_0) \) or \( \sqrt{n}(\hat{\beta}_{\text{ML}} - \beta_0) \).
Proof. Note that $n^{-1/2} T_n = \hat{\beta} - \beta_0$, where $\hat{\beta}$ is either $\hat{\beta}_{ML}$ or $\hat{\beta}_{MPL}$. All the following arguments are on the following event, which has probability 1 in view of Lemma 3:

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

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

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

Now, by part (2) of Lemma 5 and the dominated convergence theorem, we have:

$$\mathbb{E}^*_{\hat{\beta} + t,p}(X_n^p) \to m_*(\beta + t, p)^p.$$

Hence, once again by part (2) of Lemma 5, we have:

$$\mathbb{P}_{\hat{\beta}_{ML}}(\hat{\beta}_{ML} > \beta + t) \leq \mathbb{P}_{\hat{\beta}_{ML}}(X_n^p > m_*(\beta + t, p)^p - o(1)) \leq \mathbb{P}_{\hat{\beta}_{ML}}(\mathbb{E}^*_{\hat{\beta} + t,p}(X_n^p) - m_*(\beta, p)^p > \Omega(1)) = o(1).$$

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

$$\mathbb{P}_{\hat{\beta}_{ML}}(\hat{\beta}_{ML} < \beta - t) \leq \mathbb{P}_{\hat{\beta}_{ML}}(X_n^p < m_*(\beta - t, p)^p + o(1)) \leq \mathbb{P}_{\hat{\beta}_{ML}}(\mathbb{E}^*_{\hat{\beta} + t,p}(X_n^p) - m_*(\beta, p)^p > \Omega(1)) = o(1).$$

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

Now, suppose that $\hat{\beta} = \hat{\beta}_{MPL}$. By part 1 of Lemma 5, 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 [28] and Lemma 12, we conclude that $\hat{\beta}_{MPL}$ is a consistent estimator of $\beta$ under the model $\mathbb{P}^*_{\beta,p}$. This completes the proof of Lemma 6. \qed

We are now ready to prove Theorem 2, the main result of this section.

Proof of Theorem 2. We begin with the ML estimator first. Note that by Lemma 3, we have:

$$\mathbb{P}_{\beta_0,p}(X_n^p > \mathbb{E}^*_{\beta_0 + t,p} X_n^p + 6 \gamma_n) \leq \mathbb{P}_{\beta_0,p}(H_n(X) > \mathbb{E}^*_{\beta_0 + t,p} H_n(X)) \leq \mathbb{P}_{\beta_0,p}(X_n^p > \mathbb{E}^*_{\beta_0 + t,p} X_n^p - 6 \gamma_n).$$

Now, note that by part (2) of Lemma 5 and the dominated convergence theorem, $\mathbb{E}^*_{\beta_0 + t,p} X_n^p \to m_*(\beta_0 + t, p)^p$. Hence, we have:

$$\mathbb{P}_{\beta_0,p}(X_n^p > m_*(\beta_0 + t, p)^p + o(1)) \leq \mathbb{P}_{\beta_0,p}(H_n(X) > \mathbb{E}^*_{\beta_0 + t,p} H_n(X)) \leq \mathbb{P}_{\beta_0,p}(X_n^p > m_*(\beta_0 + t, p)^p + \delta(1))$$

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

$$\mathbb{P}_{\beta_0,p}(X_n^p > (m_*(\beta_0 + t, p) + \varepsilon)^p) \leq \mathbb{P}_{\beta_0,p}(H_n(X) > \mathbb{E}^*_{\beta_0 + t,p} H_n(X)) \leq \mathbb{P}_{\beta_0,p}(X_n^p > (m_*(\beta_0 + t, p) - \varepsilon)^p).$$

It now follows from Lemma 4, that:

$$\frac{1}{n} \log \mathbb{P}_{\beta_0,p}(X_n^p > (m_*(\beta_0 + t, p) + \varepsilon)^p) + o(1) \leq \frac{1}{n} \log \mathbb{P}_{\beta_0,p}(H_n(X) > \mathbb{E}^*_{\beta_0 + t,p} H_n(X)) \leq \frac{1}{n} \log \mathbb{P}_{\beta_0,p}(X_n^p > (m_*(\beta_0 + t, p) - \varepsilon)^p) + o(1).$$

Theorem 2 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 [28], that $\hat{\beta}_{MPL}$ is the least solution of the equation (in $\beta$):

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

\[
\psi_n'(\beta) = \frac{P}{n} \sum_{i=1}^{n} m_i^2(X) \text{sech}^2(p \beta m_i(X)) > 0,
\]

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

\[
\frac{1}{n} \log P^*_{\beta_0,p}(\hat{\beta}_{\text{MLE}} > \beta_0 + t) = \frac{1}{n} \log P^*_{\beta_0,p}(\psi_n(\hat{\beta}_{\text{MLE}}) > \psi_n(\beta_0 + t))
\]

\[
= \frac{1}{n} \log P^*_{\beta_0,p} \left( \frac{1}{n} H_n(X) > \frac{1}{n} \sum_{i=1}^{n} m_i(X) \tanh(p(\beta_0 + t)m_i(X)) \right)
\]

\[
= \frac{1}{n} \log P^*_{\beta_0,p} \left( X_n^p > \frac{1}{n} \sum_{i=1}^{n} m_i(X) \tanh(p(\beta_0 + t)m_i(X)) + o(1) \right)
\]

Therefore, in view of Lemma 3, Corollary 2 and Lemma 4, we have the following for every fixed \( \varepsilon > 0 \) sufficiently small:

\[
\limsup_{n \to \infty} \frac{1}{n} \log P^*_{\beta_0,p}(\hat{\beta}_{\text{MLE}} > \beta_0 + t) 
\leq \limsup_{n \to \infty} \frac{1}{n} \log P^*_{\beta_0,p} \left( X_n^p > \frac{1}{n} \sum_{i=1}^{n} X_n^{p-1} \tanh(p(\beta_0 + t)m_i(X)) + o(1) \right)
\]

\[
= \limsup_{n \to \infty} \frac{1}{n} \log P^*_{\beta_0,p} \left( X_n^p > \frac{1}{n} \sum_{i=1}^{n} X_n^{p-1} \tanh(p(\beta_0 + t)m_i(X)) + o(1) \right)
\]

\[
= \limsup_{n \to \infty} \frac{1}{n} \log P^*_{\beta_0,p} \left( \eta_p(X_n) > \beta_0 + t - \varepsilon \right)
\]

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

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

\[
\limsup_{n \to \infty} \frac{1}{n} \log P^*_{\beta_0,p}(\hat{\beta}_{\text{MLE}} > \beta_0 + t) \leq \sup_{x \in \eta^{-1}_p((\beta_0+\varepsilon,\infty))} H_{\beta_0,p}(x) - \sup_{x \in [-1,1]} H_{\beta_0,p}(x). \tag{3.8}
\]

By an exactly similar approach, we can show that:

\[
\liminf_{n \to \infty} \frac{1}{n} \log P^*_{\beta_0,p}(\hat{\beta}_{\text{MLE}} > \beta_0 + t) \geq \sup_{x \in \eta^{-1}_p((\beta_0+\varepsilon,\infty))} H_{\beta_0,p}(x) - \sup_{x \in [-1,1]} H_{\beta_0,p}(x). \tag{3.9}
\]

Theorem 2 now follows from (3.8) and (3.9).

4. Numerical Results

In this section, we provide a graphical presentation of the numerical values of the 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. In figures 2–7, we fix the level \( \delta = 0.05 \). In Figure 2, we fix \( \beta_0 = 0.7 \), a value slightly larger than \( \log 2 \), and plot the theoretical 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 optimal sample size decreases as the alternative parameter $\beta$ increases, which is expected, since the detection capability of the tests should increase as the alternative $\beta$ moves far apart from the null $\beta_0$. Another important observation is that with increase in the interaction complexity $p$, the 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 $\beta_0$ (which is fixed at 0.7) gets closer to $\beta^*(p)$, which causes a slight increase in the difficulty of the testing problem.

In Figures 3 and 4, we fix the alternative $\beta = 0.9$, and demonstrate graphically (for the cases $p = 3$ and $4$ respectively), that the optimal sample sizes for the MPL and the ML tests differ for all $\beta_0$ in a small right neighborhood of $\beta^*(p)$ below $\log 2$, and agree above that neighborhood. The figures also demonstrate that the optimal sample size for the MPL test approaches $\infty$ as the null $\beta_0$ approaches the threshold $\beta^*(p)$. In Figures 5 and 6, we fix the null $\beta_0$ to values slightly smaller than $\log 2$, and demonstrate (for the cases $p = 3$ and $4$ respectively), that although the 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 $\beta$ large enough. All these results reflect the contents of Corollary 1. Figure 7 demonstrates the decreasing nature of the optimal sample sizes of the ML and MPL tests with increase in $\delta$, for $p = 2, 3$ and $4$, holding $\beta_0$ and $\beta$ fixed at 0.7 and 0.71, respectively.

Finally, 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 closely with the theoretical sample-size value of 679 for the MPL test that one will obtain in this setting, from Theorem 1. In this case, the theoretical 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 MPL test becomes much less efficient than the ML test for $\beta_0 < \log 2$ and sufficiently high $\beta$.

5. 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 one-parameter tensor Curie-Weiss model. One of our interesting findings is that although the MPL estimator is equally efficient as the ML estimator for $p = 2$, for $p = 3$ 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 $\beta^*(p)$ and $\log 2$, then the MPL estimator is strictly less efficient than the ML estimator for all sufficiently large values of the alternative parameter. Not only so, the Bahadur asymptotic relative efficiency of the MPL estimator with respect to the ML estimator approaches 0 for every fixed alternative above the threshold, as the null parameter approaches the threshold. We also showed a certain universality property, which demonstrates that
our results hold ditto for the more general class of tensor Erős-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 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, is 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 [15], 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.

6. Acknowledgment

S.G. was supported by the National Science Foundation BIGDATA grant IIS-1837931.

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] J. Besag, Spatial interaction and the statistical analysis of lattice systems, J. Roy. Stat. Soc. B, Vol. 36, 192–236, 1974.
[9] J. Besag, Statistical analysis of non-lattice data, The Statistician, Vol. 24, No. 3, 179–195, 1975.
[10] 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.
[11] B. B. Bhattacharya and S. Mukherjee, Inference in ising models, Bernoulli, Vol. 24, No. 1, 493–525, 2018.
[12] Guy Bresler, Efficiently learning Ising models on arbitrary graphs, Proceedings Symposium on Theory of Computing (STOC), 771–782, 2015.
[13] 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.
[14] S. Chatterjee, Estimation in spin glasses: A first step, The Annals of Statistics, Vol. 35, No. 5, 1931–1946, 2007.
[15] N. Deb and S. Mukherjee, Fluctuations in mean-field Ising models, arXiv:2005.00710, 2020.
[16] 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.
[17] P. Ghosal and S. Mukherjee, Joint estimation of parameters in Ising model, arXiv:1801.06570, 2019.
[18] M. Goemans, Chernoff bounds, and some applications, Lecture Notes: 18.310, Feb 21, 2015. http://math.mit.edu/~goemans/18310S15/chernoff-notes.pdf
[19] 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.
[20] 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.
where in the last step, we used the identity:

\[ g(x) = \frac{1}{2} \{(1 + y) \log(1 + y) + (1 - y) \log(1 - y)\} \quad \text{if} \quad y \in [-1, 1], \]

\[ g(yz) = \frac{1}{2} \{ (1 + y) \log(1 + y) + (1 - y) \log(1 - y) \} \quad \text{otherwise}. \]

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:

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

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

\[ y \tanh^{-1}(y) + \frac{1}{2} \log(1 - y^2) = \frac{1}{2} \{ (1 + y) \log(1 + y) + (1 - y) \log(1 - y) \}. \]

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

\[ \lim_{z \to \infty} e^{g(z)} = \lim_{z \to \infty} \frac{2e^{yz}}{e^z + e^{-z}} = \lim_{z \to \infty} \frac{2e^{(y-1)z}}{1 + e^{-2z}} = \begin{cases} 2 & \text{if} \quad y = 1, \\ \infty & \text{if} \quad y > 1. \end{cases} \]
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 \( y \leq -1 \), and the proof of Lemma 7. \( \square \)

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

**Lemma 8.** 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 \( p \geq 3 \), then \( H_{\beta,p} \) can have at most 3 positive stationary points. Further, there exists \( m(\beta, p) \in (0, m_*(\beta, p)) \) such that \( H_{\beta,p}' \leq 0 \) on \((0, m(\beta, p))\), \( H_{\beta,p}' \geq 0 \) on \((m(\beta, p), m_*(\beta, p))\) and \( H_{\beta,p}' \leq 0 \) on \((m_*(\beta, p), 1)\).

**Proof.** To begin with, define \( N_{\beta,p}(x) := (1 - x^2)H_{\beta,p}'(x) \). Then,
\[
N_{\beta,p}'(x) = \beta p(x - 1)x^{p-3}(p - 2 - px^2).
\]

Let us first consider the case \( p \geq 3 \). 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:
\[
m(\beta, p) := \sup\{t \in (0, 1) : H_{\beta,p}' \leq 0 \text{ on } (0, t)\}.
\]

Since \( H_{\beta,p}'(x) = \beta px^{p-1} - \tanh^{-1}(x) \) and \( \lim_{x \to 0} \tanh^{-1}(x)/x = 1 \), we have \( m(\beta, p) > 0 \). Clearly, \( H_{\beta,p}' \leq 0 \) on \((0, 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 \( m(\beta, p) < m_*(\beta, p) \). Now, by definition of \( m(\beta, p) \), there must exist a sequence \( x_n \downarrow m(\beta, p) \), such that \( H_{\beta,p}'(x_n) > 0 \) and \( x_n > m(\beta, p) \) for all \( n \). Continuity of \( H_{\beta,p}' \) now implies that \( m(\beta, p) \) is a stationary point of \( H_{\beta,p} \).

We will now show that \( H_{\beta,p}' \geq 0 \) on \((m(\beta, p), m_*(\beta, p))\]. Suppose towards a contradiction, that \( H_{\beta,p}'(y) < 0 \) for some \( y \in (m(\beta, p), m_*(\beta, p)) \). Then, there exist \( y_1 \in (m(\beta, p), y) \) 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, \( H_{\beta,p}' \geq 0 \) on \((m(\beta, p), m_*(\beta, p))\].

Finally, we show that \( H_{\beta,p}' \leq 0 \) 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, \(\tilde{H}_{\beta,p}' \leq 0\) on \((m_*(\beta,p), 1)\). This completes the proof of part (2) of Lemma 8.

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 8. \(\Box\)

The next lemma is crucial for comparing the Bahadur slopes of \(\hat{\beta}_{\text{MPL}}\) and \(\hat{\beta}_{\text{ML}}\). Recall the definition of the function \(\eta_p : [-1, 1] \mapsto \mathbb{R}\):

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

**Lemma 9.** For every \(\beta > \beta_0 > \beta^*(p)\), we have:

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

**Proof.** First, suppose that \(p\) is even. Then, \(\eta_p\) 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

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

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

Now, \(x \in \eta_p^{-1}((\beta, \infty)) \cap (0,1]\) 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 \(\frac{\beta}{p}x^{p-1} - \tanh^{-1}(x) < 0\).

If \(p \geq 3\), then by Lemma 8, this region is precisely equal to \(\{(0, m_*(\beta,p)) \cup (m_*(\beta,p), 1]\} \setminus F\) 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:

\[
\sup_{x \in \eta_p^{-1}((\beta, \infty))} H_{\beta_0,p}(x) = \max \left\{ \sup_{x \in (0,m_*(\beta,p))} H_{\beta_0,p}(x) \right\}, \quad \sup_{x \in (m_*(\beta,p), 1]} H_{\beta_0,p}(x) \right\}. \tag{A.2}
\]

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

\[
\sup_{x \in (0,m_*(\beta,p))} H_{\beta,p}(x) = 0.
\]

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

\[
\sup_{x \in (0,m_*(\beta,p))} H_{\beta_0,p}(x) = 0.
\]

Lemma 9 for \(p \geq 3\) now follows from (A.2). Now, let \(p = 2\). Then, \(\eta_p^{-1}((\beta, \infty)) \cap (0,1] = (m_*(\beta,p), 1]\).
Hence,
\[ \sup_{x \in \mathcal{H}^{-1}((\beta, \infty))} H_{\beta,p}(x) = \sup_{x > m_*(\beta,p)} H_{\beta,p}(x). \]
This completes the proof of Lemma 9.

**Lemma 10.** 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 \( \{n_k\}_{k \geq 1} \) of the positive integers. This subsequence must have a further subsequence \( \{n_{k_\ell}\}_{\ell \geq 1} \) 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 \( H_{\beta_{n_{k_\ell}}, p}(m_*(\beta_{n_{k_\ell}},p)) \geq H_{\beta_{n_{k_\ell}}, p}(x) \) for all \( x \in [0,1] \) and for all large \( \ell \), we must have \( H_{\beta,p}(m') \geq H_{\beta,p}(x) \) 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 \( \xi_p \).

Next, take any \( t \in (0,1) \), whence \( H_{\beta,p}(t) = \beta pt^{p-1} - \tanh^{-1}(t) > 0 \) for all \( \beta \) large enough. On the other hand, it follows from Lemma 8, that \( H'_{\beta,p} \leq 0 \) on \([m_*(\beta,p), 1]\). This shows that \( m_*(\beta,p) > t \) for all \( \beta \) large enough, showing that \( \lim_{\beta \to \infty} \xi_p(\beta) = 1 \).

Finally, to show that \( \xi_p \) is increasing on \((\beta^*(p), \infty)\), take \( \beta_2 > \beta_1 > \beta^*(p) \). Then, by Lemma 8, \( H'_{\beta_2,p} \leq 0 \) 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 \( \xi_p \) is strictly increasing. This completes the proof of Lemma 10.

**Appendix B. 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 3.

**B.1. Proof of Lemma 3.** To begin with, for every \( x \in \{-1,1\}^n \), let us define the set:
\[ \Lambda_n(x) := \{(i_1, \ldots, i_p) \in [n]^p : x_{i_1} \ldots x_{i_p} = 1\}. \]
Also, let \( \mathcal{L}_n(x) := \sum_{(i_1, \ldots, i_p) \in \Lambda_n(x)} A_{i_1 \ldots i_p} \). In these notations, we have:
\[ H_n(x) = \alpha_n^{-1} n^{-p} \left( 2\mathcal{L}_n(x) - \sum_{(i_1, \ldots, i_p) \in [n]^p} A_{i_1 \ldots i_p} \right). \]
For each \( \gamma > 0 \), define an event:
\[ \Omega_n(\gamma) := \left\{ \sup_{x \in \{-1,1\}^n} \left| \frac{\mathcal{L}_n(x)}{\mathbb{E}\mathcal{L}_n(x)} - 1 \right| \leq \gamma \right\}. \]
Since $\sup_{x \in \{-1, 1\}^n} |E \mathcal{L}_n(x)| \leq \alpha_n n^p$, we have the following on the event $\Omega_n(\gamma_n)$:

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

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

$$\leq 2 \gamma_n + \alpha_n^{-1} n^{-p} \sum_{(i_1, \ldots, i_p) \in [n]^p} A_{i_1 \ldots i_p} - 1. \quad \text{(B.1)}$$

It follows from Theorem 4 in [18], that

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

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

$$P(\Omega_n(\gamma_n) \text{ occurs for all but finitely many } n) = 1, \quad \text{(B.3)}$$

in order to complete the proof of Lemma 3. Towards this, note that by a union bound,

$$P(\Omega_n(\gamma)^c) \leq \sum_{x \in \{-1, 1\}^n} P(\mathcal{L}_n(x) > (1 + \gamma) E \mathcal{L}_n(x)) + \sum_{x \in \{-1, 1\}^n} P(\mathcal{L}_n(x) < (1 - \gamma) E \mathcal{L}_n(x)). \quad \text{(B.4)}$$

It follows from Theorem 1 in [22], that

$$P(\mathcal{L}_n(x) > (1 + \gamma) E \mathcal{L}_n(x)) = P\left( \frac{\mathcal{L}_n(x)}{\Lambda_n(x)} > (1 + \gamma) \alpha_n \right) \leq e^{-|\Lambda_n(x)| D((1+\gamma)\alpha_n||\alpha_n)}, \quad \text{(B.5)}$$

and

$$P(\mathcal{L}_n(x) < (1 - \gamma) E \mathcal{L}_n(x)) = P\left( \frac{\mathcal{L}_n(x)}{\Lambda_n(x)} < (1 - \gamma) \alpha_n \right) \leq e^{-|\Lambda_n(x)| D((1-\gamma)\alpha_n||\alpha_n)}, \quad \text{(B.6)}$$

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 $x_n := n^{-1} \sum_{i=1}^n x_i$ can take, for some $x \in \{-1, 1\}^n$. Combining (B.4), (B.5) and (B.6), we have by Lemma 11 and Equation (2.17) in [10],

$$P(\Omega_n(\gamma)^c) \leq \sum_{x \in \{-1, 1\}^n} \left\{ e^{-|\Lambda_n(x)| D((1+\gamma)\alpha_n||\alpha_n)} + e^{-|\Lambda_n(x)| D((1-\gamma)\alpha_n||\alpha_n)} \right\}$$

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

$$= e^{-n \left[ \frac{p-1}{2} D((1+\gamma)\alpha_n||\alpha_n) - \log 2 \right]} \left[ \frac{1}{2} + O(n^{-1}) \right] \sum_{m \in \mathcal{M}_n} e^{-\frac{n m^p}{2} D((1+\gamma)\alpha_n||\alpha_n) - nI(m)}$$
With probability 1, we have the following:

\[ \sum_{m \in \mathcal{M}_n} e^{-n^{p+1} D((1+\gamma) \alpha_n \| \alpha_n) - \log 2} \leq |\mathcal{M}_n| = n + 1. \]

(Hence, we have from (B.7) and Equation (2.30) in [10],

\[ \mathbb{P}(\Omega_n(\gamma_n)^c) \leq O(\sqrt{n}) \left( e^{-n \left[ \frac{n^{p+1}}{2} D((1+\gamma) \alpha_n \| \alpha_n) - \log 2 \right]} + e^{-n \left[ \frac{n^{p+1}}{2} D((1-\gamma) \alpha_n \| \alpha_n) - \log 2 \right]} \right) \]

\[ \leq O(\sqrt{n}) \left( \exp \left\{ -n \left[ \alpha_n n^{p-1} \frac{\gamma^2}{6} - \log 2 \right] \right\} + \exp \left\{ -n \left[ \alpha_n n^{p-1} \frac{\gamma^2}{4} - \log 2 \right] \right\} \right) \]

\[ \leq O(\sqrt{n}) \exp(-0.8n). \] (B.8)

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

We now derive the cardinality of \( \Lambda_n(x) \) that was required in the proof of Lemma 3.

**Lemma 11.** For every \( x \in \{-1, 1\}^n \), we have:

\[ |\Lambda_n(x)| = \frac{1}{2} n^p (1 + \bar{x} n^p), \]

where \( \bar{x} := n^{-1} \sum_{i=1}^{n} x_i \).

**Proof.** First, note that \( (i_1, \ldots, i_p) \in \Lambda_n(x) \) 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 \( n(1 - \bar{x})/2 \). To form an \( (i_1, \ldots, i_p) \in \Lambda_n(x) \), we must thus choose an even number of these \( p \) indices from the total number of \( n(1 - \bar{x})/2 \) possible indices where we have \(-1\), and the rest of these \( p \) indices from the remaining \( n(1 + \bar{x})/2 \) number of possible indices where we have \(+1\). We thus have:

\[ |\Lambda_n(x)| = \sum_{k \in [p] \cup \{0\}: \ k \text{ is even}} \binom{p}{k} \left( \frac{n(1 - \bar{x})}{2} \right)^k \left( \frac{n(1 + \bar{x})}{2} \right)^{p-k} \]

\[ = \frac{1}{2} \left( \frac{n(1 + \bar{x})}{2} + \frac{n(1 - \bar{x})}{2} \right)^p + \frac{1}{2} \left( \frac{n(1 + \bar{x})}{2} - \frac{n(1 - \bar{x})}{2} \right)^p. \] (B.9)

Lemma 11 follows from (B.9).

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

**Lemma 12.** With probability 1, we have the following:

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

**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 [18], we have:

\[ \mathbb{P} \left( \alpha_n^{-1} n^{1-p} \sum_{(i_2, \ldots, i_p) \in [n]^{p-1}} A_{i_1 \ldots i_p} \geq 1 + \delta \right) \leq e^{-\frac{\delta^2}{2n^{2-p}}} \alpha_n n^{p-1} \]
for every $\delta > 0$. Hence,
\[
\mathbb{P}\left( \max_{1 \leq 1_1 \leq n} \sum_{(i_2, \ldots, i_p) \in [n]^{p-1}} A_{i_1 \ldots i_p} \geq 2(1 + \delta)\alpha_n n^{p-1} \right) \leq ne^{-\frac{\beta^2}{2+\delta}\alpha_n n^{p-1}} = e^{\log n - \frac{\beta^2}{2+\delta}\alpha_n n^{p-1}}.
\]
Since $\alpha_n = \Omega(n^{1-p} \log n)$, we can choose $\delta > 0$ large enough, so that $\log n - \frac{\beta^2}{2+\delta}\alpha_n n^{p-1} \leq -2\log n$ thereby ensuring that
\[
\mathbb{P}\left( \max_{1 \leq 1_1 \leq n} \sum_{(i_2, \ldots, i_p) \in [n]^{p-1}} A_{i_1 \ldots i_p} \geq 2(1 + \delta)\alpha_n n^{p-1} \right) \leq n^{-2}.
\]
It now follows by an application of the Borel-Cantelli lemma, that:
\[
\mathbb{P}\left( \max_{1 \leq 1_1 \leq n} \sum_{(i_2, \ldots, i_p) \in [n]^{p-1}} A_{i_1 \ldots i_p} \leq 2(1 + \delta)\alpha_n n^{p-1} \text{ for all large } n \right) = 1,
\]
which completes the proof of Lemma 12. \hfill \Box

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

Department of Statistics, Stanford University raswa281@stanford.edu

Department of Statistics, Columbia University js4638@columbia.edu

Department of Statistics, University of Florida souravmukherjee@ufl.edu