Bernoulli **24**(1), 2018, 493–525
DOI: 10.3150/16-BEJ886

Inference in Ising models

BHASWAR B. BHATTACHARYA[1] and SUMIT MUKHERJEE[2]

[1]*Department of Statistics, University of Pennsylvania, Philadelphia, PA, USA.*
E-mail: *bhaswar@wharton.upenn.edu*
[2]*Department of Statistics, Columbia University, New York, NY, USA. E-mail:* *sm3949@columbia.edu*

The Ising spin glass is a one-parameter exponential family model for binary data with quadratic sufficient statistic. In this paper, we show that given a single realization from this model, the maximum pseudolikelihood estimate (MPLE) of the natural parameter is $\sqrt{a_N}$-consistent at a point whenever the log-partition function has order a_N in a neighborhood of that point. This gives consistency rates of the MPLE for ferromagnetic Ising models on general weighted graphs in all regimes, extending the results of Chatterjee (*Ann. Statist.* **35** (2007) 1931–1946) where only $\sqrt{N}$-consistency of the MPLE was shown. It is also shown that consistent testing, and hence estimation, is impossible in the high temperature phase in ferromagnetic Ising models on a converging sequence of simple graphs, which include the Curie–Weiss model. In this regime, the sufficient statistic is distributed as a weighted sum of independent χ_1^2 random variables, and the asymptotic power of the most powerful test is determined. We also illustrate applications of our results on synthetic and real-world network data.

Keywords: exponential family; graph limit theory; hypothesis testing; Ising model; pseudolikelihood estimation; spin glass

1. Introduction

The Ising spin glass is a discrete random field developed in statistical physics as a model for ferromagnetism [23], and is now widely used in statistics as a model for binary data with applications in spatial modeling, image processing, and neural networks (cf. [2,20,22] and the references therein). To describe the model, suppose that the data is a vector of dependent ± 1 random variables $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_N)$, and the dependence among the coordinates of σ is modeled by a one-parameter exponential family where the sufficient statistic is a quadratic form:

$$H_N(\tau) = \tau' J_N \tau = \sum_{1 \le i, j \le N} J_N(i, j) \tau_i \tau_j \tag{1.1}$$

for any $\tau \in S_N := \{-1, 1\}^N$ and an $N \times N$ symmetric matrix J_N with zeros on the diagonals. The elements of J_N are denoted by $J_N(i, j) = J_N(j, i)$, for $1 \le i < j \le N$. Given any $\beta \ge 0$, the quadratic form (1.1) defines a parametric family of probability distributions on S_N:

$$\mathbb{P}_\beta(\sigma = \tau) = 2^{-N} \exp\left\{\frac{1}{2}\beta H_N(\tau) - F_N(\beta)\right\}, \tag{1.2}$$

where $F_N(\beta)$ is the *log-partition function* which is determined by the condition $\sum_{\tau\in S_N}\mathbb{P}_\beta\{\sigma=\tau\}=1$, that is,

$$F_N(\beta):=\log\left\{\frac{1}{2^N}\sum_{\tau\in S_N}e^{\frac{1}{2}\beta H_N(\tau)}\right\}=\log\mathbb{E}_0 e^{\frac{1}{2}\beta H_N(\sigma)}, \tag{1.3}$$

where $\mathbb{E}_0$ denotes the expectation over σ distributed as $\mathbb{P}_0$, the uniform measure on S_N. The parameter $\beta=1/T$ is often referred to as the *inverse temperature*, so the high temperature regime corresponds to small values of β. The family (1.2) includes many famous statistical physics models: the usual ferromagnetic Ising model on generals graphs, the Sherrington–Kirkpatrick mean-field model [30,33,34], and the Hopfield model for neural networks [22].

Estimating the parameter β in (1.2), given one realization from the model, is extremely difficult using likelihood-based methods because of the presence of an intractable normalizing constant $F_N(\beta)$ in the likelihood. A variety of numerical methods are known for approximately computing the likelihood [18], but they are computationally expensive and very little is known about the rate of convergence.

One alternative to using likelihood-based methods is to consider the maximum pseudolikelihood estimator (MPLE) [4,5]. Chatterjee [10] showed that given a single spin configuration from the model (1.2), the MPLE $\hat{\beta}_N$ is $\sqrt{N}$-consistent at $\beta=\beta_0$,[1] whenever $\liminf_{N\to\infty}\frac{1}{N}F_N(\beta_0)>0$. However, in many popular models such as regular graphs, random graphs, and dense graphs, the log-partition function $F_N(\beta)=o(N)$ for certain ranges of β, and Chatterjee's result does not tell us anything about the consistency of the MPLE.

In this paper, we show that the MPLE is $\sqrt{a_N}$-consistent at $\beta=\beta_0$, if the log-partition function has order a_N in a neighborhood of β_0 (Theorem 2.1), for a sequence $a_N\to\infty$. This gives the consistency rate of the MPLE for all values of $\beta>0$ away from the critical points, and shows that the rate of the MPLE undergoes phase transitions for Ising models on various graphs ensembles (Corollaries 3.1 and 3.2). We also show that *no* consistent test, and hence no estimator, exists if the log-partition function remains bounded (Theorem 2.3). As a consequence, consistent estimation is impossible in the high temperature regime in ferromagnetic Ising models on a converging sequence (in cut-metric as defined by Lovaśz and co-authors [7,8,27]) of graphs (Theorem 3.3). This strengthens previous results of Comets and Gidas [12] and Chatterjee [10] where the MLE and the MPLE was, respectively, shown to be inconsistent for $0\le\beta<1$ in the Curie–Weiss model, which corresponds to taking $J_N(i,j)=1/N$, for all $1\le i<j\le N$. Finally, using the emerging theory of graph limits [7,8,27], the limiting distribution of the sufficient statistic $H_N(\sigma)$, and the asymptotic power of the most powerful test are derived for dense graphs in the high temperature regime (Theorem 3.4).

While proving the consistency of the MPLE, we show that the asymptotic order of the sufficient statistic $H_N(\sigma)$ is same as the order of the log-partition function for general matrices J_N; a result which appears to be new and might be of independent interest. More precisely, the sequence of random variables $\frac{1}{a_N}H_N(\sigma)$ is asymptotically tight under $\mathbb{P}_{\beta_0}$, and the limiting distribution (if any) is non-zero when the log-partition function has order a_N in a neighborhood of

[1] A sequence of estimators $\{\hat{\beta}_N\}_{N\ge1}$ is said to be a_N*-consistent* at $\beta=\beta_0$ if $a_N|\hat{\beta}_N-\beta_0|=O_P(1)$, that is, $\limsup_{K\to\infty}\limsup_{N\to\infty}\mathbb{P}_{\beta_0}(a_N|\hat{\beta}_N-\beta_0|>K)=0$.

β_0 (Lemma 5.1). Moreover, simple bounds for matrices J_N with non-negative entries provide the correct order of the log-partition function in the high temperature regime for a wide class of Ising models (Lemma 7.1).

Finally, we illustrate the usefulness of the MPLE and the applicability of our results on a real dataset: In Section 4, we study the effect of gender among friends in two Facebook friendship-networks from the Stanford Large Network Dataset (SNAP) collection.

Another active area of research is high-dimensional structure estimation in a sparse Ising model, where the goal is to consistently estimate the underlying matrix J_N, under certain structural constraints from i.i.d. samples from the model (see [1,9,32,35] and the references therein). This is in contrast with the present work, where the matrix J_N is known and we estimate the natural parameter and its error rate given a *single* realization from the model.

1.1. Organization

The rest of the paper is organized as follows: The consistency of the MPLE and general inconsistency results are described in Section 2. Applications of these results to various graph ensembles including regular graphs, random graphs, and general weighted graphs, are explained in Section 3. Theorems 2.1 and 2.3 are proved in Sections 5 and 6, respectively. The proofs of Corollaries 3.1 and 3.2 are given in Section 7. The results on converging sequence of graphs are in Section 8. The analysis of the Facebook dataset is given in Section 4.

2. Consistency of the MPLE

The maximum pseudolikelihood estimator (MPLE), introduced by Besag [4,5], can be conveniently used to approximate the joint distribution of $\sigma \sim \mathbb{P}_\beta$ that avoids calculations with the normalizing constant.

Definition 2.1. *Given a random vector $(X_1, X_2, \dots, X_N)$ whose joint distribution is parametrized by a parameter $\beta \in \mathbb{R}$, the MPLE of β is defined as*

$$\hat{\beta}_N := \arg\max \prod_{i=1}^{N} f_i(\beta, X), \tag{2.1}$$

where $f_i(\beta, X)$ is the conditional probability density of X_i given $(X_j)_{j \neq i}$.

Given $\sigma \sim \mathbb{P}_\beta$ from the model (1.2), the conditional density of σ_i, given $(\sigma_j)_{j\neq i}$ can be easily computed. To this end, given $\tau \in S_N$, define the function $L_\tau : [0, \infty) \to \mathbb{R}$ as

$$L_\tau(x) := \frac{1}{N}\sum_{i=1}^{N} m_i(\tau)\big(\tau_i - \tanh\big(x m_i(\tau)\big)\big), \tag{2.2}$$

where

$$m_i(\tau) := \sum_{j=1}^{N} J_N(i,j)\tau_j. \tag{2.3}$$

Note that $m_i(\tau)$ does not depend on τ_i since the diagonal element $J_N(i,i) = 0$. Interpreting $\tanh(\pm\infty) = \pm 1$, the function L_τ can be extended to $[0,\infty]$ by defining $L_\tau(\infty) := \frac{1}{N}\sum_{i=1}^{N}(m_i(\tau)\tau_i - |m_i(\tau)|)$. Then it is easy to verify that (see Chatterjee [10], Section 1.2) $\frac{1}{N}\frac{\partial}{\partial\beta}\sum_{i=1}^{N}\log f_i(\beta,\tau) = L_\tau(\beta)$, and the function $L_\tau(\beta)$ is a decreasing function of β. Therefore, the MPLE for β in the model (1.2) is

$$\hat{\beta}_N(\sigma) := \inf\{x \geq 0 : L_\sigma(x) = 0\}, \tag{2.4}$$

where $\sigma \sim \mathbb{P}_\beta$ is a random element from (1.2). Hereafter, we suppress the dependence on σ and denote by $\hat{\beta}_N := \hat{\beta}_N(\sigma)$ the MPLE of β.

Consistency results for the MPLE in Ising models are known in the case of lattices [11,19,21,31], complete graphs [10], and spatial point processes [24]. However, for general processes where the dependence is neither local nor mean-field, it is very difficult to prove consistency results for MPLE. In a major breakthrough, Chatterjee [10] developed a remarkable technique using exchangeable pairs and showed [10], Theorem 1.1, that the MPLE $\{\hat{\beta}_N\}_{N\geq 1}$, given a single realization $\sigma \in S_N$ from (1.2), is a $\sqrt{N}$-consistent estimator at $\beta = \beta_0 > 0$, whenever $\sup_N \|J_N\| < \infty$[2] and

$$\liminf_{N\to\infty} \frac{1}{N} F_N(\beta_0) > 0. \tag{2.5}$$

To the best of our knowledge, all results regarding MPLE $\{\hat{\beta}_N\}_{N\geq 1}$ are in the regime where it is $\sqrt{N}$-consistent. However, in many examples such as the Ising model on dense graphs, $d(N)$-regular graphs with $d(N) \to \infty$, and Erdős–Rényi graphs $G(N, p(N))$, with $\frac{\log N}{N} \ll p(N) \ll 1$, the log-partition function $F_N(\beta) = o(N)$ for certain ranges for β. In these cases, the hypothesis (2.5) is not satisfied, and Chatterjee's result is not applicable for deriving the consistency of the MPLE. The following theorem (see Section 5.2 for proof) shows that the consistency of the MPLE at a point is governed by the order of the log-partition function in a neighborhood of that point. This generalizes the result of Chatterjee [10] giving the rate of consistency of the MPLE for all values β (at all temperatures) away from the critical points.

Theorem 2.1. *Let $\sup_{N\geq 1}\|J_N\| < \infty$, and $\beta_0 > 0$ be fixed. Suppose $\{a_N\}_{N\geq 1}$ is a sequence of positive reals diverging to ∞ such that for some $\delta > 0$ we have*

$$0 < \liminf_{N\to\infty} \frac{1}{a_N} F_N(\beta_0 - \delta) \leq \limsup_{N\to\infty} \frac{1}{a_N} F_N(\beta_0 + \delta) < \infty. \tag{2.6}$$

Moreover, assume that the following conditions hold:

[2] For any $N \times N$ symmetric matrix A, denote by $\|A\| = \sup_{x\in\mathbb{R}^N} \frac{\|Ax\|_2}{\|x\|_2}$ the operator norm of A.

(a) $\limsup_{K\to\infty} \limsup_{N\to\infty} \frac{1}{a_N}\mathbb{E}_{\beta_0}(\sum_{i=1}^N |m_i(\sigma)| \cdot \mathbf{1}\{|m_i(\sigma)| > K\}) = 0$, *where* $m_i(\sigma)$ *is as defined in* (2.3).

(b) $\limsup_{N\to\infty} \frac{1}{a_N} \sum_{i,j=1}^N J_N(i,j)^2 < \infty$.

Then the MPLE $\{\hat{\beta}_N\}_{N\geq 1}$ *for the model* (1.2) *is a* $\sqrt{a_N}$*-consistent sequence of estimators for* $\beta = \beta_0$.

Conditions (a) and (b) are technical requirements arising out of the proof technique, which ensure that the main contributions come from $m_i(\sigma)$ that are small, and on average the entries in J_N are not too large compared to a_N. The proof of the result is given in Section 5.1 (with technical lemmas proved in Appendix A). The proof is organized as follows: Using the two conditions of the theorem, Lemma 5.2 shows that $\mathbb{E}_{\beta_0}(L_\sigma(\beta_0)^2) = O(a_N/N^2)$, which implies that $L_\sigma(\beta_0)$ is small with high probability. To derive the rate of consistency of the pseudo-likelihood, it thus suffices to get a lower bound of the derivative $L'_\sigma(\beta)$. Again invoking the two conditions of Theorem 2.1, in Lemma A.3 we derive a lower bound on

$$\sum_{i=1}^N m_i(\sigma)^2 \mathbf{1}\{|m_i(\sigma)| \leq K\}$$

for K fixed. This translates into the desired lower bound on the derivative $L'_\sigma(\beta)$ using which the proof of the theorem is then completed.

The conditions of the theorem are satisfied in most commonly used models (see Section 3). Moreover, the result of Chatterjee [10], Theorem 1.1, is an immediate corollary of Theorem 2.1 (refer to Section 5.3 for the proof).

Corollary 2.2 ([10], Theorem 1.1). *Let* $\sup_{N\geq 1} \|J_N\| < \infty$ *and* $\beta_0 > 0$ *be such that* (2.5) *holds. Then the sequence of estimators* $\{\hat{\beta}_N\}_{N\geq 1}$ *is* $\sqrt{N}$ *consistent for* $\beta = \beta_0$.

Remark 2.1. Condition (2.6) in the Theorem 2.1 demands the right order of the log-partition function in a small neighborhood around the point β_0. This avoids the critical points, where the order of the log-partition function (and its derivative) undergoes a sharp transition. It follows from the proof of Theorem 2.1 that the following (possibly slightly weaker) condition works as well instead of (2.6):

$$0 < \lim_{\delta\to 0} \liminf_{N\to\infty} \frac{1}{a_N} F'_N(\beta_0 - \delta) \leq \limsup_{N\to\infty} \frac{1}{a_N} F'_N(\beta_0) < \infty.$$

However, for most of the applications estimates of the log-partition function are more readily available. Thus, the sufficient conditions are stated in terms of the log-partition function instead of its derivative.

Note that Theorem 2.1 does not apply to the case $F_N(\beta_0) = O(1)$. Next, we show that if $F_N(\beta_0) = O(1)$, then there is no sequence of estimators which consistently estimates β_0. In fact, we show that even testing is impossible in this regime: Given a single spin-configuration $\sigma \in S_N$ from (1.2), there exists *no* sequence of consistent tests[3] for the hypothesis testing problem:

$$H_0 : \beta = \beta_1 \quad \text{versus} \quad H_1 : \beta = \beta_2. \tag{2.7}$$

This is summarized in the following theorem (see Section 6 for proof):

Theorem 2.3. *Let* $\sup_{N\geq 1} \|J_N\| < \infty$, *and* $\beta_0 > 0$ *be fixed. Suppose*

$$\limsup_{N\to\infty} F_N(\beta_0) < \infty. \tag{2.8}$$

Then for $0 \leq \beta_1 < \beta_2 \leq \beta_0$, *there exists no consistent sequence of tests for the testing problem* (2.7). *In particular, there exists no consistent sequence of estimators for* β *in the interval* $[0, \beta_0]$.

One of the main applications of above results is in deriving the rate of the MPLE for Ising models on weighted graphs, that is, for matrices J_N with non-negative entries. For such matrices, condition (b) in Theorem 2.1 can be directly verified, and we have the following simplified corollary:

Corollary 2.4. *Consider the model* (1.2) *such that* J_N *is a sequence of matrices with non-negative entries with* $\lim_{N\to\infty} \|J_N\| = \lambda > 0$.

(a) *The sequence of estimators* $\{\hat{\beta}_N\}_{N\geq 1}$ *is* $\|J_N\|_F := \sqrt{\sum_{i,j=1}^N J_N(i,j)^2}$ *consistent at* $\beta = \beta_0$ *for any* $\beta_0 < \frac{1}{\lambda}$, *whenever condition* (a) *in Theorem* 2.1 *holds*.
(b) *If* $\limsup_{N\to\infty} \sum_{i,j=1}^N J_N(i,j)^2 < \infty$, *then exists no consistent sequence of estimators for* β *in the interval* $[0, \frac{1}{\lambda})$.

3. Applications

The $\sqrt{N}$-consistency of the MPLE in the Sherrington–Kirkpatrick (SK) model and the Hopfield model, for all values of $\beta > 0$, follows from results of Chatterjee [10]. Our results give the rate of consistency of the MPLE in the regime where it is not $\sqrt{N}$-consistent.

We begin with a simple example where the rate of the MPLE undergoes multiple phase transitions.

[3] A sequence of test functions $\phi_N : S_N \to \{0, 1\}$ is said to be *consistent* for the testing problem (2.7) if $\lim_{N\to\infty} \mathbb{E}_{\beta_1}\phi_N = 0$ and $\lim_{N\to\infty} \mathbb{E}_{\beta_2}\phi_N = 1$.

Example 1. *Consider the model* (1.2) *with*

$$J_N(i,j)=\begin{cases}\dfrac{1}{N}, & \text{if } 1\le i\neq j\le \dfrac{N}{2},\\ \dfrac{1}{\sqrt{N}}, & \text{if } \dfrac{N}{2}<i\neq j\le \dfrac{N}{2}+\sqrt{N},\\ 0, & \text{otherwise}.\end{cases}$$

Then the sequence of estimators $\{\hat\beta_N\}_{N\ge 1}$ *is inconsistent for* $\beta\in(0,1)$, $N^{1/4}$*-consistent for* $\beta\in(1,2)$, *and* $\sqrt{N}$*-consistent if* $\beta>2$.

The proof of the above example is given in Section 7.2. In fact, this example can be easily generalized to construct a K-block matrix J_N such that the consistency rate of MPLE undergoes K phase transitions. However, for most popular choices of J_N the rate of the MPLE undergoes at most one phase transition from $\|J_N\|_F$-consistent to $\sqrt{N}$-consistent.

3.1. Ising model on regular graphs

Let G_N be a sequence of d_N regular graphs. Consider the family of probability distributions (1.2) with the sufficient statistic

$$H_N(\tau)=\frac{1}{d_N}\tau' A(G_N)\tau, \tag{3.1}$$

where $A(G_N)=((a_N(i,j)))$ is the adjacency matrix of the graph G_N. This includes Ising models on lattices, complete graph, hypercube, and random regular graphs, among others, and have been extensively studied in probability and statistical physics. Dembo et al. [14,15] derived the limit of the log-partition function for random regular (and other locally-tree like) graphs. Levin *et al.* [26] showed that the mixing time of the Glauber dynamics on the complete graph exhibits the cutoff phenomenon [16] in the high temperature regime. The cutoff phenomenon for lattices was established by Lubetzky and Sly in a series of breakthrough papers (refer to [28,29] and the references therein).

The next result gives the rate of consistency of the MPLE for general regular graphs. The proofs are deferred to Section 7.

Corollary 3.1. *Fix* $\beta_0>0$ *and let* G_N *be a sequence of* d_N *regular graphs. Suppose* $\{\hat\beta_N\}_{N\ge1}$ *is the MPLE for the model* (1.2) *with sufficient statistic* (3.1).

(a) *If* $0<\beta_0<1$, $\{\hat\beta_N\}_{N\ge1}$ *is a* $\sqrt{N/d_N}$*-consistent sequence of estimators for* β_0.
(b) *If* $\beta_0>1$, $\{\hat\beta_N\}_{N\ge1}$ *is a* $\sqrt{N}$*-consistent sequence of estimators for* β_0.

The above theorem shows that the rate of the MPLE undergoes a phase transition at $\beta=1$ for general regular graphs. In particular if $d_N=d=O(1)$ remains bounded, then the above theorem shows that the MPLE is $\sqrt{N}$ for all non-negative $\beta\neq1$. However, in this case, it is easy to argue that $\liminf_{N\to\infty}\frac{1}{N}F_N(\beta)>0$, for all $\beta>0$ (see proof of lower bound in Corollary 3.1).

Theorem 2.1 then concludes that $\hat{\beta}_N$ is $\sqrt{N}$-consistent for all values of $\beta > 0$. In fact, using similar arguments as in the proof of Corollary 3.1, it follows that the MPLE $\{\hat{\beta}_N\}_{N\geq 1}$ is $\sqrt{N}$-consistent for all $\beta > 0$ in all bounded degree graphs with at least $O(N)$ edges. This shows that MPLE is $\sqrt{N}$-consistent for lattice graphs re-deriving classical results (see [21] and the references therein).

For $d_N \to \infty$, the behavior of the MPLE at $\beta = 1$ remains unclear. It is believed that the MPLE might have a non-Gaussian limiting distribution at the critical point $\beta = 1$ [10].

Remark 3.1. If $d_N = \Theta(N)$,[4] then Theorem 3.1 shows that the MPLE is $O(1)$ consistent for $0 < \beta_0 < 1$, suggesting that the MPLE might be inconsistent in this regime. Chatterjee [10] showed that this is indeed the case for the Curie–Weiss model (where $J_N(i,j) = 1/N$ for all $i \neq j$) for $0 \leq \beta < 1$. Comets and Gidas [12] showed that even the MLE of β in the Curie–Weiss model is inconsistent for $0 \leq \beta < 1$. Later, in Theorem 3.4 we strengthen this result by showing that for Ising models on arbitrary dense graphs, there exists no sequence of consistent estimators before the phase transition point. This extends the results in [10,12] and justifies the $O(1)$-rate of the MPLE in the dense case.

3.2. Ising model on Erdős–Rényi graphs

Let $G_N \sim \mathcal{G}(N, p(N))$ be a sequence of Erdős–Rényi graphs. Consider the family of probability distributions (1.2) with the sufficient statistic

$$H_N(\tau) = \frac{1}{Np(N)} \tau' A(G_N) \tau, \tag{3.2}$$

where $A(G_N) = ((a_N(i,j)))$ is the adjacency matrix of the graph G_N.

Corollary 3.2. *Fix $\beta_0 > 0$ and consider a sequence $G_N \sim \mathcal{G}(N, p(N))$ of Erdős–Rényi graphs, with $\frac{\log N}{N} \ll p(N) \leq 1$. Let $\{\hat{\beta}_N\}_{N\geq 1}$ be the MPLE for the model (1.2) with sufficient statistic (3.2).*

(a) *If $0 < \beta_0 < 1$, $\{\hat{\beta}_N\}_{N\geq 1}$ is a $\sqrt{1/p(N)}$-consistent sequence of estimators for β_0.*
(b) *If $\beta_0 > 1$, $\{\hat{\beta}_N\}_{N\geq 1}$ is a $\sqrt{N}$-consistent sequence of estimators for β_0.*

As in the regular case, the rate of the MPLE undergoes a phase transition at $\beta = 1$ for Erdős–Rényi graphs. Figure 1 shows the error bars for the MPLE for the Ising model on $G_N \sim \mathcal{G}(N, p(N))$, with $N = 2000$ and $p(N) = N^{-\frac{1}{3}}$, for a sequence of values of $\beta \in [0, 2]$.

3.3. Ising model on dense graphs

Recall that the MPLE is inconsistent in the Curie–Weiss model in the high temperature regime, $0 \leq \beta < 1$ [10]. In this section, using the emerging theory of graph limits and Theorem 2.3

[4]Given non-negative sequences $\{a_N\}_{N\geq 1}$ and $\{b_N\}_{N\geq 1}$, the notation $a_N = \Theta(b_N)$ means that there exist constants $k_1, k_2 > 0$, such that $k_1 b_N \leq a_N \leq k_2 b_N$, for all N large enough.

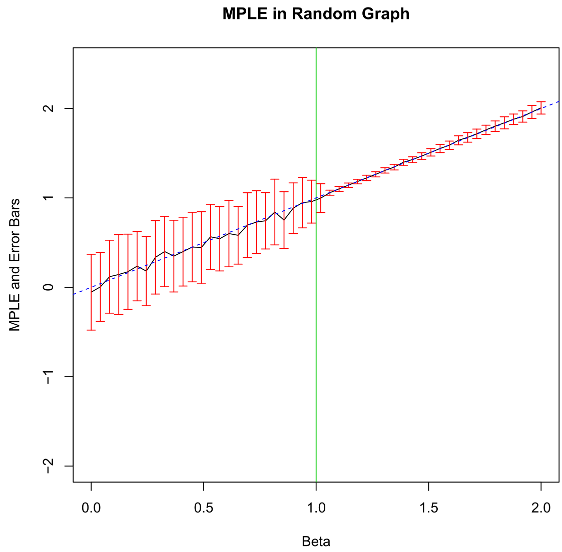

Figure 1. The MPLE and the 1-standard deviation error bar in an Ising model on $G_N \sim \mathcal{G}(N, p(N))$ with $N = 2000$ and $p(N) = N^{-\frac{1}{3}}$ averaged over 100 repetitions for a sequence of values of $\beta \in [0, 2]$. Lengths of the error bars undergo a phase transition at $\beta = 1$, as predicted by Corollary 3.2 which shows that for $0 \le \beta < 1$ the MPLE is $N^{\frac{1}{6}}$ consistent, and for $\beta > 1$, the MPLE is $\sqrt{N}$-consistent.

above, we strengthen this result to show that consistent testing is impossible in the entire high temperature regime in Ising models on a converging sequence of dense graphs. We also calculate the distribution of the most powerful test and the asymptotic power in this regime.

3.3.1. *Graph limit theory*

Let G_N be a simple graph with vertices $V(G_N) = \{1, 2, \dots, N\}$ and adjacency matrix $A(G_N)$. Lovász and co-authors [7,8] developed a limit theory of graphs, which connects various topics such as graph homomorphisms, Szemerédi regularity lemma, and extremal graph theory. In the following, we summarize the basic results for converging sequence of graphs (cf. Lovász [27] for a detailed exposition). To this end, note that any graph G_N can be represented as a function $W_{G_N} : [0, 1]^2 \to [0, 1]$ in a natural way: Define $W_{G_N}(x, y) := 1$ if and only if $(\lceil nx \rceil, \lceil ny \rceil)$ is an edge in G_N, that is, partition $[0, 1]^2$ into N^2 squares of side length $1/N$, and define $W_{G_N}(x, y) = 1$, when (x, y) is in the (a, b)th square and (a, b) is an edge in G_N. Let $\mathscr{W}$ be the space of all measurable functions from $[0, 1]^2$ into $[0, 1]$ that satisfy $W(x, y) = W(y, x)$ for all $x, y \in [0, 1]$. For every $W \in \mathscr{W}$ and any fixed simple graph $H = (V(H), E(H))$ define the *homomorphism density*

$$t(H, W) = \int_{[0,1]^{|V(H)|}} \prod_{(i,j)\in E(H)} W(x_i, x_j)\,\mathrm{d}x_1\,\mathrm{d}x_2 \cdots \mathrm{d}x_{|V(H)|}.$$

A sequence of simple graphs $\{G_N\}_{N\geq 1}$ is said to converge to $W \in \mathscr{W}$ if for every finite simple graph H,

$$\lim_{N\to\infty} t(H, G_N) = t(H, W). \tag{3.3}$$

The limit objects, that is, the elements of $\mathscr{W}$, are called *graph limits* or *graphons*. Conversely, every such function arises as the limit of an appropriate graph sequence.

It turns out that the above notion of convergence can be suitably metrized using the so-called *cut-metric* (cf. [27], Chapter 8, for details). Moreover, every function $W \in \mathscr{W}$ defines an operator $T_W : L_2[0,1] \to L_2[0,1]$:

$$(T_W f)(x) = \int_0^1 W(x, y) f(y)\,\mathrm{d}y. \tag{3.4}$$

T_W is a Hilbert–Schmidt operator with operator norm denoted by $\|W\|$, which is compact and has a discrete spectrum, that is, a countable multiset of non-zero real eigenvalues $\{\lambda_i(W)\}_{i\in\mathbb{N}}$. In particular, every non-zero eigenvalue has finite multiplicity and

$$\sum_{i=1}^{\infty} \lambda_i^2(W) = \int_{[0,1]^2} W(x, y)^2\,\mathrm{d}x\,\mathrm{d}y := \|W\|_2^2. \tag{3.5}$$

3.3.2. *Consistency and asymptotic power*

Recall that for a graph G_N, $A(G_N)$ is the adjacency matrix of G_N. Now, using graph limit theory we show the following result:

Theorem 3.3. *Let $\{G_N\}_{N\geq 1}$ be a sequence of simple graphs which converges in cut-metric to $W \in \mathscr{W}$ such that $\int_{[0,1]^2} W(x,y)\,\mathrm{d}x\,\mathrm{d}y > 0$. Consider the testing problem (2.7) given a single realization $\sigma \in S_N$ from (1.2) with sufficient statistic $H_N(\tau) = \frac{1}{N}\tau' A(G_N)\tau$.*

(a) *If $0 \leq \beta_1 < \beta_2 < \frac{1}{\|W\|}$, then there does not exist a sequence of consistent tests for (2.7).*
(b) *If $\beta_0 > \frac{1}{\|W\|}$, then the MPLE $\{\hat{\beta}_N\}_{N\geq 1}$ is a sequence of $\sqrt{N}$-consistent estimators for $\beta = \beta_0$.*

The proof of the theorem is given in Section 8. It involves showing that $F_N(\beta_0) = O(1)$ whenever $0 \leq \beta_0 < \frac{1}{\|W\|}$, for any converging sequence of graphs, which together with Theorem 2.3 proves (a). To show (b) it suffices to show that $\lim_{N\to\infty} \frac{1}{N} F_N(\beta_0) > 0$, for $\beta_0 > \frac{1}{\|W\|}$ (by Corollary 2.2). For Ising models on a convergence sequence of graphs, $\lim_{N\to\infty} \frac{1}{N} F_N(\beta_0)$ is given by a variational problem (8.1) (cf. [8], Theorem 2.14). Even though explicitly solving this variational problem for large values of β is extremely difficult, a simple argument can be used to show that the value of the variational problem is positive for $\beta > \frac{1}{\|W\|}$.

By the Neyman–Pearson lemma, the most-powerful (MP) test for (2.7) is based on the sufficient statistic $H_N(\sigma)$. By Theorem 3.3, the test based on $H_N(\sigma)$ is not consistent (see Figure 2). However, the asymptotic power of the MP-test can be derived from the limiting distribution of $H_N(\sigma)$, for any $\beta < \frac{1}{\|W\|}$.

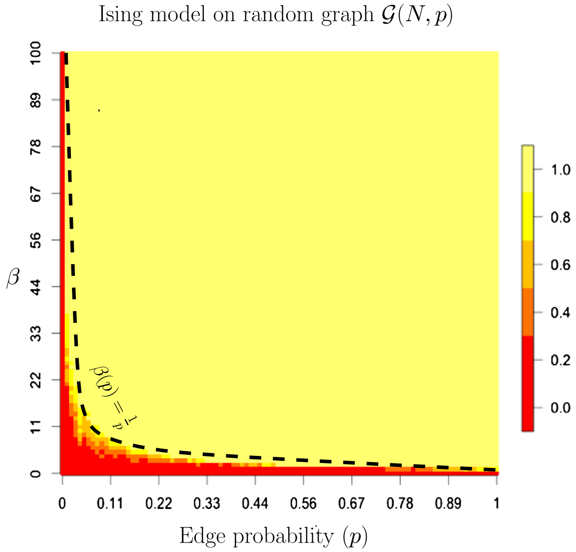

Figure 2. The power of the MP-test for the Ising model on an Erdős–Rényi random graph $\mathcal{G}(N, p)$ as a function of p and β, with $N = 500$. Every point (p, β) in the grid shows the empirical power of the MP-test averaged over 100 repetitions. Note the phase transition curve $\beta(p) = \frac{1}{p}$ above which the MP-test has power 1, as predicted by Theorem 3.4.

Theorem 3.4. *Let $\{G_N\}_{N\geq 1}$ be a sequence of simple graphs which converges in cut-metric to $W \in \mathscr{W}$, with $\int_{[0,1]^2} W(x, y)\, \mathrm{d}x\, \mathrm{d}y > 0$. If $\sigma \sim \mathbb{P}_\beta$, then for $\beta < \frac{1}{\|W\|}$*

$$H_N(\sigma) = \frac{1}{N}\sigma' A(G_N)\sigma \xrightarrow{\mathscr{D}} \sum_{i=1}^{\infty} \lambda_i(W)\left(\frac{1}{1 - \beta\lambda_i(W)}\xi_i - 1\right), \tag{3.6}$$

where $\xi_1, \xi_2, \dots,$ are i.i.d. χ_1^2 random variables.

Hereafter, the random variable in the RHS of (3.6) will be denoted by $Q_{\beta,W}$ and can be used to compute the asymptotic power for the test based on $H_N(\sigma)$ for the testing problem (2.7), when $0 \leq \beta_1 < \beta_2 < \frac{1}{\|W\|}$. To this end, we need the following definition:

Definition 3.1. *Let $W \in \mathscr{W}$ and $\beta < \frac{1}{\|W\|}$. Denote by $F_{\beta,W}$ the distribution function of the random variable $Q_{\beta,W}$ defined in (3.6). Also, let $q_{1-\alpha,\beta,W}$ be the $(1-\alpha)$th quantile of $F_{\beta,W}$, that is, $\mathbb{P}_\beta(Q_{\beta,W} \geq q_{1-\alpha,\beta,W}) = \alpha$.*

The following corollary is an immediate consequence of the Neyman–Pearson lemma and Theorem 3.4.

Corollary 3.5. *Fix $\alpha \in (0,1)$ and $0 \le \beta_1 < \beta_2 < \frac{1}{\|W\|}$. The most powerful level α test for* (2.7) *rejects H_0 when $H_N(\sigma) > q_{1-\alpha,\beta_1,W}$, and has limiting power*

$$\lim_{N\to\infty} \mathbb{P}_{\beta_2}\bigl(H_N(\sigma) > q_{1-\alpha,\beta_1,W}\bigr) = 1 - Q_{\beta_2,W}(q_{1-\alpha,\beta_1,W}). \tag{3.7}$$

In most of the relevant examples, the limiting graphon W has finitely many non-zero eigenvalues, and the expression on the RHS of (3.7) can be computed easily in terms of the quantiles of the chi-squared distribution.

Example 2. *Suppose $G_N \sim \mathcal{G}(N,p)$ be a Erdős–Rényi random graph with $0 < p \le 1$. Then G_N converges to the constant function $W_p :\equiv p$ on $[0,1]^2$, which has only one non-zero eigenvalue $\lambda_1(W_p) = p$. Therefore, consistent testing is impossible for $0 \le \beta < \frac{1}{p}$ (see Figure 2). Moreover, for $\beta < 1/p$,* (3.7) *simplifies to*

$$H_N(\sigma) \xrightarrow{\mathscr{D}} Q_{\beta,W_p} = p\left(\frac{1}{1-\beta p}\chi_1^2 - 1\right).$$

If $q_{1-\alpha}$ denotes the $(1-\alpha)$th quantile of the χ_1^2 distribution, then by (3.7), *the limiting power of the test with rejection region $\{H_N(\sigma) > c_\alpha := p(q_{1-\alpha} - 1)\}$ for the testing problem $\beta = 0$ versus $\beta = \beta_0 < 1/p$ is*

$$\lim_{N\to\infty} \mathbb{P}_{\beta_0}\bigl(H_N(\sigma) > c_\alpha\bigr) = \mathbb{P}\bigl(\chi_1^2 > (1-\beta_0 p)q_{1-\alpha}\bigr). \tag{3.8}$$

The limiting power of the MP-test for the Curie–Weiss model (*which corresponds to taking $p = 1$ in* (3.8)) *is shown in Figure 3. Note that it has a phase transition at $\beta = 1$, as stated in Theorem 3.3.*

Remark 3.2. Note that throughout the paper, the term *phase transition* has been is used to imply a change in the rate of consistency of the pseudo-likelihood estimate $\hat{\beta}_N$. Interestingly, in all our examples (Corollaries 3.1, 3.2 and Theorem 3.3) the change in the rate of consistency happens exactly at the point of thermodynamic phase transition, that is, prior to this phase transition point the log-partition function is $o(N)$, whereas after the phase-transition point the log-partition function scales linearly with N. In fact, in the setting of Corollary 3.1, the limiting log-partition function is continuous but not differentiable at the phase transition point $\beta = 1$ (see [3], Theorem 2.2(b)). Similar statements about the non-differentiability of the limiting log-partition function should also hold for the other two examples, but since they are not directly used in our calculations, this direction has not been pursued.

4. Analysis of the facebook dataset

Ising models have been widely used to understand correlations among neighboring vertices in network data with binary node attributes. Here, we use it to study the effect of gender in Facebook friendship-networks using data from the Stanford Large Network Dataset (SNAP) collection, available freely at http://snap.stanford.edu/data/egonets-Facebook.html. The nodes are

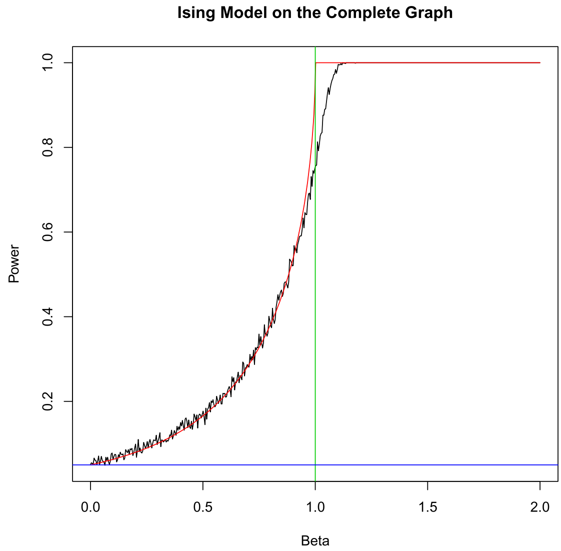

Figure 3. The power of the MP-test in the Curie–Weiss model as a function of β; the black curve is the empirical power for the Curie–Weiss model with $N = 500$ and 1000 repetitions at each point along a sequence of values (of length 500) of $\beta \in [0, 2]$. The red curve is the limiting power function (corresponds to taking $p = 1$ in (3.8)) as a function of $\beta \in [0, 2]$. The blue line corresponds to the level $\alpha = 0.05$ of the test.

groups of users from Facebook and there is an edge between two users if they are friends. The dataset also include several anonymized node features, such as hometown, gender, birthday, school, and university. We consider two networks (referred to as FB1 and FB2) with gender as the binary node feature, encoding, without loss of generality, male by 1 and female by -1. The nodes labelled 1 are colored blue and those labelled -1 are colored red. The FB1 network has 221 nodes and 3176 edges. Among the 221 nodes, 170 are labelled 1 and 51 are labelled -1. The FB2 network has 333 nodes and 2519 edges, with 213 nodes labelled 1 and 120 labelled -1.

In order to understand how gender correlates with friendship, we fit Ising models on the two networks. The MPLE for β corresponding to the two networks are given in the table in Figure 4. This can be used to test the null hypothesis that gender does not correlate with friendship. The p-values show that the null hypothesis is rejected at the 5% level in both cases, suggesting, as expected, significant correlation in the friendship-network based on gender. The MPLE in FB1 is larger, which suggests a stronger gender-based correlation in FB1, which might be due to the larger male-to-female ratio in FB1 than in FB2.

Figure 4 also shows the error bars for the MPLE calculated using parametric bootstrap: 10^5 realizations of the Ising model were resampled using the original MPLE, which then gives an estimate of the standard error of the MPLE. Note that the error bar for FB1 is slightly longer than that for FB2. This might be because the FB1 network, with average degree 28.74, is significantly dense than FB2, which has average degree 15.13.

	FB1	FB2
(Vertices, Edges)	(221, 3176)	(333, 2519)
Average Degree	28.74	15.13
MPLE	1.0518	0.8530
p-value	0.0045	0.0001

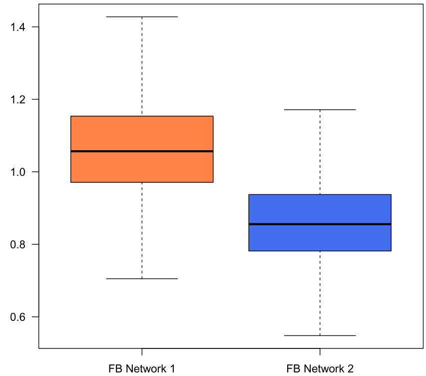

Figure 4. Facebook friendship-network: The table gives the MPLE of β for Ising models on two Facebook friendship-networks and the corresponding p-values for testing independence. The plot shows the empirical (resampled) error-bars for the MPLE in the two networks.

5. Proof of consistency of the MPLE

This section contains the proof of Theorem 2.1. The technical lemmas required for the proof are listed in Section 5.1 and proved later in Appendix A.1. Using this, we complete the proof of the theorem in Section 5.2. Corollary 2.2 is proved in Section 5.3.

5.1. Technical lemmas

The proof of Theorem 2.1 requires a few technical lemmas. We begin by showing that in Ising models satisfying (2.6), the asymptotic order of the sufficient statistic $H_N(\sigma)$ is the same as the order of the log-partition function, that is, (a) the sequence $\frac{1}{a_N}H_N(\sigma)$ does not tend to 0 in distribution, and (b) $\frac{1}{a_N}H_N(\sigma)$ is $O_P(1)$. In fact, (b) is not required in the rest of the proof, however we include it because, together with (a), it gives the correct order of $H_N(\sigma)$, which appears to be new and might be of independent interest. The proof of the lemma is given in Appendix A.1.

Lemma 5.1. *Under assumption* (2.6), *the following hold*:

(a) $\lim_{\varepsilon\to 0}\limsup_{N\to\infty}\mathbb{P}_{\beta_0}(H_N(\sigma)<\varepsilon a_N)=0$,
(b) $\lim_{K\to\infty}\limsup_{N\to\infty}\mathbb{P}_{\beta_0}(H_N(\sigma)>K a_N)=0$.

The next lemma is similar to the lemma in [10], Lemma 1.2, where it was shown that the second moment of the function $L_\sigma(\beta_0)$ is $O(1/N)$ whenever the log-partition function scales like N. Here, by a finer analysis using part (a) of Lemma 5.1 we show that the $\mathbb{E}_{\beta_0}(L_\sigma(\beta_0)^2)=O(a_N/N^2)$, if the log-partition function has order a_N. The proof of the lemma is given in Appendix A.2.

Lemma 5.2. *Let L_σ be as defined in* (2.2). *Then under the assumptions in Theorem* 2.1, *for N large enough*,

$$\limsup_{N\to\infty} \frac{N^2}{a_N} \mathbb{E}_{\beta_0}\big(L_\sigma(\beta_0)^2\big) < \infty.$$

Lemma 5.3. *Under the assumptions in Theorem* 2.1,

$$\lim_{\varepsilon\to 0} \lim_{K\to\infty} \lim_{N\to\infty} \mathbb{P}_{\beta_0}\left(\sum_{i=1}^N m_i(\sigma)^2 \mathbf{1}\big\{|m_i(\sigma)| \le K\big\} \le \varepsilon a_N\right) = 0.$$

The above lemma replaces the application of Paley–Zygmund inequality of [10], Lemma 2.2, and will be used to complete the proof of Theorem 2.1.

5.2. Completing the proof of Theorem 2.1

By Chebyshev's inequality and Lemma 5.2 there exists $C < \infty$ such that

$$\mathbb{P}_{\beta_0}\big(|L_\sigma(\beta_0)| > K_1\sqrt{a_N}/N\big) \le \frac{N^2}{a_N} \mathbb{E}_{\beta_0} L_\sigma(\beta_0)^2 \le \frac{C}{K_1^2}. \tag{5.1}$$

Now, fix $\delta > 0$. Therefore, it is possible to choose $K_1 = K_1(\delta)$ such that the RHS above is less than δ.

Also, by Lemma 5.3 there exists $\varepsilon := \varepsilon(\delta) > 0$ and $K_2 = K_2(\varepsilon, \delta) < \infty$ such that

$$\mathbb{P}_{\beta_0}\left(\sum_{i=1}^N m_i(\sigma)^2 \mathbf{1}\big\{|m_i(\sigma)| \le K_2\big\} \ge \varepsilon a_N\right) \ge 1 - \delta, \tag{5.2}$$

for N large enough. Thus, taking N large enough and setting

$$T_N(\beta_0) := \left\{\sigma \in S_N : |L_\sigma(\beta_0)| \le K_1 \frac{\sqrt{a_N}}{N}, \sum_{i=1}^N m_i(\sigma)^2 \mathbf{1}\big\{|m_i(\sigma)| \le K_2\big\} \ge \varepsilon a_N\right\},$$

we have $\mathbb{P}_{\beta_0}(T_N) \ge 1 - \delta$. For $\sigma \in T_N$,

$$\begin{aligned} |L'_\sigma(\beta_0)| &= \left|\frac{\partial}{\partial\beta} L_\sigma(\beta)\right|_{\beta=\beta_0} = \frac{1}{N}\sum_{i=1}^N m_i(\sigma)^2 \operatorname{sech}^2\big(\beta_0 m_i(\sigma)\big) \\ &\ge \frac{1}{N} \operatorname{sech}^2(\beta_0 K_2) \sum_{i=1}^N m_i(\sigma)^2 \mathbf{1}\big\{|m_i(\sigma)| \le K_2\big\} \\ &\ge \varepsilon \frac{a_N}{N} \operatorname{sech}^2(\beta_0 K_2). \end{aligned} \tag{5.3}$$

Therefore,

$$
\begin{aligned}
K_1\frac{\sqrt{a_N}}{N} &\geq \left|L_\sigma(\beta_0)\right| = \left|L_\sigma(\beta_0) - L_\sigma\left(\hat{\beta}_N(\sigma)\right)\right| \\
&\geq \int_{\beta_0\wedge\hat{\beta}_N(\sigma)}^{\beta_0\vee\hat{\beta}_N(\sigma)} L'_\sigma(\beta)\,\mathrm{d}\beta \\
&\geq \frac{\varepsilon a_N}{K_2 N}\left|\tanh\left(K_2\hat{\beta}_N(\sigma)\right) - \tanh(K_2\beta_0)\right|.
\end{aligned}
$$

Let $R = R(\delta) := \frac{K_2}{K_1\varepsilon}$. This implies that

$$
\mathbb{P}_{\beta_0}\left(\sqrt{a_N}\left|\tanh(K_2\hat{\beta}_N) - \tanh(K_2\beta_0)\right| \geq R\right) \leq \delta, \tag{5.4}
$$

and Theorem 2.1 follows.

5.3. Proof of Corollary 2.2

Note that for $\tau \in S_N$ and any $K > 0$,

$$
\sum_{i=1}^N \left|m_i(\tau)\right| \mathbf{1}\left\{\left|m_i(\tau)\right| > K\right\} \leq \frac{1}{K}\sum_{i=1}^N m_i(\tau)^2 = \frac{1}{K}\tau' J_N^2 \tau \leq \frac{N\|J_N\|^2}{K}.
$$

Therefore, condition (a) in Theorem 2.1 holds with $a_N = N$.

Moreover,

$$
\sum_{i,j=1}^N J_N^2(i,j) = \sum_{i=1}^N \|J_N \underline{e}_i\|_2^2 \leq \|J_N\|^2 \sum_{i=1}^N \|\underline{e}_i\|_2^2 = N\|J_N\|^2,
$$

that is condition (b) in Theorem 2.1 holds with $a_N = N$.

Finally, to check (2.6) note that $F'_N(\beta) = \frac{1}{2}\mathbb{E}_\beta \sigma' J_N \sigma \leq \frac{M}{2}N$, where $M := \|J_N\| < \infty$. Therefore,

$$
\lim_{\delta\to 0}\liminf_{N\to\infty}\frac{1}{N}F_N(\beta_0 - \delta) \geq \lim_{\delta\to 0}\liminf_{N\to\infty}\left(\frac{1}{N}F_N(\beta_0) - \frac{M}{2}\delta\right) > 0,
$$

by condition (2.5). Also, $\lim_{\delta\to 0}\limsup_{N\to\infty}\frac{1}{N}F_N(\beta_0 + \delta) \leq M\lim_{\delta\to 0}(\beta_0 + \delta) < \infty$. This verifies (2.6) and by an application of Theorem 2.1 the result follows.

6. Proof of Theorem 2.3

In this section, we give the proof of Theorem 2.3, which shows that consistent testing and estimation is impossible whenever the partition function is $O(1)$. This is a consequence of a general result (see Proposition 6.1 below) which shows that distinguishing two probability measures $\mathbb{P}_N$ versus $\mathbb{Q}_N$ is impossible whenever the KL divergence between the two measures $\mathbb{P}_N$ and $\mathbb{Q}_N$ remains asymptotically bounded.

6.1. Non-existence of consistent tests

For every $N \geq 1$, let $(\mathscr{X}_N, \mathcal{F}_N)$ be a measure space and $\mathbb{P}_N$ and $\mathbb{Q}_N$ two distributions on this measure space. Let μ_N be a dominating measure for both $\mathbb{P}_N$ and $\mathbb{Q}_N$, and p_N and q_N denote the respective densities with respect to this measure. Also, denote the Kullback–Leibler (KL) divergence between $\mathbb{Q}_N$ and $\mathbb{P}_N$ by

$$\begin{aligned} D(\mathbb{Q}_N \| \mathbb{P}_N) &:= \mathbb{E}_{\mathbb{Q}_N} L_N(\boldsymbol{X}) := \mathbb{E}_{\mathbb{Q}_N} \log \frac{q_N(\boldsymbol{X})}{p_N(\boldsymbol{X})} \\ &= \int_{\mathscr{X}_N} q_N(\boldsymbol{x}) \log \frac{q_N(\boldsymbol{x})}{p_N(\boldsymbol{x})} \, \mathrm{d}\mu_N. \end{aligned} \tag{6.1}$$

Consider the problem of testing $\mathbb{P}_N$ versus $\mathbb{Q}_N$. A sequence of tests ϕ_N is *consistent* for this testing problem if there exists a sequence of test functions $\{\phi_N\}_{N\geq 1}$ such that $\lim_{N\to\infty} \mathbb{E}_{\mathbb{P}_N} \phi_N = 0$, and $\lim_{N\to\infty} \mathbb{E}_{\mathbb{Q}_N} \phi_N = 1$.

Proposition 6.1. *Consider the problem of testing $\mathbb{P}_N$ versus $\mathbb{Q}_N$. If*

$$\limsup_{N\to\infty} D(\mathbb{Q}_N \| \mathbb{P}_N) < \infty, \tag{6.2}$$

then there does not exist a consistent sequence of tests for this testing problem.

The proof of the proposition is given in Appendix B. In the following, we use it to prove Theorem 2.3.

6.2. Completing the proof of Theorem 2.3

Given Proposition 6.1, it remains to verify that

$$D(\mathbb{P}_{\beta_1} \| \mathbb{P}_{\beta_2}) = F_N(\beta_2) - F_N(\beta_1) - (\beta_2 - \beta_1) F_N'(\beta_1) < \infty, \tag{6.3}$$

for $0 \leq \beta_1 < \beta_2 \leq \beta_0$ (where β_0 satisfies (2.8)).

By hypothesis (2.8) there exists $M < \infty$ such that $F_N(\beta_1) < M$ and $F_N(\beta_2) < M$, for N large enough. Moreover, by the monotonicity of $F_N'(\cdot)$,

$$(\beta_2 - \beta_1) F_N'(\beta_1) \leq \int_{\beta_1}^{\beta_2} F_N'(\theta) \, \mathrm{d}\theta = F_N(\beta_2) - F_N(\beta_1) < M,$$

proving (6.3).

7. Applications: Proofs of Corollary 2.4, 3.1 and 3.2

In this section, we prove Corollary 2.4 which will then be used to derive rates of consistency of the MPLE for Ising models on different graph ensembles, using Theorems 2.1 and 2.3. To apply these results, we need to determine the correct order of $F_N(\beta_0)$ in a neighborhood of a point $\beta_0 > 0$. However, the exact asymptotics $F_N(\beta_0)$ is known only for specific choices of the matrix J_N and for specific values of β_0.

Nevertheless, the correct order of $F_N(\beta_0)$ can be easily obtained in various examples, using, for instance, the following very useful lemma, which is of independent interest and may find other applications.

Lemma 7.1. *Consider the family of probability distributions on S_N given by (1.2). Assume that the elements of the matrix J_N are non-negative, and $\lambda_1(J_N) \le \lambda_2(J_N) \le \cdots \le \lambda_N(J_N)$ are the eigenvalues of the matrix J_N.*

(a) *For* $0 < \beta < \frac{1}{\|J_N\|}$,

$$F_N(\beta) \le -\frac{1}{2}\sum_{i=1}^{n} \log\bigl(1 - \beta\lambda_i(J_N)\bigr). \tag{7.1}$$

(b) *For any* $\beta > 0$,

$$F_N(\beta) \ge \sum_{1 \le i < j \le N} \log\cosh\bigl(\beta J_N(i,j)\bigr). \tag{7.2}$$

Proof. Let $W := (W_1, W_2, \dots, W_N)'$ be a vector of i.i.d. $N(0,1)$ random variables. Note that for any $s \ge 1$ and non-negative integers $b_1, b_2, \dots, b_s$

$$\mathbb{E}_0 \sigma_1^{b_1}\sigma_2^{b_2}\cdots\sigma_s^{b_s} \le \mathbb{E} W_1^{b_1} W_2^{b_2}\cdots W_s^{b_s}.$$

Since the matrix J_N has non-negative entries, by expanding the exponential function in power series every term can be bounded using the above inequality. This implies that

$$e^{F_N(\beta)} = \mathbb{E}_0 e^{\frac{1}{2}\beta\sigma' J_N \sigma} \le \mathbb{E} e^{\frac{1}{2}\beta W' J_N W}. \tag{7.3}$$

The RHS of (7.3) can be computed exactly as follows: Let $J_N = \sum_{i=1}^N \lambda_i(J_N) p_i p_i'$, be the spectral decomposition of J_N, where $p_1, p_2, \dots, p_N$ are the normalized eigenvectors of J_N. Then setting $p_i' W = Z_i$ for $1 \le i \le N$, we get

$$\mathbb{E} e^{\frac{1}{2}\beta W' J_N W} = \mathbb{E} e^{\frac{1}{2}\beta \sum_{i=1}^N \lambda_i(J_N)(p_i' W)^2} = \mathbb{E} e^{\frac{1}{2}\beta \sum_{i=1}^N \lambda_i(J_N) Z_i^2}. \tag{7.4}$$

Note that $Z := (Z_1, Z_2, \dots, F_N)$ is a vector of i.i.d. $N(0,1)$ random variables. Therefore, (7.3) and (7.4) implies

$$\mathbb{E} e^{\frac{1}{2}\beta W' J_N W} \le \prod_{i=1}^N \bigl(1 - \beta\lambda_i(J_N)\bigr)^{-\frac{1}{2}},$$

using the MGF of the chi-squared distribution (since $\beta\lambda_i(J_N) < 1$, for all $1 \le i \le N$). The inequality (7.1) follows by taking log on both sides.

To prove (b), let $\{Y_{ij}, 1 \le i < j \le N\}$ be i.i.d. with $\mathbb{P}(Y_{ij} = \pm 1) = \frac{1}{2}$. Then for any collection of non-negative integers $((b_{ij}))_{1 \le i < j \le N}$,

$$\mathbb{E} \prod_{1 \le i < j \le N} Y_{ij}^{b_{ij}} \le \mathbb{E}_0 \prod_{1 \le i < j \le N} (\sigma_i \sigma_j)^{b_{ij}}.$$

Indeed, this follows on noting that both the LHS and RHS are $\{0, 1\}$-valued, and the LHS is 1 if and only if b_{ij} is even for all (i, j), which is when the RHS is 1 as well. This implies,

$$\begin{aligned} e^{F_N(\beta)} &= \mathbb{E}_0 \prod_{1 \le i < j \le N} e^{\beta J_N(i,j)\sigma_i\sigma_j} \ge \mathbb{E} \prod_{1 \le i < j \le N} e^{\beta J_N(i,j) Y_{ij}} \\ &= \prod_{1 \le i < j \le N} \cosh\big(\beta J_N(i,j)\big). \end{aligned}$$

The inequality (7.2) follows on taking log on both sides. □

Remark 7.1. Note that the upper bound (7.1) is obtained by replacing the spin configuration $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_N)$ with a vector of i.i.d. $N(0, 1)$ random variables. To get the lower bound, the collection $\{\sigma_i \sigma_j\}_{1 \le i < j \le N}$ is replaced by i.i.d. Rademacher random variables. Surprisingly, the bounds obtained by these simple comparison techniques often give the correct asymptotic order of $F_N(\beta)$ in the high temperature regime $\beta < \frac{1}{\|J_N\|}$. To get the order of $F_N(\beta)$ beyond the phase transition, the standard mean-field approximation can be used (see Section 7.1 for details).

7.1. Proof of Corollary 2.4

For all $\beta > 0$, by the bound (7.2) in Lemma 7.1, we get

$$F_N(\beta) \ge \sum_{1 \le i < j \le N} \log\cosh\big(\beta J_N(i,j)\big) \ge C_1 \beta^2 \sum_{1 \le i < j \le N} J_N^2(i,j), \tag{7.5}$$

where $C_1 := \inf_{|x| \le 1} \frac{\log\cosh x}{x^2} > 0$. To get the upper bound, we use (7.1) for $\beta < \frac{1}{\lambda}$

$$\begin{aligned} F_N(\beta) &\le -\frac{1}{2} \sum_{i=1}^N \log\big(1 - \beta\lambda_i(J_N)\big) \le \frac{C_2\beta^2}{2} \sum_{i=1}^N \lambda_i\big(J_N^2\big) \\ &\le \frac{C_2\beta^2}{2} \operatorname{tr}\big(J_N^2\big) \qquad (7.6) \\ &= \frac{C_2\beta^2}{2} \sum_{i,j=1}^N J_N(i,j)^2, \end{aligned}$$

where $C_2 = C_2(\beta) := \sup_{|x|\le\beta} \frac{-\log(1-x)-x}{x^2} < \infty$ for any $\beta < \frac{1}{\lambda}$, and we use the fact that $\sum_{i=1}^N \lambda_i(J_N) = 0$. The bounds (7.5) and (7.6) together implies (2.6) with $a_N = \sum_{i,j=1}^N J_N(i,j)^2$ for $\beta = \beta_0 < \frac{1}{\lambda}$. Therefore, if $\sum_{i,j=1}^N J_N(i,j)^2 \to \infty$, part (a) follows by Theorem 2.1.

Finally, if $\limsup_{N\to\infty} \sum_{i,j=1}^N J_N(i,j)^2 < \infty$, then $F_N(\beta) = O(1)$ for $\beta < \frac{1}{\lambda}$ and by Theorem 2.3 part (b) follows.

7.2. Proof of Example 1

It is well known that in the Curie–Weiss model for $\beta > 1$ (see [8], Example 3.9)

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E}_0 e^{\frac{\beta}{2N}\sum_{1\le i\ne j\le N}\sigma_i\sigma_j} := F(\beta) \in (0,\infty). \tag{7.7}$$

Note that

$$\begin{aligned}\sigma' J_N \sigma &= \frac{1}{N}\sum_{1\le i\ne j\le \frac{N}{2}} \sigma_i\sigma_j + \frac{1}{\sqrt{N}} \sum_{\frac{N}{2}<i\ne j\le \frac{N}{2}+\sqrt{N}} \sigma_i\sigma_j \\ &= \sigma^{(1)'} A_N \sigma^{(1)} + \sigma^{(2)'} B_N \sigma^{(2)},\end{aligned}$$

where A_N is a $N/2 \times N/2$ matrix with $A_N(i,j) = 1/N$, for $i \ne j$, and B_N is a $\sqrt{N} \times \sqrt{N}$ matrix with $B_N(i,j) = 1/\sqrt{N}$, for $i \ne j$, and $\sigma = (\sigma^{(1)}, \sigma^{(2)})$. Therefore,

$$e^{F_N(\beta)} = \mathbb{E}_0 e^{\frac{\beta}{2}\sigma' J_N \sigma} = \mathbb{E}_0 e^{\frac{\beta}{2}\sigma^{(1)'} A_N \sigma^{(1)}} \mathbb{E}_0 e^{\frac{\beta}{2}\sigma^{(2)'} B_N \sigma^{(2)}}. \tag{7.8}$$

Note that $J_N\| = 1$ and by (7.1) $F_N(\beta) = O(1)$ for $\beta < 1$. Thus, there exists no sequence of consistent estimators for $\beta \in (0,1)$ by Theorem 2.3.

For $1 < \beta < 2$, by (7.7)

$$0 < \liminf_{N\to\infty} \frac{1}{\sqrt{N}} F_N(\beta) \le \limsup_{N\to\infty} \frac{1}{\sqrt{N}} F_N(\beta) < \infty,$$

since $\sigma^{(2)'} B_N \sigma^{(2)}$ is the Hamiltonian of a Curie–Weiss model on size $\sqrt{N}$. Moreover, $|m_i(\tau)| \le 1$, for all $1 \le i \le N$ and $\tau \in S_N$; so taking $K = 1$, $\frac{1}{\sqrt{N}} \sum_{i=1}^N |m_i(\sigma)| \mathbf{1}\{|m_i(\sigma)| > K\} = 0$, establishing condition (a) of Theorem 2.1. Therefore, the MPLE $\{\hat{\beta}_N\}_{N\ge 1}$ is $N^{1/4}$-consistent for $\beta \in (1,2)$ by Theorem 2.1.

Similarly, for $\beta > 2$

$$0 < \liminf_{N\to\infty} \frac{1}{N} F_N(\beta) \le \limsup_{N\to\infty} \frac{1}{N} F_N(\beta) < \infty,$$

and so the MPLE $\{\hat{\beta}_N\}_{N\ge 1}$ is $\sqrt{N}$-consistent.

7.3. Proof of Corollary 3.1

Note that when the sufficient statistic is of the form (3.1), $|m_i(\tau)| \le 1$, for all $\tau \in S_N$. Therefore, taking $K = 1$, $\frac{d_N}{N}\sum_{i=1}^N |m_i(\sigma)|\mathbf{1}\{|m_i(\sigma)| > K\} = 0$, which implies condition (a) of Theorem 2.1. Moreover, in this case, $\|J_N\| = 1$, and $\sum_{i,j=1}^N J_N(i,j)^2 = N/d_N$. Therefore, part (a) follows by Corollary 2.4.

By Corollary 2.2, to show part (b) it suffices to verify that condition (2.5) holds for all $\beta_0 > 1$. This is done using the mean field approximation of Lemma C.1. By plugging in the vector $(m, m, \ldots, m)'$ for the vector $\mathbf{z}$ in the RHS of (C.1)

$$F_N(\beta_0) \ge N \sup_{m\in[-1,1]} \left\{\frac{\beta_0 m^2}{2} - I(m)\right\}, \tag{7.9}$$

where $I(x) := \frac{1}{2}(1+x)\log(1+x) + \frac{1}{2}(1-x)\log(1-x)$ for $x \in [-1,1]$. Thus, it suffices to show that $\sup_{m\in[-1,1]} g(m) > 0$, where $g(m) := \frac{\beta_0 m^2}{2} - I(m)$. To this end, note that $g''(0) = \beta_0 - 1 > 0$, that is, $m = 0$ is not a local maximum of g. This implies the RHS of (7.9) is positive, thus verifying condition (2.5).

7.4. Proof of Corollary 3.2

Let d_i be the degree of the vertex i in G_N, for $1 \le i \le N$. Then $|m_i(\tau)| \le \frac{d_i}{Np(N)}$, for all $\tau \in S_N$. In the regime $\frac{\log N}{N} \ll p(N) \le 1$, the maximum degree $\Delta = \max_{i\in V(G_N)} d_i = Np(N)(1+o(1))$ with high probability [25]. Therefore, $|m_i(\tau)| \le 1 + o(1)$ for all $1 \le i \le N$, and by taking $K \ge 2$ it follows that $p(N)\sum_{i=1}^N |m_i(\sigma)|\mathbf{1}\{|m_i(\sigma)| > K\} = 0$, with high probability. This implies condition (a) of Theorem 2.1.

Moreover, for $\frac{\log N}{N} \ll p(N) \le 1$, $\|J_N\| = 1 + o(1)$ with high probability [25], and

$$p(N)\sum_{i,j=1}^N J_N(i,j)^2 = \frac{2}{N^2 p(N)}\big|E(G_N)\big| \xrightarrow{\mathscr{P}} 1,$$

and part (a) follows from Corollary 2.4.

To prove part (b), we use the mean field approximation as in Corollary 3.1. By plugging in the vector $(m, m, \ldots, m)'$ for the vector $\mathbf{z}$ in the RHS of (C.1), we get

$$F_N(\beta_0) \ge N \sup_{m\in[-1,1]} \left\{\frac{\beta_0 m^2 |E(G_N)|}{N^2 p(N)} - I(m)\right\}.$$

Condition (2.5) follows by arguments similar to those in Corollary 3.1 and the fact $\frac{2|E(G_N)|}{N^2 p(N)} \xrightarrow{\mathscr{P}} 1$.

8. Proofs of Theorems 3.3 and 3.4

In this section, we show the existence of a untestable/testable threshold in Ising models on converging sequence of dense graphs, and compute the distribution and asymptotic power of the most powerful test, before the phase transition.

8.1. Proof of Theorem 3.3

If G_N converges to W, then $\frac{1}{N}\|A(G_N)\|$ converges to the operator norm of $\|W\|$ (see (3.4)). Moreover,

$$\frac{1}{N^2}\sum_{i=1}^{N}\lambda_i\big(A(G_N)^2\big) \to t(C_2, W),$$

and part (a) follows by Corollary 2.4.

We now show (b). From [8], Theoem 2.14, when G_N converges to W, then $\lim_{N\to\infty}\frac{1}{N}F_N(\beta) = \mathscr{E}(W,\beta)$, where

$$\mathscr{E}(W,\beta) := \sup_{m:[0,1]\mapsto[-1,1]}\left\{\frac{\beta}{2}\int_{[0,1]^2} m(x)m(y)W(x,y)\,\mathrm{d}x\,\mathrm{d}y - \int_0^1 I\big(m(x)\big)\,\mathrm{d}x\right\}, \tag{8.1}$$

and $I(x) = \frac{1}{2}\{(1+x)\log(1+x) + (1-x)\log(1-x)\}$ as in Corollary 3.1. By Corollary 2.2, it enough to show that $\mathscr{E}(W,\beta) > 0$, for $\beta > \frac{1}{\|W\|}$.

To this end, let $v_1(x)$ to be the eigenvector corresponding to the eigenvalue $\lambda = \|W\|$. Then $|\lambda v_1(x)| = |\int_0^1 W(x,y)v_1(y)\,\mathrm{d}y| \le 1$, and $\sup_{x\in[0,1]}|v_1(x)| < \infty$. Thus, there exists $\delta > 0$ such that for $z \in (-\delta,\delta)$ we have $\sup_{x\in[0,1]}|zv_1(x)| \le 1$, and

$$\begin{aligned}\mathscr{E}(W,\beta) &\ge \sup_{|z|<\delta}\left\{\frac{\beta}{2}z^2\int_{[0,1]^2} v_1(x)v_1(y)W(x,y)\,\mathrm{d}x\,\mathrm{d}y - \int_0^1 I\big(zv_1(x)\big)\,\mathrm{d}x\right\}\\ &= \sup_{|z|<\delta}\left\{\frac{\beta z^2\lambda}{2}\int_0^1 v_1(x)^2\,\mathrm{d}x - \int_0^1 I\big(zv_1(x)\big)\,\mathrm{d}x\right\}.\end{aligned}$$

Setting $h(z) := \frac{\beta z^2\lambda}{2}\int_0^1 v_1(x)^2\,\mathrm{d}x - \int_0^1 I(zv_1(x))\,\mathrm{d}x$ it suffices to show that $z = 0$ is not a point of local maxima of the function h. This follows on noting that $h''(0) = (\beta\lambda - 1)\int_0^1 v_1(x)^2\,\mathrm{d}x > 0$.

8.2. Proof of Theorem 3.4

By Lemma D.1 (see Appendix D), the limiting distribution (3.6) is well defined.

The following proposition (proved in Appendix D) gives the limit of the log-partition function, for a converging sequence of dense graphs, for $\beta < \frac{1}{\|W\|}$.

Proposition 8.1. *Let $\{G_N\}_{N\geq 1}$ be a sequence of simple graphs converging in cut-metric to $W \in \mathscr{W}$, such that $\int_{[0,1]^2} W(x,y)^2 \,\mathrm{d}x\,\mathrm{d}y > 0$. Then for any $0 < \beta < \frac{1}{\|W\|}$*

$$\lim_{N\to\infty} F_N(\beta) = -\frac{1}{2}\sum_{i=1}^{\infty}\bigl\{\log\bigl(1-\beta\lambda_i(W)\bigr) - \beta\lambda_i(W)\bigr\}. \tag{8.2}$$

The above proposition can be used to complete the proof of Theorem 3.3 as follows: Fix $\delta > 0$ such that $\beta + \delta < \frac{1}{\|W\|}$. Then for any $t \in (-\beta, \delta)$,

$$\begin{aligned}\mathbb{E}_\beta \exp\left\{\frac{t}{2}\cdot\frac{1}{N}\sigma' W_N \sigma\right\} &= \exp\bigl\{F_N(\beta+t) - F_N(\beta)\bigr\}\\ &\to \prod_{i=1}^{\infty}\frac{e^{-\frac{1}{2}t\lambda_i(W)}}{\sqrt{1-\frac{t\lambda_i(W)}{1-\beta\lambda_i(W)}}},\end{aligned} \tag{8.3}$$

by Proposition 8.1.

By Lemma D.1 the RHS above is the MGF of the random variable $\frac{Q_{\beta,W}}{2}$ defined in (3.6).

Appendix A: Proofs of technical lemmas

In the appendix we prove the lemmas used in the proof of Theorem 2.1. The rest of the section is organized as follows: Appendix A.1 contains the proof of Lemma 5.1. The proofs of Lemmas 5.2 and 5.3 are given in Appendices A.2 and A.3, respectively.

A.1. Proof of Lemma 5.1

By (2.6) there exists $\delta \in (0, \beta_0/2)$ such that $\liminf_{N\to\infty}\frac{1}{a_N}F_N(\beta_0-\delta) > 0$. By the monotonicity of $F_N'(\cdot)$,

$$F_N(\beta_0-\delta) = \int_0^{\beta_0-\delta} F_N'(t)\,\mathrm{d}t \leq (\beta_0-\delta)F_N'(\beta_0-\delta) \leq \beta_0 F_N'(\beta_0-\delta),$$

it follows that $\liminf_{N\to\infty}\frac{1}{a_N}F_N'(\beta_0-\delta) > 0$. Thus, for any $\varepsilon > 0$

$$\mathbb{P}_{\beta_0}\bigl(H_N(\sigma) < \varepsilon a_N\bigr) = \mathbb{P}_{\beta_0}\bigl(e^{-\frac{1}{2}\delta H_N(\sigma)} > e^{-\frac{1}{2}\delta\varepsilon a_N}\bigr) \leq e^{\frac{1}{2}\delta\varepsilon a_N + F_N(\beta_0-\delta) - F_N(\beta_0)},$$

which, on taking logarithms, implies that

$$\log \mathbb{P}_{\beta_0}\bigl(H_N(\sigma) < \varepsilon a_N\bigr) \leq \frac{\varepsilon\delta a_N}{2} - \int_{\beta_0-\delta}^{\beta_0} F_N'(t)\,\mathrm{d}t \leq \frac{\varepsilon\delta a_N}{2} - F_N'(\beta_0-\delta)\delta.$$

Dividing both sides by a_N and taking limits as $N \to \infty$ followed by $\varepsilon \to 0$ we have

$$\lim_{\varepsilon \to 0} \limsup_{N\to\infty} \frac{1}{a_N} \log \mathbb{P}_{\beta_0}\big(H_N(\sigma) < \varepsilon a_N\big) \le -\liminf_{N\to\infty} \frac{1}{a_N} F_N'(\beta_0 - \delta) < 0,$$

thus completing the proof of (a).

To show (b), again invoking (2.6) there exists $\delta > 0$ such that $\limsup_{N\to\infty} F_N(\beta_0 + 2\delta) < \infty$. Since

$$F_N(\beta_0 + 2\delta) = \int_0^{\beta_0+2\delta} F_N'(t)\,\mathrm{d}t \ge \delta F_N'(\beta_0 + \delta),$$

it follows that $\limsup_{N\to\infty} \frac{1}{a_N} F_N'(\beta_0 + \delta) < \infty$. Thus, for any $K < \infty$

$$\mathbb{P}\big(H_N(\sigma) > K a_N\big) = \mathbb{P}\big(e^{\frac{1}{2}\delta H_N(\sigma)} > e^{\frac{1}{2}\delta K a_N}\big) \le e^{-\frac{1}{2}\delta K a_N + F_N(\beta_0+\delta) - F_N(\beta_0)}.$$

Taking logarithm on both sides,

$$\begin{aligned}\log \mathbb{P}\big(H_N(\sigma) > K a_N\big) &\le -\frac{\delta K a_N}{2} + \int_{\beta_0}^{\beta_0+\delta} F_N'(t)\,\mathrm{d}t \\ &\le -\frac{\delta K a_N}{2} + F_N'(\beta_0 + \delta),\end{aligned}$$

from which dividing by a_N and taking limits as $N \to \infty$ followed by $K \to \infty$ gives

$$\lim_{K\to\infty} \limsup_{N\to\infty} \frac{1}{a_N} \log \mathbb{P}\big(H_N(\sigma) > K a_N\big) = -\infty,$$

thus proving part (b).

A.2. Proof of Lemma 5.2

We begin with a technical estimate which will be needed to bound the second moment of $L_\sigma(\beta_0)$.

Lemma A.1. *Under assumption* (2.6) *and* $m_i(\sigma)$ *as defined in* (2.3),

$$\limsup_{N\to\infty} \frac{1}{a_N} \mathbb{E}_{\beta_0} \sum_{i=1}^N m_i(\sigma) \tanh\big(\beta_0 m_i(\sigma)\big) < \infty.$$

Proof. By (2.6) there exists $\delta > 0$ such that $\limsup_{N\to\infty} \frac{1}{a_N} F_N(\beta_0 + \delta) < \infty$. Therefore, $F_N(\beta_0 + \delta) = \int_0^{\beta_0+\delta} F_N'(t)\,\mathrm{d}t \ge \delta F_N'(\beta_0)$, and so

$$\frac{1}{a_N} \limsup_{N\to\infty} Z_N'(\beta_0) < \infty. \tag{A.1}$$

Now, observe that $m_i(\sigma)$ does not depend on σ_i, and $\mathbb{E}_{\beta_0}(\sigma_i|(\sigma_j)_{j\neq i}) = \tanh(\beta_0 m_i(\sigma))$. Since

$$2F_N'(\beta_0) = \mathbb{E}_{\beta_0} H_N(\sigma) = \mathbb{E}_{\beta_0}\left(\sum_{i=1}^N \sigma_i m_i(\sigma)\right)$$
$$= \mathbb{E}_{\beta_0}\left(\sum_{i=1}^N m_i(\sigma)\tanh\bigl(\beta_0 m_i(\sigma)\bigr)\right),$$

the result follows from (A.1). $\square$

The above lemma will be used to complete the proof of Lemma 5.2. To this end, for $1 \le j \le N$ and $\tau \in S_N$, let

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

and

$$p_j(\tau) = \frac{e^{-\beta_0\tau_j m_j(\tau)}}{e^{\beta_0\tau_j m_j(\tau)} + e^{-\beta_0\tau_j m_j(\tau)}}. \tag{A.2}$$

From equation (10) of Chatterjee [10] it follows that

$$\mathbb{E}_\beta\bigl(L_\sigma(\beta_0)^2\bigr) = \frac{1}{N}\mathbb{E}_\beta \sum_{j=1}^N \bigl(L_\sigma(\beta_0) - L_{\sigma^{(j)}}(\beta_0)\bigr) m_j(\sigma)\sigma_j p_j(\sigma). \tag{A.3}$$

Setting $r(x) := x\tanh(\beta_0 x)$, note that

$$L_\sigma(\beta_0) - L_{\sigma^{(j)}}(\beta_0) = \frac{2m_j(\sigma)\sigma_j}{N} + \frac{1}{N}\sum_{i=1}^N \bigl\{r\bigl(m_i\bigl(\sigma^{(j)}\bigr)\bigr) - r\bigl(m_i(\sigma)\bigr)\bigr\}.$$

Now, by a second order Taylor expansion,

$$\mathbb{E}_\beta\bigl(L_\sigma(\beta_0)^2\bigr) = \frac{a_N}{N^2}(T_1 + T_2 + T_3), \tag{A.4}$$

where

$$T_1 = \frac{2}{a_N}\sum_{j=1}^N m_j(\sigma)^2 p_j(\sigma),$$

$$T_2 = -\frac{2}{a_N}\sum_{i,j=1}^N J_N(i,j) r'\bigl(m_i(\sigma)\bigr) m_j(\sigma) p_j(\sigma)$$

and

$$T_3 = \frac{2}{a_N} \sum_{i,j=1}^{N} r''\big(\theta_{ij}(\sigma)\big) J_N(i,j)^2 m_j(\sigma) \sigma_j p_j(\sigma),$$

for some $\theta_{ij}(\sigma)$ in the interval $[m_i(\sigma^{(j)}), m_i(\sigma)]$. Therefore, to prove the lemma, it suffices to control these three terms.

To control T_1, note that

$$\mathbb{E}_{\beta_0}\big(p_j(\sigma)|(\sigma_i)_{i\neq j}\big) = \frac{2}{e^{\beta_0 m_j(\sigma)} + e^{-\beta_0 m_j(\sigma)}} = \frac{1}{2}\operatorname{sech}^2\big(\beta_0 m_j(\sigma)\big),$$

and $x^2 \operatorname{sech}^2(\beta_0 x) \leq M_1 x \tanh(\beta_0 x)$ for all $x \in \mathbb{R}$ for some $M_1 = M_1(\beta_0) < \infty$, which gives

$$\begin{aligned} \mathbb{E}_{\beta_0} T_1 &= \frac{1}{a_N}\mathbb{E}_{\beta_0} \sum_{j=1}^{N} m_j(\sigma)^2 \operatorname{sech}^2\big(\beta_0 m_j(\sigma)\big) \\ &\leq \frac{M_1}{a_N}\mathbb{E}_{\beta_0} \sum_{j=1}^{N} m_j(\sigma) \tanh\big(\beta_0 m_j(\sigma)\big), \end{aligned} \tag{A.5}$$

which is finite as $N \to \infty$ by an application of Lemma A.1.

Now, let us bound T_2. By the Cauchy–Schwarz inequality,

$$\begin{aligned} |T_2| &\leq \frac{2}{a_N}\left\{\sum_{i=1}^{N} r'\big(m_i(\sigma)\big)^2\right\}^{1/2} \left\{\sum_{i=1}^{N}\left(\sum_{j=1}^{N} J_N(i,j) m_j(\sigma)\sigma_j p_j(\sigma)\right)^2\right\}^{1/2} \\ &\leq \frac{2\|J_N\|}{a_N}\left\{\sum_{i=1}^{N} r'\big(m_i(\sigma)\big)^2\right\}^{1/2} \left\{\sum_{j=1}^{N} m_j(\sigma)^2 p_j(\sigma)^2\right\}^{1/2}. \end{aligned}$$

Taking expectation on both sides above and using Cauchy–Schwarz inequality again

$$\mathbb{E}_{\beta_0}|T_2| \leq \frac{2\|J_N\|}{a_N}\left\{\mathbb{E}_{\beta_0}\sum_{i=1}^{N} r'\big(m_i(\sigma)\big)^2 \cdot \mathbb{E}_{\beta_0}\sum_{j=1}^{N} m_j(\sigma)^2 p_j(\sigma)\right\}^{1/2}. \tag{A.6}$$

Now, since $r'(x)^2 = \{\tanh(\beta_0 x) + \beta_0 x \operatorname{sech}^2(\beta_0 x)\}^2 \leq M_2 x \tanh(\beta_0 x)$, for some constant $M_2 = M_2(\beta_0)$, by Lemma A.1

$$\limsup_{N\to\infty} \frac{1}{a_N}\mathbb{E}_{\beta_0}\sum_{i=1}^{N} r'\big(m_i(\sigma)\big)^2 < \infty.$$

Using this along with (A.5) in (A.6) gives $\limsup_{N\to\infty} \mathbb{E}_{\beta_0}|T_2| < \infty$.

It remains to bound T_3. Since $M_3 = M_3(\beta_0) := \sup_{x\in\mathbb{R}} |r''(x)| < \infty$, we have

$$\begin{aligned}|T_3| &\le \frac{2M_3}{a_N}\sum_{i,j=1}^N J_N(i,j)^2\big|m_j(\sigma)\big|p_j(\sigma)\\ &\le \frac{2M_3}{a_N}\left\{\sum_{j=1}^N\left(\sum_{i=1}^N J_N(i,j)^2\right)^2\right\}^{1/2}\left\{\sum_{j=1}^N m_j(\sigma)^2 p_j(\sigma)\right\}^{1/2} \\ &\le \frac{2M_3\|J_N\|}{a_N}\left\{\sum_{i,j=1}^N J_N(i,j)^2\right\}^{1/2}\left\{\sum_{j=1}^N m_j(\sigma)^2 p_j(\sigma)\right\}^{1/2},\end{aligned} \tag{A.7}$$

where the last step uses $\sum_{i=1}^N J_N(i,j)^2 = \|J_N e_j\|^2 \le \|J_N\|^2$. Finally, taking expectations on both sides in (A.7), and using condition (b) on the first term, and (A.5) on the second term, gives $\limsup_{N\to\infty} \mathbb{E}_{\beta_0}|T_3| < \infty$.

A.3. Proof of Lemma 5.3

Fixing $\delta > 0$ by Lemma 5.1(a) there exists $\varepsilon = \varepsilon(\delta) > 0$ such that

$$\mathbb{P}_{\beta_0}\big(H_N(\sigma) < 3\varepsilon\beta_0 a_N\big) \le \delta, \tag{A.8}$$

for N large enough. Also, using Lemma 5.2 and Chebyshev's inequality, for $K_1 = K_1(\delta) := \sqrt{\frac{C(\beta_0)}{\delta}}$ we have

$$\mathbb{P}_{\beta_0}\big(\big|L_\sigma(\beta_0)\big| > K_1\sqrt{a_N}/N\big) \le \frac{N^2}{a_N}\mathbb{E}_{\beta_0} L_\sigma(\beta_0)^2 \le \delta. \tag{A.9}$$

Moreover, by condition (a) in Theorem 2.1 there exists $K_2 = K_2(\delta) < \infty$ such that for all N large enough we have

$$\mathbb{E}_{\beta_0}\sum_{i=1}^N\big|m_j(\sigma)\big|\mathbf{1}\big\{\big|m_j(\sigma)\big| > K_2\big\} \le \varepsilon\delta\beta_0 a_N$$

and so by Markov's inequality

$$\begin{aligned}&\mathbb{P}_{\beta_0}\left(\sum_{i=1}^N\big|m_j(\sigma)\big|\mathbf{1}\big\{\big|m_j(\sigma)\big| > K_2\big\} > \varepsilon\beta_0 a_N\right)\\ &\qquad\le \frac{\mathbb{E}_{\beta_0}\sum_{i=1}^N |m_j(\sigma)|\mathbf{1}\{|m_j(\sigma)| > K_2\}}{\varepsilon\beta_0 a_N} \le \delta.\end{aligned} \tag{A.10}$$

Defining

$$A_N(\delta) := \Bigg\{\sigma \in S_N : H_N(\sigma) \geq 3\varepsilon\beta_0 a_N, \big|L_\sigma(\beta_0)\big| \leq K_1 \frac{\sqrt{a_N}}{N},$$

$$\sum_{i=1}^{N} \big|m_j(\sigma)\big| \mathbf{1}\big\{|m_j(\sigma)| > K_2\big\} > \varepsilon\beta_0 a_N\Bigg\},$$

we have $\mathbb{P}_{\beta_0}(A_N(\delta)) \geq 1 - 3\delta$, for N large enough (by combining (A.8), (A.9), and (A.10)).

Now, on the set $A_N(\delta)$ using the bounds $\tanh x \leq x$ on $x \leq K_2$, and $\tanh x \leq 1$ on $x > K_2$,

$$\begin{aligned}\beta_0 \sum_{i=1}^{N} m_i(\sigma)^2 \mathbf{1}\big\{\big|m_i(\sigma)\big| \leq K_2\big\} + \varepsilon\beta_0 a_N &\geq \sum_{i=1}^{N} m_i(\sigma) \tanh\big(\beta_0 m_i(\sigma)\big) \\ &= H_N(\sigma) - N L_\sigma(\beta_0) \geq 3\varepsilon\beta_0 a_N - K_1\sqrt{a_N}.\end{aligned}$$

Thus, on the set $A_N(\delta)$,

$$\sum_{i=1}^{N} m_i(\sigma)^2 \mathbf{1}\big\{\big|m_i(\sigma)\big| \leq K_2\big\} \geq 2\varepsilon a_N - \frac{K_1}{\beta_0}\sqrt{a_N} > \varepsilon a_N$$

for all N large, completing the proof.

Appendix B: Proof of Lemma 6.1

For every $N \geq 1$, let $(\mathscr{X}_N, \mathcal{F}_N)$ be a measure space and $\mathbb{P}_N$ and $\mathbb{Q}_N$ two distributions on this measure space. Recall the definition of Kullback–Leibler divergence $D(\mathbb{Q}_N \| \mathbb{P}_N)$ from (6.1), and consider the problem of testing $\mathbb{P}_N$ versus $\mathbb{Q}_N$ such that condition (6.2) holds. Since $D(\mathbb{Q}_N \| \mathbb{P}_N) = \mathbb{E}_{\mathbb{Q}_N} L_N$, by assumption (6.2)

$$0 \leq \mathbb{E}_{\mathbb{Q}_N} L_N = \mathbb{E}_{\mathbb{Q}_N} L_N^+ - \mathbb{E}_{\mathbb{Q}_N} L_N^- \leq M_1, \tag{B.1}$$

for some $M_1 < \infty$ and all large N. Also, there exists $M_2 < \infty$ such that $\mathbb{E}_{Q_N} L_N^- \leq M_2$, for all N. To see this, note that

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}_N} L_N^- &= -\sum_{s=1}^{\infty} \mathbb{E}_{\mathbb{Q}_N} L_N \mathbf{1}\{-s \leq L_N < -s+1\} \\ &\leq \sum_{s=1}^{\infty} s e^{-(s-1)} \mathbb{P}_N(-s \leq L_N < -s+1) \qquad \text{(B.2)} \\ &\leq \sum_{s=1}^{\infty} s e^{-(s-1)} := M_2 < \infty.\end{aligned}$$

Hence, by (B.1) and (B.2), $\mathbb{E}_{\mathbb{Q}_N}|L_N| = \mathbb{E}_{\mathbb{Q}_N} L_N^+ + \mathbb{E}_{\mathbb{Q}_N} L_N^- \le M_1 + 2M_2 =: M < \infty$. Therefore, by Markov's inequality, for any $\varepsilon > 0$

$$\mathbb{Q}_N\big(|L_N| > M/\varepsilon\big) \le \frac{\varepsilon}{M}\mathbb{E}_{\mathbb{Q}_N}\big(|L_N|\big) \le \varepsilon.$$

Now, suppose there exists a sequence of test functions ϕ_N such that $\mathbb{E}_{\mathbb{P}_N}\phi_N \to 0$. Then

$$\mathbb{E}_{\mathbb{Q}_N}\phi_N \le \mathbb{Q}_N\big(|L_N| > M/\varepsilon\big) + \mathbb{E}_{\mathbb{Q}_N}\big(\phi_N \mathbf{1}\{|L_N| \le M/\varepsilon\}\big) \le \varepsilon + e^{M/\varepsilon}\mathbb{E}_{\mathbb{P}_N}\phi_N.$$

Taking limits on both sides gives, $\limsup_{N\to\infty} \mathbb{E}_{\mathbb{Q}_N}\phi_N \le \varepsilon$. Since $\varepsilon > 0$ is arbitrary $\lim_{N\to\infty} \mathbb{E}_{\mathbb{Q}_N}\phi_N = 0$, that is, ϕ_N is not a consistent sequence of test functions.

Appendix C: The mean-field approximation

A standard technique to derive a lower bound on the log-partition function is the mean-field approximation (refer to [13] for details). Here, we give a short proof for the sake of completeness.

Lemma C.1. *Consider the family of probability distributions on S_N given by* (1.2). *Then for any matrix*

$$F_N(\beta) \ge \sup_{\mathbf{z}\in[-1,1]^N} \left\{ \frac{\beta}{2}\mathbf{z}' J_N \mathbf{z} - \sum_{i=1}^N I(z_i) \right\}, \tag{C.1}$$

where $I(x) = \frac{1}{2}[(1+x)\log(1+x) + (1-x)\log(1-x)]$ for $x \in [-1,1]$.

Proof. Let $D(\cdot\|\cdot)$ be the Kullback–Leibler divergence between two probability measures. By a direction computation, for any probability mass function ν on $S_N = [-1,1]^N$ we have

$$D(\nu\|\mathbb{P}_\beta) = F_N(\beta) + N\log 2 + \mathbb{E}_\nu \log \nu(\sigma) - \frac{1}{2}\mathbb{E}_\nu H_N(\sigma).$$

Now, since $D(\nu\|\mathbb{P}_\beta) \ge 0$ we have

$$F_N(\beta) \ge \frac{\beta}{2}\mathbb{E}_\nu H_N(\sigma) - \mathbb{E}_\nu \log \nu(\sigma) - N\log 2.$$

One can obtain a lower bound on $F_N(\beta)$ by taking supremum in LHS over product measures, that is $\nu(\sigma) = \prod_{i=1}^N \nu_i(\sigma_i)$. Hence, setting $z_i = \mathbb{E}_{\nu_i}\sigma = \nu_i(1) - \nu_i(-1) \in [-1,1]$, the bound in (C.1) follows. □

Appendix D: Proof of Proposition 8.1

We begin by deriving the MGF of the limiting distribution (3.6). The proof involves straightforward calculations using the MGF of the chi-squared distribution, similar to [6], Proposition 7.1.

Lemma D.1. *Let* $\{a_i\}_{i\geq 1}, \{b_i\}_{i\geq 1}$ *be a sequence of real numbers such that* $\sum_{i=1}^\infty a_i^2 < \infty$ *and* $\sum_{i=1}^\infty (a_i - b_i) = \mu$ *for some finite real number* μ. *Suppose* $\xi_1, \xi_2, \dots$ *be i.i.d.* χ_1^2 *random variables.*

(a) *Then the sum* $S := \frac{1}{2}\sum_{i=1}^\infty (a_i\xi_i - b_i)$ *converges almost surely and in* L^1.
(b) *Moroever, if* $M := \sup_{i\geq 1} |a_i| < \infty$, *then for* $0 < t < \frac{1}{M}$,

$$\mathbb{E}e^{\frac{1}{2}t\sum_{i=1}^\infty (a_i\xi_i - b_i)} = \prod_{i=1}^\infty \frac{e^{-\frac{1}{2}tb_i}}{\sqrt{1-ta_i}}. \tag{D.1}$$

Proof. By defining $S_N := \frac{1}{2}\sum_{i=1}^N (a_i\xi_i - b_i)$ and $\mathscr{F}_N := \sigma(\{\xi_j\}_{j=1}^N)$, it follows that $(S_N, \mathscr{F}_N)$ is a martingale, with

$$\limsup_N \mathbb{E}S_N^2 = \frac{1}{4}\left(\mu^2 + \sum_{j=1}^\infty a_j^2\right) < \infty,$$

and so S_N converges almost surely and in L^1 [17].

To compute the moment generating function of S, first note that $e^{tS_N} \stackrel{\mathscr{P}}{\to} e^{tS}$. Thus if the collection of random variables $\{e^{tS_N}\}$ is uniformly integrable, then we have

$$\mathbb{E}e^{tS} = \lim_{N\to\infty} \mathbb{E}e^{tS_N} = \lim_{N\to\infty}\prod_{i=1}^N \frac{e^{-\frac{1}{2}tb_i}}{\sqrt{1-ta_i}} = \prod_{i=1}^\infty \frac{e^{-\frac{1}{2}tb_i}}{\sqrt{1-ta_i}},$$

thus completing the proof of the lemma. It thus remains to prove uniform integrability, for which it suffices to show that for some $\delta > 0$ we have $\limsup_{N\to\infty} \mathbb{E}e^{(t+\delta)S_N} < \infty$. Since $t < \frac{1}{M}$ there exists $\delta > 0$ such that $t + \delta < \frac{1}{M}$. For this δ setting $t' := t + \delta$ we have

$$\log \mathbb{E}e^{t'S_N} = \frac{1}{2}\sum_{i=1}^N \bigl\{-t'b_i - \log\bigl(1 - t'a_i\bigr)\bigr\}. \tag{D.2}$$

Now setting $C := \sup_{|x|\leq t'M} \frac{-\log(1-x)-x}{x^2} < \infty$ we have $-\log(1-x) - x \leq Cx^2$ for $|x| < t'M$, and so the RHS of (D.2) can be bounded by $\frac{1}{2}\sum_{i=1}^N \{t'(a_i - b_i) + Ct'^2 a_i^2\}$, which converges to $e^{t'\mu + Ct'^2\sum_{i=1}^\infty a_i^2}$. Therefore, e^{tS_N} is uniformly integrable, thus completing the proof of the lemma. □

The above lemma can be used to complete the proof of Proposition 8.1. To this end, let $W_N := A(G_N)$. Then, by [6], Theorem 1.4, it follows that

$$\frac{1}{N}\sigma' W_N \sigma \stackrel{\mathscr{D}}{\to} \sum_{i=1}^\infty \lambda_i(W)(\xi_i - 1),$$

where $\xi_1, \xi_2, \dots,$ are i.i.d. χ_1^2 random variables. Thus,

$$\exp\left\{\frac{\beta}{2} \cdot \frac{1}{N}\sigma' W_N \sigma\right\} \xrightarrow{\mathscr{D}} \exp\left\{\frac{\beta}{2} \sum_{i=1}^{\infty} \lambda_i(W)(\xi_i - 1)\right\}. \tag{D.3}$$

If the LHS in (D.3) is uniformly integrable, then

$$\begin{aligned} \lim_{N\to\infty} \mathbb{E}\exp\left\{\frac{\beta}{2} \cdot \frac{1}{N}\sigma' W_N \sigma\right\} &= \mathbb{E}\exp\left\{\frac{\beta}{2} \sum_{i=1}^{\infty} \lambda_i(W)(\xi_i - 1)\right\} \\ &= \prod_{i=1}^{\infty} \frac{e^{-\frac{1}{2}\beta\lambda_i(W)}}{\sqrt{1-\beta\lambda_i(W)}}, \end{aligned} \tag{D.4}$$

where the last equality uses Lemma D.1. The proof of part (a) then follows on taking log of both sides of the above equality.

It remains to show that the LHS in (D.3) is uniformly integrable, that is,

$$\limsup_{N\to\infty} \log \mathbb{E}_0 \exp\left\{\frac{\beta+\delta}{2} \cdot \frac{1}{N}\sigma' W_N \sigma\right\} = \lim_{N\to\infty} F_N(\beta+\delta) < \infty, \tag{D.5}$$

for some $\delta > 0$. To this end, note that if $0 < \beta < 1/\|W\|$, there exists $\delta > 0$ such that $\gamma := \beta + \delta < 1/\|W\|$. Now, using (7.3) and the fact $\sum_{i=1}^N \lambda_i(G_N) = 0$, we have

$$F_N(\gamma) \le \sum_{i=1}^{N} \left\{-\frac{1}{2}\log\left(1 - \frac{\gamma\lambda_i(W_N)}{N}\right) - \frac{\gamma}{2} \cdot \frac{\lambda_i(W_N)}{N}\right\}. \tag{D.6}$$

Since $W_N \Rightarrow W$ in the cut metric, $\lim_{N\to\infty} \frac{\gamma\lambda_i(W_N)}{N} = \gamma\|W\| < 1$, and so there exists $\varepsilon > 0$ such that for all N large enough $\frac{\gamma\lambda_i(W_N)}{N} \le 1 - \varepsilon$. For $x \le 1 - \varepsilon$ there exists $M = M(\varepsilon)$ such that $-\log(1-x) - x \le Mx^2$. Using this the RHS of (D.6) can be bounded by $\frac{M\gamma^2 \sum_{i=1}^N \lambda_i^2(W_N)}{N^2}$ which converges to $M\gamma^2\|W\|_2^2 = M\gamma^2$, as $N \to \infty$. This proves (D.5) and completes the proof of the proposition.

Acknowledgements

The authors thank Sourav Chatterjee, Persi Diaconis, and Qingyuan Zhao for helpful discussions. The authors also thank the anonymous referees for their careful comments which improved the presentation of the paper.

References

[1] Anandkumar, A., Tan, V.Y.F., Huang, F. and Willsky, A.S. (2012). High-dimensional structure estimation in Ising models: Local separation criterion. *Ann. Statist.* **40** 1346–1375. MR3015028

[2] Banerjee, S., Carlin, B.P. and Gelfand, A.E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: Chapman & Hall.

[3] Basak, A. and Mukherjee, S. (2016). Universality of the mean-field for the Potts model. *Probab. Theory Related Fields.* To appear. Available at arXiv:1508.03949.

[4] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **36** 192–236. MR0373208

[5] Besag, J. (1975). Statistical analysis of non-lattice data. *Statistician* **24** 179–195.

[6] Bhattacharya, B.B., Diaconis, P. and Mukherjee, S. (2016). Universal Poisson and normal limit theorems in graph coloring problems with connections to extremal combinatorics. *Ann. Appl. Probab.* To appear. Available at arXiv:1310.2336.

[7] Borgs, C., Chayes, J.T., Lovász, L., Sós, V.T. and Vesztergombi, K. (2008). Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.* **219** 1801–1851. MR2455626

[8] Borgs, C., Chayes, J.T., Lovász, L., Sós, V.T. and Vesztergombi, K. (2012). Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math.* (2) **176** 151–219. MR2925382

[9] Bresler, G. (2015). Efficiently learning Ising models on arbitrary graphs [extended abstract]. In *STOC'*15*—Proceedings of the* 2015 *ACM Symposium on Theory of Computing* 771–782. New York: ACM. MR3388257

[10] Chatterjee, S. (2007). Estimation in spin glasses: A first step. *Ann. Statist.* **35** 1931–1946. MR2363958

[11] Comets, F. (1992). On consistency of a class of estimators for exponential families of Markov random fields on the lattice. *Ann. Statist.* **20** 455–468. MR1150354

[12] Comets, F. and Gidas, B. (1991). Asymptotics of maximum likelihood estimators for the Curie–Weiss model. *Ann. Statist.* **19** 557–578. MR1105836

[13] Dembo, A. and Montanari, A. (2010). Gibbs measures and phase transitions on sparse random graphs. *Braz. J. Probab. Stat.* **24** 137–211. MR2643563

[14] Dembo, A. and Montanari, A. (2010). Ising models on locally tree-like graphs. *Ann. Appl. Probab.* **20** 565–592. MR2650042

[15] Dembo, A., Montanari, A. and Sun, N. (2013). Factor models on locally tree-like graphs. *Ann. Probab.* **41** 4162–4213. MR3161472

[16] Diaconis, P. (1996). The cutoff phenomenon in finite Markov chains. *Proc. Natl. Acad. Sci. USA* **93** 1659–1664. MR1374011

[17] Durrett, R. (2010). *Probability: Theory and Examples*, 4th ed. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge Univ. Press. MR2722836

[18] Geyer, C.J. and Thompson, E.A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **54** 657–699. MR1185217

[19] Gidas, B. (1988). Consistency of maximum likelihood and pseudolikelihood estimators for Gibbs distributions. In *Stochastic Differential Systems*, *Stochastic Control Theory and Applications* (*Minneapolis*, *MN*, 1986) (M. Fleming and P.-L. Lions, eds.). *IMA Vol. Math. Appl.* **10** 129–145. New York: Springer. MR0934721

[20] Green, P.J. and Richardson, S. (2002). Hidden Markov models and disease mapping. *J. Amer. Statist. Assoc.* **97** 1055–1070. MR1951259

[21] Guyon, X. and Künsch, H.R. (1992). Asymptotic comparison of estimators in the Ising model. In *Stochastic Models*, *Statistical Methods*, *and Algorithms in Image Analysis* (*Rome*, 1990). *Lecture Notes in Statist.* **74** 177–198. Berlin: Springer. MR1188486

[22] Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* **79** 2554–2558. MR0652033

[23] Ising, E. (1925). Beitrag zur theorie der ferromagnetismus. *Zeitschrift Für Physik* **31** 253–258.

[24] Jensen, J.L. and Møller, J. (1991). Pseudolikelihood for exponential family models of spatial point processes. *Ann. Appl. Probab.* **1** 445–461. MR1111528
[25] Krivelevich, M. and Sudakov, B. (2003). The largest eigenvalue of sparse random graphs. *Combin. Probab. Comput.* **12** 61–72. MR1967486
[26] Levin, D.A., Luczak, M.J. and Peres, Y. (2010). Glauber dynamics for the mean-field Ising model: Cut-off, critical power law, and metastability. *Probab. Theory Related Fields* **146** 223–265. MR2550363
[27] Lovász, L. (2012). *Large Networks and Graph Limits*. *American Mathematical Society Colloquium Publications* **60**. Providence, RI: Amer. Math. Soc. MR3012035
[28] Lubetzky, E. and Sly, A. (2013). Cutoff for the Ising model on the lattice. *Invent. Math.* **191** 719–755. MR3020173
[29] Lubetzky, E. and Sly, A. (2016). Information percolation and cutoff for the stochastic Ising model. *J. Amer. Math. Soc.* **29** 729–774. MR3486171
[30] Panchenko, D. (2013). *The Sherrington–Kirkpatrick Model*. *Springer Monographs in Mathematics*. New York: Springer. MR3052333
[31] Pickard, D.K. (1987). Inference for discrete Markov fields: The simplest nontrivial case. *J. Amer. Statist. Assoc.* **82** 90–96. MR0883337
[32] Ravikumar, P., Wainwright, M.J. and Lafferty, J.D. (2010). High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *Ann. Statist.* **38** 1287–1319. MR2662343
[33] Talagrand, M. (2011). *Mean Field Models for Spin Glasses*. *Volume I*: *Basic Examples*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete*. 3. *Folge*. *A Series of Modern Surveys in Mathematics* **54**. Berlin: Springer. MR2731561
[34] Talagrand, M. (2011). *Mean Field Models for Spin Glasses*: *Advanced Replica-Symmetry and Low Temperature*. *Volume II*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete*. 3. *Folge*. *A Series of Modern Surveys in Mathematics* **55**. Heidelberg: Springer. MR3024566
[35] Xue, L., Zou, H. and Cai, T. (2012). Nonconcave penalized composite conditional likelihood estimation of sparse Ising models. *Ann. Statist.* **40** 1403–1429. MR3015030

Received November 2015 and revised July 2016