GLOBAL TESTING AGAINST SPARSE ALTERNATIVES UNDER ISING MODELS

BY RAJARSHI MUKHERJEE[*], SUMIT MUKHERJEE[†,1] AND MING YUAN[†,2]

University of California, Berkeley[*] and Columbia University[†]

In this paper, we study the effect of dependence on detecting sparse signals. In particular, we focus on global testing against sparse alternatives for the means of binary outcomes following an Ising model, and establish how the interplay between the strength and sparsity of a signal determines its detectability under various notions of dependence. The profound impact of dependence is best illustrated under the Curie–Weiss model where we observe the effect of a "thermodynamic" phase transition. In particular, the critical state exhibits a subtle "blessing of dependence" phenomenon in that one can detect much weaker signals at criticality than otherwise. Furthermore, we develop a testing procedure that is broadly applicable to account for dependence and show that it is asymptotically minimax optimal under fairly general regularity conditions.

1. Introduction. Motivated by applications in a multitude of scientific disciplines, statistical analysis of "sparse signals" in a high dimensional setting, be it large-scale multiple testing or screening for relevant features, has drawn considerable attention in recent years. For more discussions on sparse signal detection type problems see, for example, Addario-Berry et al. (2010), Arias-Castro, Donoho and Huo (2005), Arias-Castro and Wang (2015), Arias-Castro et al. (2008), Cai and Yuan (2014), Donoho and Jin (2004), Hall and Jin (2010), Ingster, Tsybakov and Verzelen (2010), Mukherjee, Pillai and Lin (2015), and references therein. A critical assumption often made in these studies is that the observations are independent. Recognizing the potential limitation of this assumption, several recent attempts have been made to understand the implications of dependence in both theory and methodology; see, for example, Arias-Castro, Candès and Plan (2011), Hall and Jin (2008, 2010), Jin and Ke (2016), Wu et al. (2014). These earlier efforts, setting in the context of Gaussian sequence or regression models, show that it is important to account for dependence among observations, and under suitable conditions, doing so appropriately may lead to tests that are as powerful as if the observations were independent. However, it remains largely unknown how the dependence may affect our ability to detect sparse signals beyond Gaussian models. The main goal

of the present work is to fill in this void. In particular, we investigate the effect of dependence on detection of sparse signals for Bernoulli sequences, a class of problems arising naturally in many genomics applications [e.g., Mukherjee, Pillai and Lin (2015)].

Let $\mathbf{X} = (X_1, \ldots, X_n)^\top \in \{\pm 1\}^n$ be a random vector such that $\mathbb{P}(X_i = +1) = p_i$. In a canonical multiple testing setup, we want to test collectively that $H_0 : p_i = 1/2, i = 1, 2, \ldots, n$. Of particular interest here is the setting when X_i's may be dependent. A general framework to capture the dependence among a sequence of binary random variables is the so-called Ising models, which have been studied extensively in the literature [Ellis and Newman (1978), Ising (1925), Majewski, Li and Ott (2001), Onsager (1944), Stauffer (2008), Mézard and Montanari (2009)]. An Ising model specifies the joint distribution of $\mathbf{X}$ as

$$(1) \qquad \mathbb{P}_{\mathbf{Q}, \boldsymbol{\mu}}(\mathbf{X} = \mathbf{x}) := \frac{1}{Z(\mathbf{Q}, \boldsymbol{\mu})} \exp\left(\frac{1}{2} \mathbf{x}^\top \mathbf{Q} \mathbf{x} + \boldsymbol{\mu}^\top \mathbf{x}\right), \qquad \forall \mathbf{x} \in \{\pm 1\}^n,$$

where $\mathbf{Q}$ is an $n \times n$ symmetric and hollow matrix, $\boldsymbol{\mu} := (\mu_1, \ldots, \mu_n)^\top \in \mathbb{R}^n$ and $Z(\mathbf{Q}, \boldsymbol{\mu})$ is a normalizing constant. Throughout the rest of the paper, the expectation operator corresponding to (1) will be analogously denoted by $\mathbb{E}_{\mathbf{Q}, \boldsymbol{\mu}}$. It is clear that the matrix $\mathbf{Q}$ characterizes the dependence among the coordinates of $\mathbf{X}$, and X_i's are independent if $\mathbf{Q} = \mathbf{0}$. Under model (1), the relevant null hypothesis can be expressed as $\boldsymbol{\mu} = \mathbf{0}$. More specifically, we are interested in testing it against a sparse alternative:

$$(2) \qquad\qquad H_0 : \boldsymbol{\mu} = \mathbf{0} \quad \text{vs} \quad H_1 : \boldsymbol{\mu} \in \Xi(s, B),$$

where

$$\Xi(s, B) := \left\{\boldsymbol{\mu} \in \mathbb{R}^n : |\operatorname{supp}(\boldsymbol{\mu})| = s, \text{ and } \min_{i \in \operatorname{supp}(\boldsymbol{\mu})} \mu_i \geq B > 0\right\}$$

and

$$\operatorname{supp}(\boldsymbol{\mu}) := \{1 \leq i \leq n : \mu_i \neq 0\}.$$

Our goal here is to study the impact of $\mathbf{Q}$ in doing so.

To this end, we adopt an asymptotic minimax framework that can be traced back at least to Burnašev (1979), Ingster (1994, 1998). See Ingster and Suslina (2003) for further discussions. Let a statistical test for H_0 versus H_1 be a measurable $\{0, 1\}$ valued function of the data $\mathbf{X}$, with 1 indicating rejecting the null hypothesis H_0 and 0 otherwise. The worst case risk of a test $T : \{\pm 1\}^n \to \{0, 1\}$ can be given by

$$(3) \qquad \operatorname{Risk}(T, \Xi(s, B), \mathbf{Q}) := \mathbb{P}_{\mathbf{Q}, \mathbf{0}}(T(\mathbf{X}) = 1) + \sup_{\boldsymbol{\mu} \in \Xi(s, B)} \mathbb{P}_{\mathbf{Q}, \boldsymbol{\mu}}(T(\mathbf{X}) = 0),$$

where $\mathbb{P}_{\mathbf{Q}, \boldsymbol{\mu}}$ denotes the probability measure as specified by (1). We say that a sequence of tests T indexed by n corresponding to a sequence of model-problem

pair (1) and (3), to be asymptotically powerful (resp., asymptotically not powerful) against $\Xi(s, B)$ if

(4)
$$\limsup_{n \to \infty} \mathrm{Risk}\big(T, \Xi(s, B), \mathbf{Q}\big)$$
$$= 0 \qquad \Big(\text{resp.,} \liminf_{n \to \infty} \mathrm{Risk}\big(T, \Xi(s, B), \mathbf{Q}\big) > 0\Big).$$

The goal of the current paper is to characterize how the sparsity s and strength B of the signal ($\boldsymbol{\mu}$) jointly determine if there is a powerful test, and how the behavior changes with $\mathbf{Q}$. In particular,

- for a general class of Ising models, we provide tests for detecting arbitrary sparse signals and show that they are asymptotically rate optimal for Ising models on regular graphs in the high temperature regime;
- for Ising models on the cycle graph, we establish rate optimal results for all regimes of temperature, and show that the detection thresholds are the same as the independent case;
- for the Curie–Weiss model [Kac (1969), Nishimori (2001)], we provide sharp asymptotic detection thresholds for detecting arbitrarily sparse signals, which reveal an interesting phenomenon at the thermodynamic phase transition point of a Curie–Weiss magnet.

Our tools for analyzing the rate optimal tests depend on the method of exchangeable pairs [Chatterjee (2007b)], which might be of independent interest.

The rest of the paper is organized as follows. In Section 2, we study in detail the optimal detection thresholds for the Curie–Weiss model and explore the effects of the presence of a "thermodynamic phase transition" in the model. Section 3 is devoted to developing and analyzing testing procedures in the context of more general Ising models where we also show that under some conditions on $\mathbf{Q}$, the proposed testing procedure is indeed asymptotically optimal. Finally, we conclude with some discussions in Section 5. The proof of the main results is relegated to Section 6. The proof of additional technical arguments can be found in Mukherjee, Mukherjee and Yuan (2018).

2. Sparse testing under Curie–Weiss model. In most statistical problems, dependence reduces effective sample size and, therefore, makes inference harder. This, however, turns out not necessarily to be the case in our setting. The effect of dependence on sparse testing under Ising model is more profound. To make this more clear, we first consider one of the most popular examples of Ising models, namely the Curie–Weiss model. In the Curie–Weiss model,

(5)
$$\mathbb{P}_{\theta, \boldsymbol{\mu}}(\mathbf{X} = \mathbf{x}) := \frac{1}{Z(\theta, \boldsymbol{\mu})} \exp\bigg(\frac{\theta}{n} \sum_{1 \le i < j \le n} x_i x_j + \sum_{i=1}^{n} \mu_i x_i\bigg),$$

where in this section, with slight abuse of notation, we rename $\mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}$, $\mathbb{E}_{\mathbf{Q},\boldsymbol{\mu}}$, and $Z(\mathbf{Q},\boldsymbol{\mu})$ by $\mathbb{P}_{\theta,\boldsymbol{\mu}}$, $\mathbb{E}_{\theta,\boldsymbol{\mu}}$ and $Z(\theta,\boldsymbol{\mu})$, respectively, for brevity. The Curie–Weiss model is deceivingly simple and one of the classical examples that exhibit the so-called "thermodynamic" phase transition at $\theta = 1$; see, for example, Kac (1969), Nishimori (2001). It turns out that such a phase transition directly impacts how well a sparse signal can be detected. Following the convention, we shall refer to $\theta = 1$ as the critical state, $\theta > 1$ the low temperature states and $\theta < 1$ the high temperature states.

2.1. *High temperature states.* We consider first the high temperature case, that is, $0 \le \theta < 1$. It is instructive to begin with the case when $\theta = 0$, that is, $X_1, \ldots, X_n$ are independent Bernoulli random variables. By the central limit theorem,

$$\sqrt{n}\left(\bar{X} - \frac{1}{n}\sum_{i=1}^{n}\tanh(\mu_i)\right) \to_d N\left(0, \frac{1}{n}\sum_{i=1}^{n}\operatorname{sech}^2(\mu_i)\right),$$

where

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n}X_i.$$

In particular, under the null hypothesis,

$$\sqrt{n}\bar{X} \to_d N(0,1).$$

This immediately suggests a test that rejects H_0 if and only $\sqrt{n}\bar{X} \ge L_n$ for a diverging sequence $L_n = o(n^{-1/2}s\tanh(B))$ is asymptotic powerful, in the sense of (4), for testing (2) whenever $s\tanh(B) \gg n^{1/2}$. This turns out to be the best one can do in that there is no powerful test for testing (2) if $s\tanh(B) = O(n^{1/2})$; see, for example, Mukherjee, Pillai and Lin (2015). An immediate question of interest is what happens if there is dependence, that is $0 < \theta < 1$. This is answered by Theorem 1 below.

THEOREM 1. *Consider testing* (2) *based on* $\mathbf{X}$ *following the Curie–Weiss model* (5) *with* $0 \le \theta < 1$. *If* $s\tanh(B) \gg n^{1/2}$, *then the test that rejects H_0 if and only if $\sqrt{n}\bar{X} \ge L_n$ for a diverging L_n such that $L_n = o(n^{-1/2}s\tanh(B))$ is asymptotically powerful for* (2). *Conversely, if $s\tanh(B) = O(n^{1/2})$, then there is no asymptotically powerful test for* (2).

Theorem 1 shows that, under high temperature states, the sparse testing problem (2) behaves similarly to the independent case. Not only the detection limit remains the same, but also it can be attained even if one neglects the dependence while constructing the test.

2.2. *Low temperature states.* Now consider the low temperature case when $\theta > 1$. The naïve test that rejects H_0 whenever $\sqrt{n}\bar{X} \geq L_n$ is no longer asymptotically powerful in these situations. In particular, $\bar{X}$ concentrates around the roots of $x = \tanh(\theta x)$ and $\sqrt{n}\bar{X}$ is larger than any $L_n = O(n^{1/2})$ with a nonvanishing probability, which results in an asymptotically strictly positive probability of Type I error for a test based on rejecting H_0 if $\sqrt{n}\bar{X} \geq L_n$.

To overcome this difficulty, we shall consider a slightly modified test statistic:

$$\tilde{X} = \frac{1}{n}\sum_{i=1}^{n}\left[X_i - \tanh\left(\frac{\theta}{n}\sum_{j\neq i}X_j\right)\right].$$

Note that

$$\tanh\left(\frac{\theta}{n}\sum_{j\neq i}X_j\right) = \mathbb{E}_{\theta,\mathbf{0}}(X_i|X_j : j \neq i)$$

is the conditional mean of X_i given $\{X_j : j \neq i\}$ under the Curie–Weiss model with $\boldsymbol{\mu} = \mathbf{0}$. In other words, we average after centering each observation X_i by its conditional mean, instead of the unconditional mean, under H_0. The idea of centering by the conditional mean is similar in spirit to the pseudo-likelihood estimate of Besag (1974, 1975); see also Bhattacharya and Mukherjee (2018), Chatterjee (2007a), Guyon (1995).

We can then proceed to reject H_0 if and only if $\sqrt{n}\tilde{X} \geq L_n$. The next theorem shows that this procedure is indeed optimal with the appropriate choice of L_n.

THEOREM 2. *Consider testing* (2) *based on* **X** *following the Curie–Weiss model* (5) *with* $\theta > 1$. *If* $s\tanh(B) \gg n^{1/2}$, *then the test that rejects* H_0 *if and only if* $\sqrt{n}\tilde{X} \geq L_n$ *for a diverging* L_n *such that* $L_n = o(n^{-1/2}s\tanh(B))$ *is asymptotically powerful for* (2). *Conversely, if* $s\tanh(B) = O(n^{1/2})$, *then there is no asymptotically powerful test for* (2).

Theorem 2 shows that the detection limits for low temperature states remain the same as that for high temperature states, but a different test is required to achieve it.

2.3. *Critical state.* The situation however changes at the critical state $\theta = 1$, where a much weaker signal could still be detected. This is made precise by our next theorem, where we show that detection thresholds, in terms of $s\tanh(B)$, for the corresponding Curie–Weiss model at criticality scales as $n^{-3/4}$ instead of $n^{-1/2}$ as in either low or high temperature states. Moreover, it is attainable by the test that rejects H_0 whenever $n^{1/4}\bar{X} \geq L_n$ for appropriately chosen L_n.

THEOREM 3. *Consider testing* (2) *based on* **X** *following the Curie–Weiss model* (5) *with* $\theta = 1$. *If* $s\tanh(B) \gg n^{1/4}$, *then a test that rejects* H_0 *if and only if*

$n^{1/4}\bar{X} \geq L_n$ *for a suitably chosen diverging sequence* L_n, *is asymptotically powerful for* (2). *Conversely, if* $s\tanh(B) = O(n^{1/4})$, *then there is no asymptotically powerful test for* (2).

A few comments are in order about the implications of Theorem 3 in contrast to Theorems 1 and 2. Previously, the distributional limits for the total magnetization $\sum_{i=1}^{n} X_i$ has been characterized in all the three regimes of high ($\theta < 1$), low ($\theta > 1$) and critical ($\theta = 1$) temperatures [Ellis and Newman (1978)] when $\boldsymbol{\mu} = \mathbf{0}$. More specifically, they show that

$$\sqrt{n}\bar{X} \xrightarrow{d} N\left(0, \frac{1}{1-\theta}\right) \qquad \text{if } \theta < 1,$$

$$n^{1/4}\bar{X} \xrightarrow{d} W \qquad \text{if } \theta = 1,$$

$$(\sqrt{n}(\bar{X} - m(\theta))|\bar{X} > 0) \xrightarrow{d} N\left(0, \frac{1}{1 - \theta(1 - m(\theta)^2)}\right) \qquad \text{if } \theta > 1,$$

where W is a random variable on $\mathbb{R}$ with density proportional to $e^{-x^4/12}$ with respect to Lebesgue measure, and $m(\theta)$ is the unique positive root of the equation $z = \tanh(\theta z)$ for $\theta > 1$. A central quantity of their analysis is studying the roots of this equation. Our results demonstrate parallel behavior in terms of detection of sparse external magnetization $\boldsymbol{\mu}$. In particular, if the vector $\boldsymbol{\mu} = (B, \ldots, B, 0, 0, \ldots, 0)$ with the number of nonzero components equal to s, we obtain the fixed-point equation $z = p\tanh(\theta z + B) + (1 - p)\tanh(\theta z)$, where $p := s/n$. One can get an informal explanation of the detection boundary for the various cases from this fixed-point equation. As, for example, in the critical case when $\theta = 1$, we get the equation

$$z = p\tanh(z + B) + (1 - p)\tanh(z) \quad \Rightarrow$$

$$z - \tanh(z) = p[\tanh(z + B) - \tanh(z)].$$

The LHS of the second equality is of order z^3 for $z \approx 0$, and the RHS is of order $p\tanh(B)$. This gives the relation $z^3 \sim (p\tanh B)$, which gives the asymptotic order of the mean of $\bar{X}$ under the alternative as $z \sim (p\tanh B)^{1/3}$. Since under H_0 the fluctuation of $\bar{X}$ is $n^{-1/4}$, for successful detection we need $n^{-1/4} \ll (p\tanh B)^{1/3}$, which is equivalent to $s\tanh(B) \gg n^{1/4}$ on recalling that $s = np$. Similar heuristic justification holds for other values of θ as well.

Interestingly, both below and above phase transition the detection problem considered here behaves similar to that in a disordered system of i.i.d. random variables, in spite having different asymptotic behavior of the total magnetization in the two regimes. However, an interesting phenomenon continues to emerge at $\theta = 1$ where one can detect a much smaller signal or external magnetization [magnitude of $s\tanh(B)$]. In particular, according to Theorem 1 and Theorem 2, no

signal is detectable of sparsity $s \ll \sqrt{n}$, when $\theta \neq 1$. In contrast, Theorem 3 establishes signals satisfying $s \tanh(B) \gg n^{1/4}$ is detectable for $n^{1/4} \lesssim s \ll \sqrt{n}$, where $a_n \lesssim b_n$ means $a_n = O(b_n)$. As mentioned before, it is well known that the Curie–Weiss model undergoes a phase transition at $\theta = 1$. Theorem 3 provides a rigorous verification of the fact that the phase transition point $\theta = 1$ can reflect itself in terms of detection problems, even though θ is a nuisance parameter. In particular, the detection is easier than at noncriticality. This is interesting in its own right since the concentration of $\bar{X}$ under the null hypothesis is weaker than that for $\theta < 1$ [Chatterjee and Dey (2010)] and yet a smaller amount of signal enables us to break free of the null fluctuations. We shall make this phenomenon more transparent in the proof of the theorem.

3. Sparse testing under general Ising models. As we can see from the previous section, the effect of dependence on sparse testing under Ising models is more subtle than the Gaussian case. It is of interest to investigate to what extent the behavior we observed for the Curie–Weiss model applies to the more general Ising model, and whether there is a more broadly applicable strategy to deal with the general dependence structure. To this end, we further explore the idea of centering by the conditional mean we employed to treat low temperature states under the Curie–Weiss model, and argue that it indeed works under fairly general situations.

3.1. *Conditional mean centered tests.* Note that under the Ising model (1),

$$\mathbb{E}_{\mathbf{Q},\mathbf{0}}(X_i | X_j : j \neq i) = \tanh\big(m_i(\mathbf{X})\big),$$

where

$$m_i(\mathbf{X}) = \sum_{j=1}^{n} Q_{ij} X_j.$$

Following the same idea as before, we shall consider a test statistic

$$\tilde{X} = \frac{1}{n} \sum_{i=1}^{n} \big[X_i - \tanh\big(m_i(\mathbf{X})\big)\big],$$

and proceed to reject H_0 if and only if $\sqrt{n}\tilde{X} \geq L_n$. The following result shows that the same detection limit $s \tanh(B) \gg n^{1/2}$ can be achieved by this test as long as $\|\mathbf{Q}\|_{\ell_\infty \to \ell_\infty} = O_p(1)$, where $\|\mathbf{Q}\|_{\ell_p \to \ell_q} = \max_{\|x\|_{\ell_p} \leq 1} \|\mathbf{Q}x\|_{\ell_q}$ for $p, q > 0$.

THEOREM 4. *Let* $\mathbf{X}$ *follow an Ising model* (1) *with* $\mathbf{Q}$ *such that* $\|\mathbf{Q}\|_{\ell_\infty \to \ell_\infty} = O(1)$. *Consider testing hypotheses about* $\boldsymbol{\mu}$ *as described by* (2). *If* $s \tanh(B) \gg n^{1/2}$, *then the test that rejects* H_0 *if and only if* $\sqrt{n}\tilde{X} \geq L_n$ *for any* $L_n \to \infty$ *such that* $L_n = o(n^{-1/2}s \tanh(B))$ *is asymptotically powerful.*

The condition $\|\mathbf{Q}\|_{\ell_\infty \to \ell_\infty} = O(1)$ is a regularity condition which holds for many common examples of the Ising model in the literature. In particular, $\mathbf{Q}$ often times can be associated with a certain graph $\mathcal{G} = (V, E)$ with vertex set $V = [n] := \{1, \ldots, n\}$ and edge set $E \subseteq [n] \times [n]$ so that $\mathbf{Q} = (n\theta)G/(2|E|)$, where G is the adjacency matrix for $\mathcal{G}$, $|E|$ is the cardinality of E, and $\theta \in \mathbb{R}$ is a parameter independent of n deciding the degree of dependence in the spin-system. Below we provide several more specific examples that are commonly studied in the literature.

Dense graphs. Recall that

$$\|\mathbf{Q}\|_{\ell_\infty \to \ell_\infty} = \max_{1 \le i \le n} \sum_{j=1}^{n} |Q_{ij}| \le \frac{n^2 |\theta|}{2|E|}.$$

If the dependence structure is guided by densely labeled graphs so that $|E| = \Theta(n^2)$, then $\|\mathbf{Q}\|_{\ell_\infty \to \ell_\infty} = O(1)$.

Regular graphs. When the dependence structure is guided by a regular graph of degree d_n, we can write $\mathbf{Q} = \theta G/d_n$. Therefore,

$$\|\mathbf{Q}\|_{\ell_\infty \to \ell_\infty} = \max_{1 \le i \le n} \sum_{j=1}^{n} |Q_{ij}| = \frac{|\theta|}{d_n} \cdot d_n = |\theta|,$$

and again obeying the condition $\|\mathbf{Q}\|_{\ell_\infty \to \ell_\infty} = O(1)$.

Erdős–Rényi graphs. Another example is the Erdős–Rényi graph where an edge between each pair of nodes is present with probability p_n independent of each other. It is not hard to derive from Chernoff bound and union bounds that the maximum degree $d_{\max}$ and the totally number of edges $|E|$ of an Erdős–Rényi graph satisfy with high probability:

$$d_{\max} \le np_n(1 + \delta), \quad \text{and} \quad |E| \ge \frac{n(n-1)}{2} p_n \cdot (1 - \delta)$$

for any $\delta \in (0, 1)$, provided that $np_n \gg \log n$. This immediately implies that $\|\mathbf{Q}\|_{\ell_\infty \to \ell_\infty} = O_p(1)$.

In other words, the detection limit established in Theorem 4 applies to all of these types of Ising models. In particular, it suggests that, under the Curie–Weiss model, the $\sqrt{n}\tilde{X}$ based test can detect sparse external magnetization $\boldsymbol{\mu} \in \Xi(s, B)$ if $s \tanh(B) \gg n^{1/2}$, for any $\theta \in \mathbb{R}$, which, in the light of Theorems 1 and 2, is optimal in both high and low temperature states.

3.2. *Optimality.* The detection limit presented in Theorem 4 matches those obtained for independent Bernoulli sequence model. It is of interest to understand to what extent the upper bounds in Theorem 4 are sharp. The answer to this question might be subtle. In particular, as we see in the Curie–Weiss case, the optimal rates of detection thresholds depend on the presence of thermodynamic phase transition in the null model. To further illustrate the role of criticality, we now consider an example of the Ising model without phase transition and the corresponding behavior of the detection problem (2) in that case. Let

$$\mathbf{Q}_{i,j} = \frac{\theta}{2}\mathbb{I}\{|i - j| = 1 \mod n\}$$

so that the corresponding Ising model can be identified with a cycle graph of length n. Our next result shows that the detection threshold remains the same for any θ, and is the same as the independent case, that is, $\theta = 0$.

THEOREM 5.   *Suppose* $\mathbf{X} \sim \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}$, *where* $\mathbf{Q}$ *is the scaled adjacency matrix of the cycle graph of length* n, *that is,* $\mathbf{Q}_{i,j} = \frac{\theta}{2}1\{|i - j| = 1 \mod n\}$ *for some* $\theta \in \mathbb{R}$. *If* $s \tanh(B) \leq C\sqrt{n}$ *for some* $C > 0$, *then no test is asymptotically powerful for the testing problem* (2).

In view of Theorem 4, if $s \tanh(B) \gg n^{1/2}$, then the test that rejects H_0 if and only if $\sqrt{n}\tilde{X} \geq L_n$ for any $L_n \to \infty$ such that $L_n = o(n^{-1/2}s\tanh(B))$ is asymptotically powerful for the testing problem (2). Together with Theorem 5, this shows that for the Ising model on the cycle graph of length n, which is a physical model without thermodynamic phase transitions, the detection thresholds mirror those obtained in independent Bernoulli sequence problems [Mukherjee, Pillai and Lin (2015)].

The difference between these results and those for the Curie–Weiss model demonstrates the difficulty of a unified and complete treatment to general Ising models. We offer here, instead, a partial answer and show that the test described earlier in the section (Theorem 4) is indeed optimal under fairly general weak dependence for reasonably regular graphs.

THEOREM 6.   *Suppose* $\mathbf{X} \sim \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}$ *as in* (1) *and consider testing hypotheses about* $\boldsymbol{\mu}$ *as described by* (2). *Assume* $\mathbf{Q}_{i,j} \geq 0$ *for all* (i, j) *such that* $\|\mathbf{Q}\|_{\ell_\infty \to \ell_\infty} \leq \rho < 1$ *for some constant* $\rho > 0$, $\|\mathbf{Q}\|_{\mathrm{F}}^2 = O(\sqrt{n})$, *and*

$$\left\|\mathbf{Q}\mathbf{1} - \frac{\mathbf{1}^\top\mathbf{Q}\mathbf{1}}{n}\mathbf{1}\right\|^2 = O(1).$$

If $s \tanh(B) \leq C\sqrt{n}$ *for some constant* $C > 0$, *then no test is asymptotically powerful for* (2).

Theorem 6 provides rate optimal lower bound to certain instances pertaining to Theorem 4. One essential feature of Theorem 6 is the implied impossibility result for the $s \ll \sqrt{n}$ regime. More precisely, irrespective of signal strength, no tests are asymptotically powerful when the number of signals drop below $\sqrt{n}$ in asymptotic order. This is once again in parallel to results in Mukherjee, Pillai and Lin (2015), and provides further evidence that low dependence/high temperature regimes (as encoded by $\|\mathbf{Q}\|_{\ell_\infty \to \ell_\infty} \le \rho < 1$) resemble independent Bernoulli ensembles. Theorem 6 immediately implies the optimality of the conditional mean centered tests for a couple of common examples.

High degree regular graphs. When the dependence structure is guided by a regular graph, that is, $\mathbf{Q} = \frac{\theta}{d_n} G$, it is clear that

$$\left\| \mathbf{Q}\mathbf{1} - \frac{\mathbf{1}^\top \mathbf{Q}\mathbf{1}}{n}\mathbf{1} \right\|^2 = 0.$$

If $0 \le \theta < 1$ and $d_n \gtrsim \sqrt{n}$, then one can easily verify the conditions of Theorem 6 since

$$\|\mathbf{Q}\|_{\ell_\infty \to \ell_\infty} = \theta < 1, \quad \text{and} \quad \|\mathbf{Q}\|_{\mathrm{F}}^2 = n\theta^2/d_n.$$

Dense Erdős–Rényi graphs. When the dependence structure is guided by a Erdős–Rényi graph on n vertices with parameter p_n, that is, $\mathbf{Q} = \theta/(np_n)G$ with $G_{i,j} \sim \text{Bernoulli}(p_n)$ independently for all $1 \le i < j \le n$, we can also verify that the conditions of Theorem 6 hold with probability tending to one if $0 \le \theta < 1$ and p_n bounded away from 0. As before, by Chernoff bounds, we can easily derive that with probability tending to one,

$$\|\mathbf{Q}\|_{\ell_\infty \to \ell_\infty} = \theta \frac{d_{\max}}{np_n} \le \frac{\theta(1+\delta)np_n}{np_n} = \theta(1+\delta)$$

and

$$\|\mathbf{Q}\|_{\mathrm{F}}^2 = \frac{\theta^2}{n^2 p_n^2} \sum_{1 \le i < j \le n} G_{i,j} \le \frac{\theta^2(1+\delta)n(n-1)p_n}{2n^2 p_n^2} \le \frac{\theta^2}{2p_n}(1+\delta),$$

for any $\delta > 0$. Finally, denote by d_i the degree of the ith node, then

$$\left\| \mathbf{Q}\mathbf{1} - \frac{\mathbf{1}^\top \mathbf{Q}\mathbf{1}}{n}\mathbf{1} \right\|^2 = \frac{\theta^2}{n^2 p_n^2} \sum_{i=1}^{n} \left(d_i - \frac{1}{n}\sum_{j=1}^{n} d_j \right)^2$$

$$\le \frac{\theta^2}{n^2 p_n^2} \sum_{i=1}^{n} (d_i - (n-1)p_n)^2$$

$$= O_p(1),$$

by Markov inequality and the fact that

$$\mathbb{E}\left[\sum_{i=1}^{n}(d_i - (n-1)p_n)^2\right] = n(n-1)p_n(1 - p_n).$$

4. Simulation results.   We now present results from a set of numerical experiments to further demonstrate the behavior of the various tests in finite samples. To fix ideas, we shall focus on the Curie–Weiss model since it exhibits the most interesting behavior in terms of the effect of thermodynamic phase transitions reflecting itself on the detection thresholds for the presence of sparse magnetization. In order to demonstrate the detection thresholds cleanly in the simulation, we parametrized sparsity s as $s = n^{1-\alpha}$ for $\alpha \in (0, 1)$. In this parametrization, the theoretical detection thresholds obtained for the Curie–Weiss model can be restated as follows. For $\theta \neq 1$, Theorem 1 and Theorem 2 suggest that the critical signal strength equals $\tanh(B) \sim n^{-(\frac{1}{2}-\alpha)}$. In particular, if $\tanh(B) = n^{-r}$, then no test is asymptotically powerful when $r > \frac{1}{2} - \alpha$; whereas the test based on conditionally centered magnetization is asymptotically powerful when $r < \frac{1}{2} - \alpha$. Moreover, for $\alpha > 1/2$, all tests are asymptotically powerless irrespective of the amount of signal strength. However, $\theta = 1$, Theorem 3 demonstrates that the critical signal strength equals $\tanh(B) \sim n^{-(\frac{3}{4}-\alpha)}$. In particular, if $\tanh(B) = n^{-r}$, then no test is asymptotically powerful when $r > \frac{3}{4} - \alpha$; whereas the test based on total magnetization is asymptotically powerful when $r < \frac{3}{4} - \alpha$. Moreover, for $\alpha > 3/4$, all tests are asymptotically powerless irrespective of the amount of the signal strength. The simulation presented below is designed to capture the different scenarios where nontrivial detection is possible, that is, $\alpha \leq 1/2$ for $\theta \neq 1$ and $\alpha \leq 3/4$ for $\theta = 1$.

We evaluated the power of the two tests, based on total magnetization and the conditionally centered magnetization, respectively, at the significance level of 5% and sample size $n = 1000$. We generated the test statistics 500 times under the null and take the 95%-quantile as the critical value. The power against different alternatives are then obtained empirically from 500 repeats each. The simulation from a Curie–Weiss model in the presence of magnetization is done using the Gaussian trick or the auxiliary variable approach as demonstrated by Lemma 3. In particular, for a given θ and $\boldsymbol{\mu}$ in the simulation parameter set, we generated a random variable Z (using package `rstan` in R) with density proportional to $f_{n,\boldsymbol{\mu}}(z) := \frac{n\theta z^2}{2} - \sum_{i=1}^{n} \log \cosh(\theta z + \mu_i)$. Next, given this realization of $Z = z$ we generated each component of $\mathbf{X} = (X_1, \ldots, X_n)$ independently taking values in ± 1 with $\mathbb{P}_{\theta,\boldsymbol{\mu}}(X_i = x_i) = \frac{e^{(\mu_i + z\theta)x_i}}{e^{\mu_i + z\theta} + e^{-\mu_i - z\theta}}$. Thereafter, Lemma 3 guarantees the joint distribution of $\mathbf{X}$ indeed follows a Curie–Weiss model with temperature parameter θ and magnetization $\boldsymbol{\mu}$. We believe that this method is much faster than the one-spin at a time Glauber dynamics which updates the whole chain $\mathbf{X}$ one location at

a time. We have absorbed all issues regarding mixing time in the simulation of Z, which being a one-dimensional continuous random variable behaves much better in simulation.

In Figure 1, we plot the power of both tests for $\theta = 0.5$ (high temperature, conditionally centered magnetization), $\theta = 1$ (critical temperature, total magnetization) and $\theta = 1.5$ (low temperature, conditionally centered magnetization). Each plot was produced by repeating the experiment for a range of equally spaced signal sparsity-strength pairs (α, r) with an increment of size 0.05. In addition, we plot in red the theoretical detection boundary given by $r = 1/2 - \alpha$ for noncritical temperature ($\theta \neq 1$) and $r = 3/4 - \alpha$ for critical temperature ($\theta = 1$). These simulation results agree very well with our theoretical development.

5. Discussions. In this paper, we study the asymptotic minimax rates of detection for arbitrary sparse signals in Ising models, considered as a framework to study dependency structures in binary outcomes. We show that the detection thresholds in Ising models might depend on the presence of a "thermodynamic" phase transition in the model. In the context of a Curie–Weiss–Ising model, the presence of such a phase transition results in substantial faster rates of detection of sparse signals at criticality. On the other hand, lack of such phase transitions, in the Ising model on the line graph, yields results parallel to those in independent Bernoulli sequence models, irrespective of the level of dependence. We further show that for Ising models defined on graphs enjoying certain degree of regularity, detection thresholds parallel those in independent Bernoulli sequence models in the low dependence/high temperature regime. It will be highly interesting to consider other kinds of graphs left out by Theorem 6 in the context of proving matching lower bounds to Theorem 4. This seems highly challenging and might depend heavily on the sharp asymptotic behavior of the partition function of more general Ising model under low-magnetization regimes. The issue of unknown dependency structure $\mathbf{Q}$, and especially the estimation of unknown temperature parameter θ for Ising models defined on given underlying graphs, is also subtle as shown in Bhattacharya and Mukherjee (2018). In particular, the rate of consistency of an estimator of θ under the null model (i.e., $\boldsymbol{\mu} = 0$) depends crucially on the position of θ with respect to the point of criticality and in particular high temperature regimes (i.e., low positive values of θ) may preclude the existence of any consistent estimator. The situation becomes even more complicated in presence of external magnetization (i.e., $\boldsymbol{\mu} \neq 0$). Finally, this paper opens up several interesting avenues of future research. In particular, investigating the effect of dependence on detection of segment type structured signals deserves special attention.

6. Proof of main results. In this section, we collect the proofs of our main results. It is convenient to first prove the general results, namely the upper bound given by Theorem 4 and lower bound by Theorem 6, and then consider the special cases of the Ising model on a cycle graph, and the Curie–Weiss model.
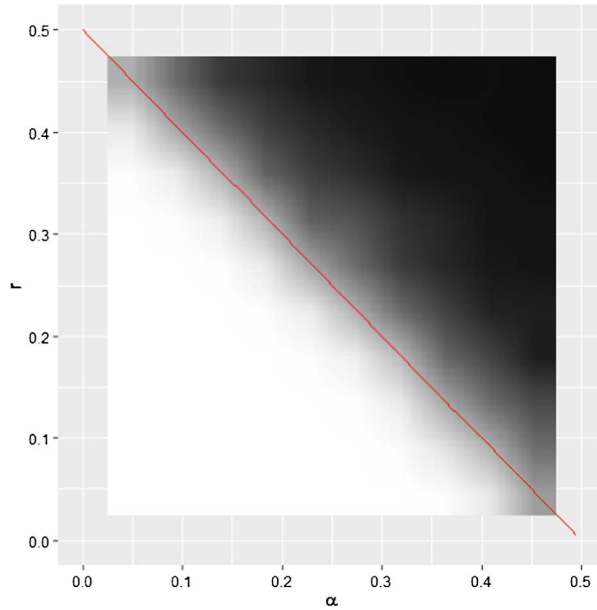
(a)
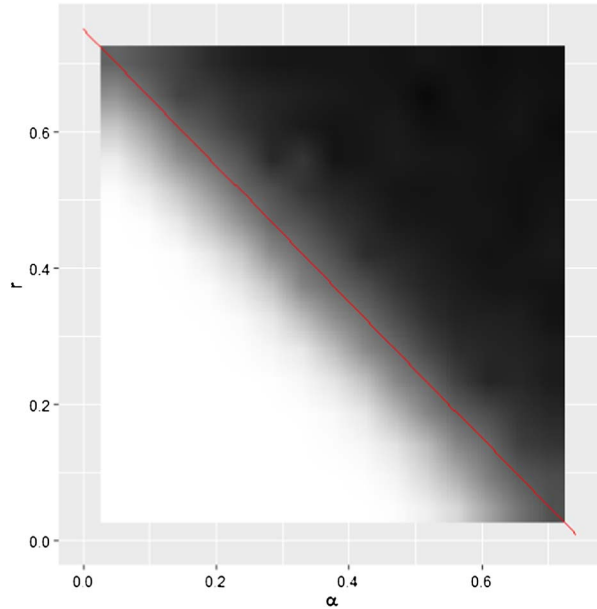
(b)

FIG. 1. *The power of testing procedures in the dense signal setup.* (a) *shows the power of the conditionally centered magnetization test for* $\theta = 0.5$, (b) *shows the power of the total magnetization test for* $\theta = 1$ *and* (c) *shows the power of the conditionally centered magnetization test for* $\theta = 1.5$. *The theoretical detection threshold is drawn in red.*
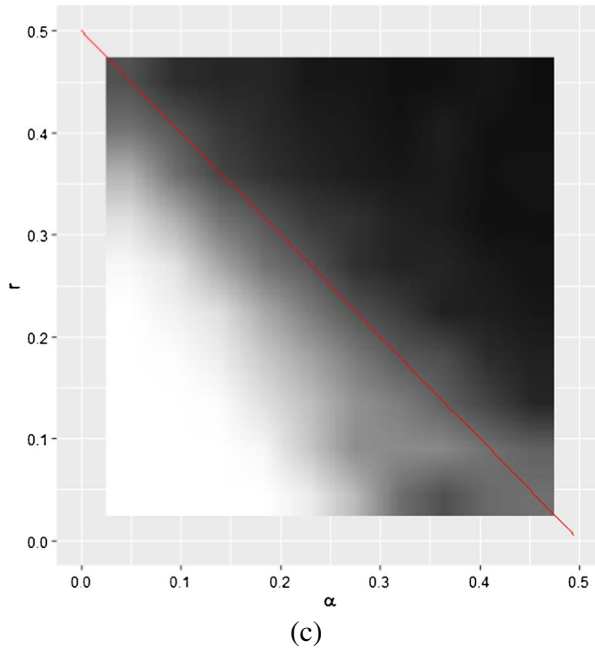
(c)

FIG. 1. (*Continued*).

6.1. *Proof of Theorem* 4. The key to the proof is the tail behavior of

$$f_{\mathbf{Q},\boldsymbol{\mu}}(\mathbf{X}) := \frac{1}{n}\sum_{i=1}^{n}[X_i - \mathbb{E}_{\mathbf{Q},\boldsymbol{\mu}}(X_i|X_j : j \neq i)]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left[X_i - \tanh\left(\sum_{j\neq i}\mathbf{Q}_{ij}X_j + \mu_j\right)\right],$$

where $\mathbb{E}_{\mathbf{Q},\boldsymbol{\mu}}$ means the expectation is taken with respect to the Ising model (1). In particular, we shall make use of the following concentration bound for $f_{\mathbf{Q},\boldsymbol{\mu}}(\mathbf{X})$.

LEMMA 1. *Let* $\mathbf{X}$ *be a random vector following the Ising model* (1). *Then for any* $t > 0$,

$$\mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}\big(|f_{\mathbf{Q},\boldsymbol{\mu}}(\mathbf{X})| \geq t\big) \leq 2\exp\left\{-\frac{nt^2}{4(1 + \|\mathbf{Q}\|_{\ell_\infty \to \ell_\infty})^2}\right\}.$$

Lemma 1 follows from a standard application of Stein's method for concentration inequalities [Chatterjee (2005, 2007b), Chatterjee and Dey (2010)]. We defer the detailed proof to the Appendix.

We are now in position to prove Theorem 4. We first consider the Type I error. By Lemma 1, there exists a constant $C > 0$ such that

$$\mathbb{P}_{\mathbf{Q},\mathbf{0}}(\sqrt{n}\tilde{X} \geq L_n) \leq 2\exp(-CL_n^2) \to 0.$$

It remains to consider the Type II error. Note that

$$\tilde{X} - f_{\mathbf{Q},\boldsymbol{\mu}}(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n}\left[\tanh\left(\sum_{j\neq i}\mathbf{Q}_{ij}X_j + \mu_i\right) - \tanh\left(\sum_{j\neq i}\mathbf{Q}_{ij}X_j\right)\right]$$

$$= \frac{1}{n}\sum_{i\in\text{supp}(\boldsymbol{\mu})}\left[\tanh\left(\sum_{j\neq i}\mathbf{Q}_{ij}X_j + \mu_i\right) - \tanh\left(\sum_{j\neq i}\mathbf{Q}_{ij}X_j\right)\right]$$

$$\geq \frac{1}{n}\sum_{i\in\text{supp}(\boldsymbol{\mu})}\left[\tanh\left(\sum_{j\neq i}\mathbf{Q}_{ij}X_j + B\right) - \tanh\left(\sum_{j\neq i}\mathbf{Q}_{ij}X_j\right)\right],$$

where the inequality follows from the monotonicity of tanh.

Observe that for any $x \in \mathbb{R}$ and $y > 0$,

(6)
$$\tanh(x + y) - \tanh(x) = \frac{[1 - \tanh^2(x)]\tanh(y)}{1 + \tanh(x)\tanh(y)}$$
$$\geq [1 - \tanh(x)]\tanh(y),$$

where the inequality follows from the fact that $|\tanh(x)| \leq 1$. Thus,

$$\tilde{X} - f_{\mathbf{Q},\boldsymbol{\mu}}(\mathbf{X}) \geq \frac{\tanh(B)}{n}\sum_{i\in\text{supp}(\boldsymbol{\mu})}\left[1 - \tanh\left(\sum_{j\neq i}\mathbf{Q}_{ij}X_j\right)\right].$$

Because

$$\sum_{j\neq i}\mathbf{Q}_{ij}X_j \leq \|\mathbf{Q}\|_{\ell_\infty\to\ell_\infty},$$

we get

$$\tilde{X} - f_{\mathbf{Q},\boldsymbol{\mu}}(\mathbf{X}) \geq \frac{s\tanh(B)}{n}[1 - \tanh(\|\mathbf{Q}\|_{\ell_\infty\to\ell_\infty})].$$

Therefore,

$$\sqrt{n}\tilde{X} - \sqrt{n}f_{\mathbf{Q},\boldsymbol{\mu}}(\mathbf{X}) \geq \frac{s\tanh(B)}{\sqrt{n}}[1 - \tanh(\|\mathbf{Q}\|_{\ell_\infty\to\ell_\infty})] \gg L_n.$$

This, together with another application of Lemma 1, yields the desired claim.

6.2. *Proof of Theorem 6.* The proof is somewhat lengthy and we break it into several steps.

6.2.1. *Reduction to magnetization.* We first show that a lower bound can be characterizing the behavior of $\bar{X}$ under the alternative. To this end, note that for any test T and a distribution π over $\Xi(s, B)$, we have

$$\text{Risk}(T, \Xi_{s,B}, \mathbf{Q}) = \mathbb{P}_{\mathbf{Q},\mathbf{0}}(T(\mathbf{X}) = 1) + \sup_{\boldsymbol{\mu} \in \Xi(s,B)} \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}(T(\mathbf{X}) = 0)$$

$$\geq \mathbb{P}_{\mathbf{Q},\mathbf{0}}(T(\mathbf{X}) = 1) + \int \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}(T(\mathbf{X}) = 0) \, d\pi(\boldsymbol{\mu}).$$

The rightmost-hand side is exactly the risk when testing H_0 against a simple alternative where $\mathbf{X}$ follows a mixture distribution:

$$\mathbb{P}_\pi(\mathbf{X} = \mathbf{x}) := \int \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}(\mathbf{X} = \mathbf{x}) \, d\pi(\boldsymbol{\mu}).$$

By the Neymann–Pearson lemma, this can be further lower bounded by

$$\text{Risk}(T, \Xi(s, B), \mathbf{Q}) \geq \mathbb{P}_{\mathbf{Q},\mathbf{0}}(L_\pi(\mathbf{X}) > 1) + \int \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}(L_\pi(\mathbf{X}) \leq 1) \, d\pi(\boldsymbol{\mu}),$$

where

$$L_\pi(\mathbf{X}) = \frac{\mathbb{P}_\pi(\mathbf{X})}{\mathbb{P}_{\mathbf{Q},\mathbf{0}}(\mathbf{X})}$$

is the likelihood ratio.

We can now choose a particular prior distribution π to make L_π a monotone function of $\bar{X}$. To this end, let π be supported over

$$\widetilde{\Xi}(s, B) = \{\boldsymbol{\mu} \in \{0, B\}^n : |\text{supp}(\boldsymbol{\mu})| = s\},$$

so that

$$\pi(\boldsymbol{\mu}) \propto Z(\mathbf{Q}, \boldsymbol{\mu}) \qquad \forall \boldsymbol{\mu} \in \widetilde{\Xi}(s, B).$$

It is not hard to derive that, with this particular choice,

$$L_\pi(\mathbf{X}) \propto \sum_{\boldsymbol{\mu} \in \widetilde{\Xi}(s,B)} \exp(\boldsymbol{\mu}^\top \mathbf{X}) = \mathbb{E}_S \exp\left(B \sum_{i \in S} X_i\right),$$

where $\mathbb{E}_S$ means expectation over S, a uniformly sampled subset of $[n]$ of size s. It is clear, by symmetry, that the rightmost-hand side is invariant to the permutation of the coordinates of $\mathbf{X}$. In addition, it is an increasing function of

$$|\{i \in [n] : X_i = 1\}| = \frac{1}{2}\left(n + \sum_{i=1}^n X_i\right),$$

and hence an increasing function of $\bar{X}$.

The observation that $L_\pi(\mathbf{X})$ is an increasing function of $\bar{X}$ implies that there exists a sequence κ_n such that

$$\text{Risk}(T, \Xi(s, B), \mathbf{Q}) \geq \mathbb{P}_{\mathbf{Q},\mathbf{0}}(L_\pi(\mathbf{X}) > 1) + \int \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}(L_\pi(\mathbf{X}) \leq 1) \, d\pi(\boldsymbol{\mu})$$

$$= \mathbb{P}_{\mathbf{Q},\mathbf{0}}\left(\sum_{i=1}^n X_i > \kappa_n\right) + \int \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}\left(\sum_{i=1}^n X_i \leq \kappa_n\right) d\pi(\boldsymbol{\mu})$$

$$\geq \mathbb{P}_{\mathbf{Q},\mathbf{0}}\left(\sum_{i=1}^n X_i > \kappa_n\right) + \inf_{\boldsymbol{\mu} \in \tilde{\Xi}(s, B)} \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}\left(\sum_{i=1}^n X_i \leq \kappa_n\right).$$

It now remains to study the behavior of $\bar{X}$.

In particular, it suffices to show that, for any fixed $x > 0$,

(7)
$$\liminf_{n \to \infty} \mathbb{P}_{\mathbf{Q},\mathbf{0}}\left\{\sum_{i=1}^n X_i > x\sqrt{n}\right\} > 0,$$

and for any $x_n \to \infty$,

(8)
$$\limsup_{n \to \infty} \sup_{\boldsymbol{\mu} \in \tilde{\Xi}(s, B)} \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}\left(\sum_{i=1}^n X_i > x_n\sqrt{n}\right) = 0.$$

Assuming (7) holds, then for any test T to be asymptotic powerful, we need $\kappa_n \gg \sqrt{n}$ to ensure that

$$\mathbb{P}_{\mathbf{Q},\mathbf{0}}\left\{\sum_{i=1}^n X_i > \kappa_n\right\} \to 0.$$

But, in the light of (8), this choice necessarily leads to

$$\inf_{\boldsymbol{\mu} \in \tilde{\Xi}(s, B)} \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}\left\{\sum_{i=1}^n X_i \leq \kappa_n\right\} \to 1,$$

so that

$$\text{Risk}(T, \Xi(s, B), \mathbf{Q}) \to 1.$$

In other words, there is no asymptotic powerful test if both (7) and (8) hold. We now proceed to prove them separately.

6.2.2. *Proof of* (8). Recall that $m_i(\mathbf{X}) = \sum_{j=1}^n \mathbf{Q}_{ij} X_j$ and assume $\boldsymbol{\mu} \in \tilde{\Xi}(s, B)$ with $s \tanh(B) \leq C\sqrt{n}$. Also let $\mathbf{r} = (r_1, \ldots, r_n)^\top$ where $\mathbf{r} = \mathbf{r}(\mathbf{Q}) := \mathbf{Q}\mathbf{1}$. We split the proof into two cases, depending on whether $B \leq 1$ or $B > 1$.

The case of $B \in [0, 1]$. Write

$$\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} [X_i - \tanh(m_i(\mathbf{X}) + \mu_i)]$$

$$+ \sum_{i=1}^{n} [\tanh(m_i(\mathbf{X}) + \mu_i) - \tanh(m_i(\mathbf{X}))]$$

$$+ \sum_{i=1}^{n} [\tanh(m_i(\mathbf{X})) - m_i(\mathbf{X})] + \sum_{i=1}^{n} m_i(\mathbf{X}).$$

Observe that

$$\sum_{i=1}^{n} m_i(\mathbf{X}) = \mathbf{1}^\top \mathbf{Q} \mathbf{X} = \sum_{i=1}^{n} r_i X_i = \rho^* \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} (r_i - \rho_*) X_i,$$

where $\rho_* = \frac{1}{n} \mathbf{1}^\top \mathbf{r} = \frac{1}{n} \mathbf{1}^\top \mathbf{Q} \mathbf{1}$. Thus,

$$(1 - \rho_*) \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} [X_i - \tanh(m_i(\mathbf{X}) + \mu_i)]$$

$$+ \sum_{i=1}^{n} [\tanh(m_i(\mathbf{X}) + \mu_i) - \tanh(m_i(\mathbf{X}))]$$

$$+ \sum_{i=1}^{n} [\tanh(m_i(\mathbf{X})) - m_i(\mathbf{X})] + \sum_{i=1}^{n} (r_i - \rho_*) X_i$$

$$=: \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4.$$

It is clear that

$$\mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}} \left\{ \sum_{i=1}^{n} X_i > x_n \sqrt{n} \right\}$$

$$\leq \sum_{j=1}^{4} \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}} \left\{ \Delta_j > \frac{1}{4(1 - \rho_*)} x_n \sqrt{n} \right\}.$$

We now argue that for any $x_n \to \infty$,

$$(9) \qquad \sup_{\boldsymbol{\mu} \in \widetilde{\Xi}(s, B)} \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}} \left\{ \Delta_j > \frac{1}{4(1 - \rho_*)} x_n \sqrt{n} \right\} \to 0, \qquad j = 1, \ldots, 4.$$

The case for Δ_4 follows from our assumption ($\|\mathbf{Q}\mathbf{1} - \frac{\mathbf{1}^\top \mathbf{Q}\mathbf{1}}{n} \mathbf{1}\|^2 = O(1)$) upon the Cauchy–Schwarz inequality. The case Δ_1 follows immediately from Lemma 1. On

the other hand, we note that

$$\sum_{i=1}^{n}\left[\tanh\left(m_i(\mathbf{X}) + \mu_i\right) - \tanh\left(m_i(\mathbf{X})\right)\right]$$

$$\leq \sum_{i=1}^{n}\left|\tanh\left(m_i(\mathbf{X}) + \mu_i\right) - \tanh\left(m_i(\mathbf{X})\right)\right|$$

$$\leq \sum_{i=1}^{n}\tanh(\mu_i) = s\tanh(B),$$

where the second inequality follows from the subadditivity of tanh. The bound (9) for Δ_2 then follows from the fact that $s\tanh(B) = O(\sqrt{n})$.

We now consider Δ_3. Recall that $|x - \tanh(x)| \leq x^2$. It suffices to show that, as $x_n \to \infty$,

$$(10) \qquad \sup_{\boldsymbol{\mu} \in \widetilde{\Xi}(s,B)} \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}\left\{\sum_{i=1}^{n} m_i^2(\mathbf{X}) > \frac{1}{4}x_n\sqrt{n}\right\} \to 0,$$

which follows from Markov inequality and the following lemma.

LEMMA 2. *Let* $\mathbf{X}$ *be a random vector following the Ising model* (1). *Assume that* $\mathbf{Q}_{i,j} \geq 0$ *for all* (i, j) *such that* $\|\mathbf{Q}\|_{\ell_\infty \to \ell_\infty} \leq \rho$ *for some constant* $\rho < 1$, *and* $\|\mathbf{Q}\|_F^2 = O(\sqrt{n})$. *Then for any fixed* $C > 0$,

$$\limsup_{n \to \infty} \sup_{\substack{\boldsymbol{\mu} \in [0,1]^n \\ \sum_{i=1}^{n} \mu_i \leq C\sqrt{n}}} \frac{1}{\sqrt{n}} \mathbb{E}_{\mathbf{Q},\boldsymbol{\mu}}\left(\sum_{i=1}^{n} m_i^2(\mathbf{X})\right) < \infty.$$

The proof of Lemma 2 is deferred to the Appendix in Mukherjee, Mukherjee and Yuan (2018).

The case of $B > 1$. In this case, $s\tanh(B) \leq C\sqrt{n}$ implies $s \leq C'\sqrt{n}$, where $C' := C/\tanh(1)$. Also, since the statistic $\sum_{i=1}^{n} X_i$ is stochastically nondecreasing in B, without loss of generality it suffices to show that, for a fixed $S \subset [n]$ obeying $|S| = s$,

$$(11) \qquad \limsup_{K \to \infty} \limsup_{n \to \infty} \limsup_{B \to \infty} \sup_{\substack{\boldsymbol{\mu} \in \widetilde{\Xi}(s,B) \\ \text{supp}(\boldsymbol{\mu})=S}} \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}\left\{\sum_{i \in S^c} X_i > K\sqrt{n}\right\} = 0.$$

Now, for $i \in S$ we have for $\boldsymbol{\mu} \in \widetilde{\Xi}(s, B)$

$$\mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}(X_i = 1 | X_j = x_j, j \neq i) = \frac{e^{B+m_i(\mathbf{x})}}{e^{B+m_i(\mathbf{x})} + e^{-B-m_i(\mathbf{x})}}$$

$$= \frac{1}{1 + e^{-2m_i(\mathbf{x})-2B}} \geq \frac{1}{1 + e^{2-2B}},$$

and so $\lim_{B\to\infty} \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}(X_i = 1, i \in S) = 1$ uniformly in $\boldsymbol{\mu} \in \tilde{\Xi}(s, B)$ with $s \leq C'\sqrt{n}$. Also note that for any configuration $(x_j, j \in S^c)$ we have

$$
\text{(12)} \qquad
\begin{aligned}
&\mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}\big(X_i = x_i, i \in S^c | X_i = 1, i \in S\big) \\
&\qquad \propto \exp\bigg(\frac{1}{2}\sum_{i,j \in S^c} x_i x_j \mathbf{Q}_{ij} + \sum_{i \in S^c} x_i \tilde{\mu}_{S,i}\bigg),
\end{aligned}
$$

where $\tilde{\mu}_{S,i} := \sum_{j \in S} \mathbf{Q}_{ij} \leq \|\mathbf{Q}\|_{\ell_\infty \to \ell_\infty} \leq \rho$. Further we have

$$
\text{(13)} \qquad \sum_{i=1}^{n} \tilde{\mu}_{S,i} = \sum_{i=1}^{n}\sum_{j \in S} \mathbf{Q}_{ij} = \sum_{j \in S}\sum_{i=1}^{n} \mathbf{Q}_{ij} \leq C'\rho\sqrt{n}.
$$

We shall refer to the distribution in (12) as $\mathbb{P}_{\tilde{\mathbf{Q}}_S, \tilde{\boldsymbol{\mu}}_S}$ where $\tilde{\mathbf{Q}}_S$ is the $(n-s) \times (n-s)$ principle matrix of $\mathbf{Q}$ by restricting the index in S^c. Therefore, we simply need to verify that $\tilde{\mathbf{Q}}_S$ satisfy the conditions for $\mathbf{Q}$ in Theorem 6. Trivially, $\tilde{\mathbf{Q}}_{ij} \geq 0$ for all i, j and $\|\tilde{\mathbf{Q}}\|_{\ell_\infty \to \ell_\infty} \leq \|\mathbf{Q}\|_{\ell_\infty \to \ell_\infty} \leq \rho$. For verifying the third condition, that is,

$$
\left\|\tilde{\mathbf{Q}}\mathbf{1} - \frac{\mathbf{1}^\top \tilde{\mathbf{Q}}\mathbf{1}}{n}\mathbf{1}\right\|^2 = O(1),
$$

note that

$$
\begin{aligned}
O(1) &= \left\|\mathbf{Q}\mathbf{1} - \frac{\mathbf{1}^\top \mathbf{Q}\mathbf{1}}{n}\mathbf{1}\right\|^2 \\
&= \frac{1}{2n}\sum_{i,j=1}^{n}\big(r_i(\mathbf{Q}) - r_j(\mathbf{Q})\big)^2 \\
&\geq \frac{1}{2n}\sum_{i,j \in S^c}\big(r_i(\mathbf{Q}) - r_j(\mathbf{Q})\big)^2 \\
&= \frac{n-s}{n} \times \frac{1}{2(n-s)}\sum_{i,j \in S^c}\big(r_i(\mathbf{Q}) - r_j(\mathbf{Q})\big)^2 \\
&\geq \frac{n-s}{n}\left\|\tilde{\mathbf{Q}}\mathbf{1} - \frac{\mathbf{1}^\top \tilde{\mathbf{Q}}\mathbf{1}}{n}\mathbf{1}\right\|^2.
\end{aligned}
$$

Therefore, with $o_B(1)$ denoting a sequence of real numbers that converges to 0 uniformly over $\boldsymbol{\mu} \in \tilde{\Xi}(s, B)$,

$$
\begin{aligned}
&\limsup_{B\to\infty} \sup_{\substack{\boldsymbol{\mu} \in \tilde{\Xi}(s,B) \\ \text{supp}(\boldsymbol{\mu})=S}} \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}\bigg\{\sum_{i \in S^c} X_i > K\sqrt{n}\bigg\} \\
&\qquad \leq \limsup_{B\to\infty} \sup_{\substack{\boldsymbol{\mu} \in \tilde{\Xi}(s,B) \\ \text{supp}(\boldsymbol{\mu})=S}} \bigg\{\mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}\bigg(\sum_{i \in S^c} X_i > K\sqrt{n}\big|X_j = 1, j \in S\bigg) + o_B(1)\bigg\}
\end{aligned}
$$

$$= \limsup_{B \to \infty} \sup_{\substack{\boldsymbol{\mu} \in \tilde{\Xi}(s,B) \\ \text{supp}(\boldsymbol{\mu}) = S}} \mathbb{P}_{\tilde{\mathbf{Q}}_S, \tilde{\boldsymbol{\mu}}_S} \left(\sum_{i \in S^c} X_i > K\sqrt{n} \right)$$

$$\leq \sup_{S \subset [n]} \sup_{\substack{\tilde{\boldsymbol{\mu}}_S: \\ \sum_{i \in S^c} \tilde{\mu}_{S,i} \leq C' \rho \sqrt{n}}} \mathbb{P}_{\tilde{\mathbf{Q}}_S, \tilde{\boldsymbol{\mu}}_S} \left(\sum_{i \in S^c} X_i > K\sqrt{n} \right),$$

where the last line follows from (13). The proof of the claim (11) thereafter follows using the same argument as that for the case when $B < 1$ since $\tilde{\mu}_{S,i} \leq \rho < 1$ for each $i \in S^c$.

6.2.3. *Proof of* (7). It is clear that, by symmetry,

$$(14) \qquad \mathbb{P}_{\mathbf{Q},\mathbf{0}} \left(\left| \sum_{i=1}^n X_i \right| > K\sqrt{n} \right) = 2\mathbb{P}_{\mathbf{Q},\mathbf{0}} \left(\sum_{i=1}^n X_i > K\sqrt{n} \right).$$

In establishing (8), we essentially proved that

$$(15) \qquad \limsup_{K \to \infty} \limsup_{n \to \infty} \sup_{\boldsymbol{\mu} \in \tilde{\Xi}(s,B)} \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}} \left(\sum_{i=1}^n X_i > K\sqrt{n} \right) = 0.$$

By choosing K large enough, we can make the right-hand side of (14) less than $1/2$. This gives

$$(16) \qquad \sum_{\mathbf{x} \in \{-1,1\}^n} e^{\mathbf{x}^\top \mathbf{Q} \mathbf{x}/2} \leq 2 \sum_{\mathbf{x} \in D_{n,K}} e^{\mathbf{x}^\top \mathbf{Q} \mathbf{x}/2},$$

where $D_{n,K} := \{|\sum_{i=1}^n X_i| \leq K\sqrt{n}\}$. Then, setting $C_n := \{\sum_{i=1}^n X_i > \lambda\sqrt{n}\}$, for any $K > \lambda$ we have

$$\mathbb{P}_{\mathbf{Q},\mathbf{0}}(C_n) \geq \mathbb{P}_{\mathbf{Q},\mathbf{0}}(C_n \cap D_{n,K})$$

$$= \frac{\sum_{\mathbf{x} \in C_n \cap D_{n,K}} e^{\mathbf{x}' \mathbf{Q} \mathbf{x}/2}}{\sum_{\mathbf{x} \in \{-1,1\}^n} e^{\mathbf{x}' \mathbf{Q} \mathbf{x}/2}}$$

$$\geq \frac{1}{2} \frac{\sum_{\mathbf{x} \in C_n \cap D_{n,K}} e^{\mathbf{x}' \mathbf{Q} \mathbf{x}/2}}{\sum_{\mathbf{x} \in D_{n,K}} e^{\mathbf{x}' \mathbf{Q} \mathbf{x}/2}}$$

$$\geq \frac{e^{-2Kt}}{2} \frac{\sum_{\mathbf{x} \in C_n \cap D_{n,K}} e^{\mathbf{x}' \mathbf{Q} \mathbf{x}/2 + \frac{t}{\sqrt{n}} \sum_{i=1}^n x_i}}{\sum_{\mathbf{x} \in D_{n,K}} e^{\mathbf{x}' \mathbf{Q} \mathbf{x}/2}}$$

$$= \frac{e^{-2Kt}}{2} \frac{\mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}(t)}(C_n \cap D_{n,K})}{\mathbb{P}_{\mathbf{Q},\mathbf{0}}(D_{n,K})} \frac{Z(\mathbf{Q}, \boldsymbol{\mu}(t))}{Z(\mathbf{Q}, \mathbf{0})}$$

$$\geq \frac{e^{-2Kt}}{2} \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}(t)}(C_n \cap D_{n,K}),$$

where $\boldsymbol{\mu}(t) = tn^{-1/2}\mathbf{1}$. In the last inequality we use the fact that the function $t \mapsto Z(\mathbf{Q}, \boldsymbol{\mu}(t))$ is nonincreasing in t on $[0, \infty)$, as

$$\frac{\partial}{\partial t} Z(\mathbf{Q}, \boldsymbol{\mu}(t)) = \frac{1}{\sqrt{n}} \mathbb{E}_{\mathbf{Q}, \boldsymbol{\mu}(t)} \sum_{i=1}^{n} X_i \geq \frac{1}{\sqrt{n}} \mathbb{E}_{\mathbf{Q}, \mathbf{0}} \sum_{i=1}^{n} X_i = 0.$$

To show (7), it thus suffices to show that there exists K large enough and $t > 0$ such that

$$\liminf_{n \to \infty} \mathbb{P}_{\mathbf{Q}, \boldsymbol{\mu}(t)}(C_n \cap D_{n,K}) > 0.$$

To this end, it suffices to show that for any $\lambda > 0$ there exists t such that

$$(17) \qquad \liminf_{n \to \infty} \mathbb{P}_{\mathbf{Q}, \boldsymbol{\mu}(t)} \left(\sum_{i=1}^{n} X_i > \lambda \sqrt{n} \right) > 0.$$

If (17) holds, then there exists $t > 0$ such that

$$\liminf_{n \to \infty} \mathbb{P}_{\mathbf{Q}, \boldsymbol{\mu}(t)}(C_n) > 0.$$

It now suffices to show that for any t fixed one has

$$\limsup_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}_{\mathbf{Q}, \boldsymbol{\mu}(t)}(D_{n,K}^c) = 0,$$

which follows from (15).

It now remains to show (17). To begin, note that for $h > 0$,

$$\mathbb{E}_{\mathbf{Q}, \boldsymbol{\mu}(h)} X_i = \mathbb{E}_{\mathbf{Q}, \boldsymbol{\mu}(h)} \tanh\left(m_i(\mathbf{X}) + \frac{h}{\sqrt{n}} \right)$$

$$= \mathbb{E}_{\mathbf{Q}, \boldsymbol{\mu}(h)} \frac{\tanh(m_i(\mathbf{X})) + \tanh(\frac{h}{\sqrt{n}})}{1 + \tanh(m_i(\mathbf{X})) \tanh(\frac{h}{\sqrt{n}})}$$

$$\geq \frac{1}{2} \left[\mathbb{E}_{\mathbf{Q}, \boldsymbol{\mu}(h)} \tanh(m_i(\mathbf{X})) + \tanh\left(\frac{h}{\sqrt{n}} \right) \right]$$

$$\geq \frac{1}{2} \tanh\left(\frac{h}{\sqrt{n}} \right).$$

In the last inequality, we use Holley inequality [e.g., Theorem 2.1 of Grimmett (2006)] for the two probability measures $\mathbb{P}_{\mathbf{Q}, \mathbf{0}}$ and $\mathbb{P}_{\mathbf{Q}, \boldsymbol{\mu}(h)}$ to conclude

$$\mathbb{E}_{\mathbf{Q}, \boldsymbol{\mu}(h)} \tanh(m_i(\mathbf{X})) \geq \mathbb{E}_{\mathbf{Q}, 0} \tanh(m_i(\mathbf{X})) = 0,$$

in the light of (2.7) of Grimmett (2006). Adding over $1 \leq i \leq n$ gives

$$(18) \qquad F_n'(h) = \frac{1}{\sqrt{n}} \mathbb{E}_{\mathbf{Q}, \boldsymbol{\mu}(h)} \sum_{i=1}^{n} X_i \geq \frac{\sqrt{n}}{2} \tanh\left(\frac{h}{\sqrt{n}} \right),$$

where $F_n(h)$ is the log normalizing constant for the model $\mathbb{P}_{\mathbf{Q}, \boldsymbol{\mu}(h)}$. Thus, using Markov's inequality one gets

$$\mathbb{P}_{\mathbf{Q}, \boldsymbol{\mu}(t)}\left(\sum_{i=1}^n X_i \leq \lambda \sqrt{n}\right) = \mathbb{P}_{\mathbf{Q}, \boldsymbol{\mu}(t)}\left(e^{-\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i} \geq e^{-\lambda}\right)$$

$$\leq \exp\{\lambda + F_n(t-1) - F_n(t)\}.$$

Using (18), the exponent in the rightmost-hand side can be estimated as

$$\lambda + F_n(t-1) - F_n(t) = \lambda - \int_{t-1}^t F_n'(h)\, dh \leq \lambda - \frac{\sqrt{n}}{2} \tanh\left(\frac{t-1}{\sqrt{n}}\right),$$

which is negative and uniformly bounded away from 0 for all n large for $t = 4\lambda + 1$, from which (17) follows.

6.3. *Proof of Theorem* 5. We set $m_i(\mathbf{X}) = \sum_{j=1}^n \mathbf{Q}_{ij} X_j$ and assume $\boldsymbol{\mu} \in \tilde{\Xi}(s, B)$ with $s \tanh(B) \leq C\sqrt{n}$. By the same argument as that of Section 6.2.1, it suffices to show that there does not exist a sequence of positive reals $\{L_n\}_{n \geq 1}$ such that

$$\mathbb{P}_{\mathbf{Q}, \mathbf{0}}\left(\sum_{i=1}^n X_i > L_n\right) + \mathbb{P}_{\mathbf{Q}, \boldsymbol{\mu}}\left(\sum_{i=1}^n X_i < L_n\right) \to 0.$$

Suppose, to the contrary, that there exists such a sequence. For any $t \in \mathbb{R}$, we have

$$\mathbb{E}_{\mathbf{Q}, \mathbf{0}} \exp\left\{\frac{t}{\sqrt{n}} \sum_{i=1}^n X_i\right\} = \frac{Z(\mathbf{Q}, \frac{t}{\sqrt{n}}\mathbf{1})}{Z(\mathbf{Q}, \mathbf{0})} = \lambda_1\left(\frac{t}{\sqrt{n}}\right)^n + \lambda_2\left(\frac{t}{\sqrt{n}}\right)^n,$$

where

$$\lambda_i(t) := \frac{e^\theta \cosh(t) + (-1)^{i+1} \sqrt{e^{2\theta} \sinh(t)^2 + e^{-2\theta}}}{e^\theta + e^{-\theta}}.$$

This computation for the normalizing constants for the Ising model on the cycle graph of length n is standard [Ising (1925)]. By a direct calculation we have

$$\lambda_1(0) = 1 > \lambda_2(0) = \tanh(\theta),$$
$$\lambda_1'(0) = \lambda_2'(0) = 0,$$
$$c(\theta) := \lambda_1''(0) > 0,$$

and so

$$\mathbb{E}_{\mathbf{Q}, \mathbf{0}} e^{\frac{t}{\sqrt{n}} \sum_{i=1}^n X_i} = \lambda_1\left(\frac{t}{\sqrt{n}}\right)^n + \lambda_2\left(\frac{t}{\sqrt{n}}\right)^n \xrightarrow{n \to \infty} e^{\frac{c(\theta)t^2}{2}}.$$

This implies that under H_0

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \xrightarrow{d} N(0, c(\theta)),$$

which for any $\lambda > 0$ gives

$$\liminf_{n \to \infty} \mathbb{P}_{\mathbf{Q},\mathbf{0}} \left(\sum_{i=1}^{n} X_i > \lambda \sqrt{n} \right) > 0.$$

Therefore, $L_n \gg \sqrt{n}$. Now invoking Lemma 1, for any $K > 0$ we have

$$\mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}} \left\{ \left| \sum_{i=1}^{n} (X_i - \tanh(m_i(\mathbf{X}) + \mu_i)) \right| > K\sqrt{n} \right\} \le 2e^{-K^2/4(1+\theta)^2}.$$

On this set, we have for a universal constant $C < \infty$

$$\left| \sum_{i=1}^{n} (X_i - \tanh(m_i(\mathbf{X}))) \right| \le \left| \sum_{i=1}^{n} (X_i - \tanh(m_i(\mathbf{X}) + \mu_i)) \right|$$

$$+ \left| \sum_{i=1}^{n} (\tanh(m_i(\mathbf{X}) + \mu_i) - \tanh(m_i(\mathbf{X}))) \right|$$

$$\le K\sqrt{n} + C \sum_{i=1}^{n} \tanh(\mu_i)$$

$$\le K\sqrt{n} + Cs \tanh(B),$$

and so

$$\mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}} \left\{ \left| \sum_{i=1}^{n} (X_i - \tanh(m_i(\mathbf{X}))) \right| > K\sqrt{n} + Cs \tanh(B) \right\}$$

(19)

$$\le 2e^{-K^2/4(1+\theta)^2}.$$

Also, setting $g(t) := t/\theta - \tanh(t)$, we get

$$\sum_{i=1}^{n} (X_i - \tanh(m_i(\mathbf{X}))) = \sum_{i=1}^{n} g(m_i(\mathbf{X})) = \{Q_n(\mathbf{X}) - R_n(\mathbf{X})\} g(\theta),$$

where

$$Q_n(\mathbf{X}) := |\{1 \le i \le n : m_i(\mathbf{X}) = \theta\}|,$$

$$R_n(\mathbf{X}) := |\{1 \le i \le n : m_i(\mathbf{X}) = -\theta\}|.$$

Indeed, this holds, as in this case $m_i(\mathbf{X})$ can take only three values $\{-\theta, 0, \theta\}$, and $g(\cdot)$ is an odd function. Thus using (19) gives

$$\mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}_n} \left\{ |Q_n(\mathbf{X}) - R_n(\mathbf{X})| > \frac{K\sqrt{n} + Cs \tanh(B)}{g(\theta)} \right\} \le 2e^{-K^2/4(1+\theta)^2}.$$

But then we have

$$\mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}_n}\left\{\sum_{i=1}^n X_i > L_n\right\} = \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}\left\{\sum_{i=1}^n m_i(\mathbf{X}) > \theta L_n\right\}$$

$$= \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}\{Q_n(\mathbf{X}) - R_n(\mathbf{X}) > L_n\}$$

$$\le 2e^{-K^2/4(1+\theta)^2},$$

as

$$L_n \gg \frac{K\sqrt{n} + Cs\tanh(B)}{g(\theta)}.$$

This immediately yields the desired result.

6.4. *Proof of Theorem* 1.   By Theorem 6, there is no asymptotically powerful test if $s\tanh(B) = O(n^{1/2})$. It now suffices to show that the naïve test is indeed asymptotically powerful. To this end, we first consider the Type I error. By Theorem 2 of Ellis and Newman (1978),

$$\sqrt{n}\bar{X} \to_d N\left(0, \frac{1}{1-\theta}\right),$$

which immediately implies that Type I error

$$\mathbb{P}_{\theta,\mathbf{0}}(\sqrt{n}\bar{X} \ge L_n) \to 0.$$

Now consider Type II error. Observe that

$$\bar{X} - f_{\mathbf{Q},\boldsymbol{\mu}}(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^n \tanh\left(\sum_{j\neq i}\mathbf{Q}_{ij}X_j + \mu_i\right)$$

$$= \frac{1}{n}\sum_{i=1}^n \tanh(\theta\bar{X} + \mu_i - \theta X_i/n)$$

$$= \frac{1}{n}\sum_{i=1}^n \tanh(\theta\bar{X} + \mu_i) + O(n^{-1}),$$

where the last equality follows from the fact that tanh is Lipschitz. In addition,

$$\frac{1}{n}\sum_{i=1}^n \tanh(\theta\bar{X} + \mu_i) = \tanh(\theta\bar{X}) + \frac{1}{n}\sum_{i\in\text{supp}(\boldsymbol{\mu})}[\tanh(\theta\bar{X} + \mu_i) - \tanh(\theta\bar{X})]$$

$$\ge \tanh(\theta\bar{X}) + \frac{1}{n}\sum_{i\in\text{supp}(\boldsymbol{\mu})}[\tanh(\theta\bar{X} + B) - \tanh(\theta\bar{X})]$$

$$\ge \tanh(\theta\bar{X}) + \frac{s\tanh(B)}{n}[1 - \tanh(\theta\bar{X})],$$

$$\ge \tanh(\theta\bar{X}) + \frac{s\tanh(B)}{n}[1 - \tanh(\theta)],$$

where the second to last inequality follows from (6). In other words,

$$\sqrt{n}\big(\bar{X} - \tanh(\theta\bar{X})\big) - \sqrt{n}\, f_{\mathbf{Q},\boldsymbol{\mu}}(\mathbf{X}) \geq \frac{s\tanh(B)}{\sqrt{n}}\big[1 - \tanh(\theta)\big].$$

Since $\sup_{x\in\mathbb{R}} \frac{x-\tanh(\theta x)}{x} < \infty$, an application of Lemma 1, together with the fact that $L_n = o(n^{-1/2}s\tanh(B))$ yields

$$\mathbb{P}_{\theta,\boldsymbol{\mu}}(\sqrt{n}\bar{X} \geq L_n) \to 1.$$

6.5. *Proof of Theorem* 2. The proof of attainability follows immediately from Theorem 4. Therefore, here we focus on the proof of the lower bound. As before, by the same argument as those following Section 6.2.1, it suffices to show that there does not exist a sequence of positive reals $\{L_n\}_{n\geq 1}$ such that

$$\mathbb{P}_{\mathbf{Q},\mathbf{0}}\left(\sum_{i=1}^{n} X_i > L_n\right) + \mathbb{P}_{\mathbf{Q},\boldsymbol{\mu}}\left(\sum_{i=1}^{n} X_i < L_n\right) \to 0.$$

From the proof of Lemma 1 and the inequality $|\tanh(x) - \tanh(y)| \leq |x - y|$, for any fixed $t < \infty$ and $\boldsymbol{\mu} \in \widetilde{\Xi}(s, B)$ we have

$$\mathbb{P}_{\theta,\boldsymbol{\mu}}\left(\bar{X} > \frac{s}{n}\tanh(\theta\bar{X} + B) + \frac{n-s}{n}\tanh(\theta\bar{X}) + \frac{\theta}{n} + \frac{t}{\sqrt{n}}\right) \leq 2e^{-\frac{t^2}{2na_n}},$$

where

$$a_n := \frac{2}{n} + \frac{2\theta}{n} + \frac{2\theta}{n^2}.$$

Also note that

$$\frac{s}{n}\tanh(\theta\bar{X} + B) + \frac{n-s}{n}\tanh(\theta\bar{X}) \leq \tanh(\theta\bar{X}) + C\frac{s}{n}\tanh(B),$$

for some constant $C < \infty$. Therefore,

$$\mathbb{P}_{\theta,\boldsymbol{\mu}}\left\{\bar{X} - \tanh(\theta\bar{X}) > C\frac{s}{n}\tanh(B) + \frac{\theta}{n} + \frac{t}{\sqrt{n}}\right\} \leq 2\exp(-t^2/2na_n).$$

Since $s\tanh(B) = O(n^{1/2})$, we have

(20) $$\sup_{\boldsymbol{\mu}\in\widetilde{\Xi}(s,B)} \mathbb{P}_{\theta,\boldsymbol{\mu}}\left\{\bar{X} - \tanh(\theta\bar{X}) > \frac{C(t)}{\sqrt{n}}\right\} \leq 2\exp(-t^2/2na_n)$$

for some finite positive constant $C(t)$. Now, invoking Theorem 1 of Ellis and Newman (1978), under $H_0 : \boldsymbol{\mu} = \mathbf{0}$ we have

$$\sqrt{n}(\bar{X} - m)|\bar{X} > 0 \xrightarrow{d} N\left(0, \frac{1-m^2}{1-\theta(1-m^2)}\right),$$

where m is the unique positive root of $m = \tanh(\theta m)$. The same argument as that from Section 6.2.1 along with the requirement to control the Type I error then

imply that without loss of generality one can assume the test ϕ_n rejects if $\bar{X} > m + L_n$, where $L_n \gg n^{-1/2}$.

Now, note that $g(x) = x - \tanh(\theta x)$ implies that $g'(x)$ is positive and increasing on the set $[m, \infty)$ and, therefore,

$$g(x) \geq g(m) + (x - m)g'(m).$$

This gives

$$\mathbb{P}_{\theta,\boldsymbol{\mu}}\left(\bar{X} > m + L_n, \bar{X} - \tanh(\theta\bar{X}) \leq \frac{C(t)}{\sqrt{n}}\right)$$

$$\leq \mathbb{P}_{\theta,\boldsymbol{\mu}}\left(\bar{X} > m + L_n, \bar{X} - m \leq \frac{C(t)}{g'(m)\sqrt{n}}\right),$$

which is 0 for all large n, as $L_n \gg n^{-1/2}$. This, along with (20) gives

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\mu}\in\tilde{\Xi}(s,B)} \mathbb{E}_{\theta,\boldsymbol{\mu}}(1 - \phi_n) \geq 1,$$

thus concluding the proof.

6.6. *Proof of Theorem* 3. The proof of Theorem 3 is based on an auxiliary variable approach known as Kac's Gaussian transform [Kac (1959)], which basically says that the moment generating function of $N(0,1)$ is $e^{t^2/2}$. This trick has already been used in computing asymptotics of log partition functions [Comets and Gidas (1991), Mukherjee (2013), Park and Newman (2004)].

In particular, the proof relies on the following two technical lemmas. The proof to both lemmas is relegated to the Appendix in Mukherjee, Mukherjee and Yuan (2018) for brevity.

LEMMA 3. *Let* $\mathbf{X}$ *follow a Curie–Weiss model of* (5) *with* $\theta > 0$. *Given* $\mathbf{X} = \mathbf{x}$ *let* Z_n *be a normal random variable with mean* $\bar{x}$ *and variance* $1/(n\theta)$. *Then*:

(a) *Given* $Z_n = z$ *the random variables* $(X_1, \ldots, X_n)$ *are mutually independent, with*

$$\mathbb{P}_{\theta,\boldsymbol{\mu}}(X_i = x_i) = \frac{e^{(\mu_i + z\theta)x_i}}{e^{\mu_i + z\theta} + e^{-\mu_i - z\theta}},$$

where $x_i \in \{-1, 1\}$.

(b) *The marginal density of* Z_n *is proportional to* $e^{-f_{n,\boldsymbol{\mu}}(z)}$, *where*

$$(21) \qquad f_{n,\boldsymbol{\mu}}(z) := \frac{n\theta z^2}{2} - \sum_{i=1}^{n} \log\cosh(\theta z + \mu_i).$$

(c)

$$\sup_{\boldsymbol{\mu}\in[0,\infty)^n} \mathbb{E}_{\theta,\boldsymbol{\mu}}\left(\sum_{i=1}^{n}(X_i - \tanh(\mu_i + \theta Z_n))\right)^2 \leq n.$$

While the previous lemma applies to all $\theta > 0$, the next one specializes to the case $\theta = 1$ and gives crucial estimates which will be used in proving Theorem 3.

For any $\boldsymbol{\mu} \in (\mathbb{R}^+)^n$, set

$$A(\boldsymbol{\mu}) := \frac{1}{n} \sum_{i=1}^{n} \tanh(\boldsymbol{\mu}_i).$$

This can be thought of as the total amount of signal present in the parameter $\boldsymbol{\mu}$. In particular, note that for $\boldsymbol{\mu} \in \Xi(s, B)$ we have

$$A(\boldsymbol{\mu}) \geq \frac{s \tanh(B)}{n},$$

and for $\boldsymbol{\mu} \in \tilde{\Xi}(s, B)$ we have

$$A(\boldsymbol{\mu}) = \frac{s \tanh(B)}{n}.$$

In the following, we abbreviate $s \tanh(B)/n := A_n$.

LEMMA 4.   (a) *If* $\theta = 1$, *for any* $\boldsymbol{\mu} \in \Xi(s, B)$ *the function* $f_{n,\boldsymbol{\mu}}(\cdot)$ *defined by* (21) *is strictly convex, and has a unique global minimum* $m_n \in (0, 1]$, *such that*

(22) $$m_n^3 = \Theta(A(\boldsymbol{\mu})).$$

(b)

$$\limsup_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}_{\theta,\boldsymbol{\mu}}(Z_n - m_n > K n^{-1/4}) = 0.$$

(c) *If* $A_n \gg n^{-3/4}$ *then there exists* $\delta > 0$ *such that*

$$\limsup_{n \to \infty} \sup_{\boldsymbol{\mu}:A(\boldsymbol{\mu}) \geq A_n} \mathbb{P}_{\theta,\boldsymbol{\mu}}(Z_n \leq \delta m_n) = 0.$$

The proof of Lemma 4 can be found in the Appendix in Mukherjee, Mukherjee and Yuan (2018). We now come back to the proof of Theorem 3. To establish the upper bound, define a test function ϕ_n by $\phi_n(\mathbf{X}) = 1$ if $\bar{X} > 2\delta A_n^{1/3}$, and 0 otherwise, where δ is as in part (c) of Lemma 4. By Theorem 1 of Ellis and Newman (1978), under $H_0 : \boldsymbol{\mu} = 0$ we have

(23) $$n^{1/4} \bar{X} \xrightarrow{d} Y,$$

where Y is a random variable on $\mathbb{R}$ with density proportional to $e^{-y^4/12}$. Since $A_n \gg n^{-3/4}$ we have

$$\mathbb{P}_{\theta,\mathbf{0}}(\bar{X} > 2\delta A_n^{1/3}) = o(1),$$

and so it suffices to show that

(24) $$\sup_{\boldsymbol{\mu}:A(\boldsymbol{\mu}) \geq A_n} \mathbb{P}_{\theta,\boldsymbol{\mu}}(\bar{X} \leq 2\delta A_n^{1/3}) = o(1).$$

To this effect, note that

$$\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} (X_i - \tanh(\mu_i + Z_n)) + \sum_{i=1}^{n} \tanh(\mu_i + Z_n)$$

$$\geq \sum_{i=1}^{n} (X_i - \tanh(\mu_i + Z_n)) + n \tanh(Z_n).$$

Now by Part (c) of Lemma 3 and Markov's inequality,

$$\left| \sum_{i=1}^{n} (X_i - \tanh(\mu_i + Z_n)) \right| \leq \delta n A_n^{1/3}$$

with probability converging to 1 uniformly over $\boldsymbol{\mu} \in [0, \infty)^n$. Thus it suffices to show that

$$\sup_{\boldsymbol{\mu}: A(\boldsymbol{\mu}) \geq A_n} \mathbb{P}_{\theta, \boldsymbol{\mu}} (n Z_n \leq 3\delta n A_n^{1/3}) = o(1).$$

But this follows on invoking Parts (a) and (c) of Lemma 4, and so the proof of the upper bound is complete.

To establish the lower bound, by the same argument as that from Section 6.2.1, it suffices to show that there does not exist a sequence of positive reals $\{L_n\}_{n \geq 1}$ such that

$$\mathbb{P}_{\mathbf{Q}, \mathbf{0}} \left(\sum_{i=1}^{n} X_i > L_n \right) + \mathbb{P}_{\mathbf{Q}, \boldsymbol{\mu}} \left(\sum_{i=1}^{n} X_i < L_n \right) \to 0.$$

If $\lim_{n \to \infty} n^{-3/4} L_n < \infty$, then (23) implies

$$\liminf_{n \to \infty} \mathbb{E}_{\theta, \mathbf{0}} \phi_n > 0,$$

and so we are done. Thus assume without loss of generality that $n^{-3/4} L_n \to \infty$. In this case, we have

$$\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} (X_i - \tanh(\mu_i + Z_n)) + \sum_{i=1}^{n} \tanh(\mu_i + Z_n)$$

$$\leq \sum_{i=1}^{n} (X_i - \tanh(\mu_i + Z_n)) + \sum_{i=1}^{n} \tanh(\mu_i) + n |Z_n|,$$

and so

$$\mathbb{P}_{\theta, \boldsymbol{\mu}} \left(\sum_{i=1}^{n} X_i > L_n \right)$$

$$\leq \mathbb{P}_{\theta, \boldsymbol{\mu}} \left\{ \left| \sum_{i=1}^{n} X_i - \tanh(\mu_i + Z_n) \right| > L_n/3 \right\}$$

$$+ \mathbb{P}_{\theta, \boldsymbol{\mu}} \{ n Z_n > L_n/3 \} + \mathbb{P}_{\theta, \boldsymbol{\mu}} \{ n Z_n < -L_n/3 \},$$

where we use the fact that

$$\sum_{i=1}^{n} \tanh(\mu_i) = O(n^{1/4}) \ll L_n.$$

Now by Part (c) of Lemma 3 and the Markov inequality, the first term above converges to 0 uniformly over all $\boldsymbol{\mu}$. Also by Parts (a) and (b) of Lemma 4, $\mathbb{P}_{\theta,\boldsymbol{\mu}}\{nZ_n > L_n/3\}$ converges to 0 uniformly over all $\boldsymbol{\mu}$ such that $A(\boldsymbol{\mu}) = O(n^{-3/4})$. Finally, note that the distribution of Z_n is stochastically increasing in $\boldsymbol{\mu}$, and so

$$\mathbb{P}_{\theta,\boldsymbol{\mu}}\{nZ_n < -L_n/3\} \leq \mathbb{P}_{\theta,\boldsymbol{0}}\{nZ_n < -L_n/3\},$$

which converges to 0 by (23). This completes the proof of the lower bound.

SUPPLEMENTARY MATERIAL

Supplement to "Global testing against sparse alternatives under Ising models" (DOI: 10.1214/17-AOS1612SUPP; .pdf). The supplementary material contain the proofs of additional technical results.

REFERENCES

ADDARIO-BERRY, L., BROUTIN, N., DEVROYE, L. and LUGOSI, G. (2010). On combinatorial testing problems. *Ann. Statist.* **38** 3063–3092. MR2722464

ARIAS-CASTRO, E., CANDÈS, E. J. and PLAN, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist.* **39** 2533–2556. MR2906877

ARIAS-CASTRO, E., DONOHO, D. L. and HUO, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inform. Theory* **51** 2402–2425. MR2246369

ARIAS-CASTRO, E. and WANG, M. (2015). The sparse Poisson means model. *Electron. J. Stat.* **9** 2170–2201. MR3406276

ARIAS-CASTRO, E., CANDÈS, E. J., HELGASON, H. and ZEITOUNI, O. (2008). Searching for a trail of evidence in a maze. *Ann. Statist.* **36** 1726–1757. MR2435454

BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* **36** 192–236. MR0373208

BESAG, J. (1975). Statistical analysis of non-lattice data. *Amer. Statist.* 179–195.

BHATTACHARYA, B. B. and MUKHERJEE, S. (2018). Inference in Ising models. *Bernoulli* **24** 493–525. MR3706767

BURNAŠEV, M. V. (1979). Minimax detection of an imperfectly known signal against a background of Gaussian white noise. *Teor. Veroyatn. Primen.* **24** 106–118. MR0522240

CAI, T. T. and YUAN, M. (2014). Rate-optimal detection of very short signal segments. Preprint. Available at arXiv:1407.2812.

CHATTERJEE, S. (2005). Concentration Inequalities with Exchangeable Pairs. Ph.D. thesis, Stanford University. Available at arXiv:math/0507526. MR2707160

CHATTERJEE, S. (2007a). Estimation in spin glasses: A first step. *Ann. Statist.* **35** 1931–1946. MR2363958

CHATTERJEE, S. (2007b). Stein's method for concentration inequalities. *Probab. Theory Related Fields* **138** 305–321. MR2288072

CHATTERJEE, S. and DEY, P. S. (2010). Applications of Stein's method for concentration inequalities. *Ann. Probab.* **38** 2443–2485. MR2683635

COMETS, F. and GIDAS, B. (1991). Asymptotics of maximum likelihood estimators for the Curie–Weiss model. *Ann. Statist.* **19** 557–578. MR1105836

DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195

ELLIS, R. S. and NEWMAN, C. M. (1978). The statistics of Curie–Weiss models. *J. Stat. Phys.* **19** 149–161. MR0503332

GRIMMETT, G. (2006). *The Random-Cluster Model. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **333**. Springer, Berlin. MR2243761

GUYON, X. (1995). *Random Fields on a Network*: *Modeling*, *Statistics*, *and Applications*. Springer, New York. MR1344683

HALL, P. and JIN, J. (2008). Properties of higher criticism under strong dependence. *Ann. Statist.* **36** 381–402. MR2387976

HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* **38** 1686–1732. MR2662357

INGSTER, Y. I. (1994). Minimax detection of a signal in l_p metrics. *J. Math. Sci.* **68** 503–515.

INGSTER, Y. I. (1998). Minimax detection of a signal for l^n-balls. *Math. Methods Statist.* **7** 401–428. MR1680087

INGSTER, Y. I. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models. Lecture Notes in Statistics* **169**. Springer, New York. MR1991446

INGSTER, Y. I., TSYBAKOV, A. B. and VERZELEN, N. (2010). Detection boundary in sparse regression. *Electron. J. Stat.* **4** 1476–1526. MR2747131

ISING, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift Für Physik A Hadrons and Nuclei* **31** 253–258.

JIN, J. and KE, Z. T. (2016). Rare and weak effects in large-scale inference: Methods and phase diagrams. *Statist. Sinica* **26** 1–34. MR3468343

KAC, M. (1959). On the Partition Function of a One-Dimensional Gas. *Phys. Fluids* **2** 8–12.

KAC, M. (1969). Mathematical Mechanisms of Phase Transitions. Technical Report, Rockefeller Univ., New York.

MAJEWSKI, J., LI, H. and OTT, J. (2001). The Ising model in physics and statistical genetics. *Am. J. Hum. Genet.* **69** 853–862.

MÉZARD, M. and MONTANARI, A. (2009). *Information*, *Physics*, *and Computation*. Oxford Univ. Press, Oxford. MR2518205

MUKHERJEE, S. (2013). Consistent estimation in the two star exponential random graph model. Preprint. Available at arXiv:1310.4526.

MUKHERJEE, R., MUKHERJEE, S. and YUAN, M. (2018). Supplement to "Global testing against sparse alternatives under Ising models." DOI:10.1214/17-AOS1612SUPP.

MUKHERJEE, R., PILLAI, N. S. and LIN, X. (2015). Hypothesis testing for high-dimensional sparse binary regression. *Ann. Statist.* **43** 352–381. MR3311863

NISHIMORI, H. (2001). *Statistical Physics of Spin Glasses and Information Processing*: *An Introduction*. Oxford Univ. Press, New York. MR2250384

ONSAGER, L. (1944). Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Phys. Rep.* **65** 117–149. MR0010315

PARK, J. and NEWMAN, M. E. J. (2004). Solution of the two-star model of a network. *Phys. Rev. E* (3) **70** 066146. MR2133810

STAUFFER, D. (2008). Social applications of two-dimensional Ising models. *Am. J. Phys.* **76** 470–473.

WU, Z., SUN, Y., HE, S., CHO, J., ZHAO, H. and JIN, J. (2014). Detection boundary and higher criticism approach for rare and weak genetic effects. *Ann. Appl. Stat.* **8** 824–851. MR3262536

R. MUKHERJEE
DIVISION OF BIOSTATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY
HAVILAND HALL
BERKELEY, CALIFORNIA 94720
USA
E-MAIL: rmukherj@berkeley.edu

S. MUKHERJEE
M. YUAN
DEPARTMENT OF STATISTICS
COLUMBIA UNIVERSITY
1255 AMSTERDAM AVENUE
NEW YORK, NEW YORK 10027
USA
E-MAIL: sm3949@columbia.edu
        ming.yuan@columbia.edu