INFERENC E IN ISING M ODELS ON DENSE REGULAR GRAPHS

BY YUANZHE XU\textsuperscript{1,a} AND SUMIT MUKHERJEE\textsuperscript{2,b}

\textsuperscript{1}Department of Statistics, Columbia University, a\textsuperscript{yuanzhe.xu@columbia.edu}
\textsuperscript{2}Department of Statistics, Columbia University, b\textsuperscript{sm3949@columbia.edu}

In this paper, we derive the limit of experiments for one parameter Ising models on dense regular graphs. In particular, we show that the limiting experiment is Gaussian in the “low temperature” regime, and non Gaussian in the “critical” regime. We also derive the limiting distributions of the maximum likelihood and maximum pseudo-likelihood estimators, and study limiting power for tests of hypothesis against contiguous alternatives. To the best of our knowledge, this is the first attempt at establishing the classical limits of experiments for Ising models (and more generally, Markov random fields).

1. Introduction. The Ising Model is possibly the most well-known discrete graphical model, originating in statistical physics (see [19, 20, 25]), and henceforth studied in depth across several disciplines, including statistics and machine learning (c.f. [2, 6, 14, 13, 17, 24] and the references therein). Under this model, we observe a vector of dependent Rademacher (i.e. $\pm 1$ valued) random variables, where the dependency is controlled by a coupling matrix, and an “inverse temperature” parameter $\theta > 0$ (borrowing statistical physics terminology). Very often this coupling matrix is taken to be the (scaled) adjacency matrix of a graph. Some of the common graph ensembles on which the Ising model has been studied include the complete graph (the corresponding Ising model is known as the Curie-Weiss model), the $d$ dimensional grid, Erdos-Rényi graphs, and random regular graphs ([15, 16, 21, 25]). Note that all the graph ensembles in the above list are (approximately) regular graphs.

In this paper, we will study the behavior of Ising models on a sequence of “dense” regular graphs converging in cut metric (see section 2.2 for a brief introduction to the theory of dense graphs/graphons). Given an Ising model on a dense regular graph parametrized by the inverse temperature parameter $\theta > 0$ (see (1)), we study limits of experiments in the sense of Lucien Le Cam ([22], see also [26]). In particular, we show that the Ising model is locally asymptotically normal (LAN) in the low temperature regime ($\theta > 1$), whereas the limiting experiment is very different in the critical ($\theta = 1$) temperature regime. Using this framework, we derive the limiting power of tests involving the parameter $\theta$, based on the maximum likelihood estimate, pseudo-likelihood estimate, and the sample mean, across all regimes of $\theta$. We also study asymptotic limiting distributions of the maximum likelihood estimate and pseudo-likelihood estimate in the regime $\theta \geq 1$ (where consistent estimation of $\theta$ is possible), and compare their asymptotic performances. Prior to our work, limit distribution of the maximum likelihood estimate was known only for the Curie-Weiss model ([12]), and limit distribution for the pseudo-likelihood estimator was not known in any example (to the best of our knowledge). Thus, we give a complete toolbox for inference regarding the parameter $\theta$, for Ising model on dense graphs.

MSC2020 subject classifications: Primary 62H22; secondary 62E20, 62F05, 62F12.

Keywords and phrases: Ising model, Limits of experiments, Phase transition, Asymptotic power, Asymptotic efficiency.
2. Main results. In this section we formally introduce the Ising model (section 2.1), and state our main results (section 2.6), which are essentially of three types, (i) limits of experiments, (ii) asymptotic performance of estimators, and (iii) asymptotic performance of tests of hypothesis. We recall the notion of limits of experiments in section 2.3. To obtain convergence in experiments, we require the sequence of coupling matrices for the Ising model to converge in cut metric, a notion which we recall in section 2.2. We also introduce the estimators and test statistics that we study in sections 2.4 and 2.5 respectively. Section 2.7 illustrates our results with two concrete examples. Finally, section 2.8 discusses the main contributions of this paper, and possible avenues of future research.

2.1. Formal set up. Let $n$ be a positive integer, and let $Q_n$ be a (known) symmetric $n \times n$ matrix with non-negative entries, and 0 on the diagonal. Let $\theta \geq 0$ be an unknown real valued parameter. Then the Ising model with inverse temperature parameter $\theta$ and coupling matrix $Q_n$ is a probability distribution on $\{-1, 1\}^n$, defined by the probability mass function

$$P_{\theta, Q_n}(X = x) := \exp \left( \frac{\theta}{2} x^T Q_n x - Z_n(\theta, Q_n) \right),$$

for $x \in \{-1, 1\}^n$.

Here $Z(\theta, Q_n)$ is the log normalizing constant which makes (1) into a probability distribution. One of the major challenges in analyzing Ising models is that $Z_n(\theta, Q_n)$ is typically intractable, both analytically and computationally. Consequently, computing and analyzing the maximum likelihood estimate is extremely challenging.

Throughout the paper we will assume that the matrix $Q_n$ satisfies the following assumptions:

• The matrix $Q_n$ is regular, i.e. setting $[n] := \{1, 2, \ldots, n\}$ we have

$$\sum_{j=1}^n Q_n(i, j) = 1, \text{ for all } i \in [n].$$

• There exists a finite positive constant $C_w$ free of $n$ such that

$$\max_{i,j \in [n]} Q_n(i, j) < \frac{C_w}{n}.$$  

• The Frobenius norm of $Q_n$ converges, i.e.

$$\|Q_n\|_F := \sqrt{\sum_{i,j=1}^n Q_n^2(i, j)} \to \gamma.$$  

2.2. Graphon Convergence. Below we will briefly introduce some of the basics of cut metric theory needed for our purposes, referring the audience for more details to [8, 9, 23].

By a graphon, we will mean a symmetric bounded measurable function $f : [0, 1]^2 \to [0, C_w]$. Let $W$ denote the space of all graphons. Equip the space $W$ by the cut distance, defined by

$$d_{\square}(f, g) := \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} (f(x,y) - g(x,y)) dx dy \right|.$$  

Let $M$ denote the space of all measurable measure preserving maps $\iota : [0, 1] \mapsto [0, 1]$. Define an equivalence relation $\sim$ on the space $W$ by setting

$$f \sim g \text{ if } f(x,y) \overset{a.s.}{=} g_\iota(x,y) := g(\iota(x), \iota(y)), \text{ for some } \iota \in M.$$
Let \( \tilde{W} \) denote the quotient space \( W/\sim \) under the above equivalence relation. Equip \( \tilde{W} \) with the cut metric, defined as follows:

\[
\delta_{\square}(\tilde{f}, \tilde{g}) := \inf_{i \in \mathcal{M}} d_{\square}(f_i, g_i) = \inf_{i \in \mathcal{M}} d_{\square}(f_{i_1}, g_{i_2}) = \inf_{i_1, i_2 \in \mathcal{M}} d_{\square}(f_{i_1}, g_{i_2}).
\]

Then it follows from [8] that \( \delta_{\square} \) is well defined, and \( (\tilde{W}, \delta_{\square}) \) is a compact metric space. We say a sequence of graphons \( \{ f_n \}_{n \geq 1} \) converge to a graphon \( f \) in cut metric, if

\[
\delta_{\square}(f_n, f) \to 0.
\]

Given a symmetric \( n \times n \) matrix \( A_n \) with 0 on the diagonal, define a corresponding graphon \( f^A_n \) by setting

\[
f^A_n(x, y) := A_{\lfloor nx \rfloor, \lfloor ny \rfloor}.
\]

We say \( \{ A_n \}_{n \geq 1} \) converge in cut metric to a graphon \( f \), if the corresponding sequence of graphons \( \{ f^A_n \}_{n \geq 1} \) converge to \( f \) in cut metric, i.e.

\[
\delta_{\square}(f^A_n, f) \to 0.
\]

Given \( f \in \mathcal{W} \), define a Hilbert-Schmidt operator \( T_f \) from \( L_2[0,1] \) to \( L_2[0,1] \) by setting

\[
T_f(g)(\cdot) := \int_{[0,1]} f(\cdot, y) g(y) dy.
\]

This is a compact operator, and hence it has (at most) countably many eigenvalues. Let \( \{ \lambda_j \}_{j \geq 1} \) be the eigenvalues of \( T_f \), arranged in decreasing order of absolute value, i.e.

\[
|\lambda_1| \geq |\lambda_2| \geq \cdots.
\]

Throughout the paper we assume that the sequence of matrices \( \{ nQ_n \}_{n \geq 1} \) converge in cut metric to a graphon \( f \in \mathcal{W} \) such that \( \sup_{i \geq 2} \lambda_i < \lambda_1 \), i.e.

\[
\delta_{\square}(f^{nQ_n}, f) \to 0, \quad \sup_{i \geq 2} \lambda_i < \lambda_1.
\]

Note that (2) and (5) together imply \( \lambda_1 = 1 \) (see [23, Theorem 11.54]), and so we require \( \sup_{i \geq 2} \lambda_i < 1 \). This last requirement essentially demands that the eigenvalue 1 is simple, i.e. it has multiplicity 1. We point out that we allow for the possibility that \( |\lambda_i| = 1 \) for some \( i \), which can only happen if \( \lambda_i = -1 \). And indeed, this happens for regular connected bipartite graphs (see (8) for an example). We emphasize that we do not require expander type assumptions typically made in the graph theory literature, which requires \( \max_{i \geq 2} |\lambda_i| < 1 \). A similar spectral gap as in (5) was utilized in [13, Equation 1.7], where the authors study universal limiting distribution of \( \mathbf{X} \) for Ising models on dense regular graphs. In particular, it was shown in [13, Example 1.1] that universality can fail without such an assumption. The same counter-example works in our setting as well, and demonstrates that the limiting experiment may be different without this assumption.

**Remark 2.1.** We now briefly discuss the assumptions on \( Q_n \) made in this paper. Throughout we have assumed that \( Q_n \) is non-negative entry-wise (and \( \theta \geq 0 \)), which ensures that the resulting model has positive association (ferro-magnetic, in statistical physics parlance). We expect that the limiting experiment (and performance of estimators/tests) under the model (1) will be very different in the anti-ferromagnetic regime (\( \theta < 0 \)), or for spin glass models (\( Q_n \) can take both positive and negative entries). This intuition is guided by the fact that the asymptotics of the log normalizing constant \( Z_n(\theta, Q_n) \) is no longer universal.
under these settings (see [1, Theorem 2.3] and [1, Section 1.3.2] respectively).

The first condition (2) demands that the row sums of $Q_n$ are all equal. The fact that the common value of these row sums equal 1 is just a normalization, which ensures that the “critical value” for the corresponding Ising model in (1) is at 1. For non-regular matrices, we expect the limit experiment to be different, and possibly non-universal. The second condition (3), which demands that all the entries of the matrix $nQ_n$ are uniformly bounded, is in essence a compactness assumption. This ensures that the sequence of functions $\{f^{nQ_n}\}_{n \geq 1}$ is uniformly bounded. Without this assumption, we again do not expect universality of the limit experiment. Assumption (3) also implies that assumptions (4) and (5) hold along sub-sequences (since $\|Q_n\|_F$ is bounded, and the space of bounded functions is compact in cut metric, as mentioned above in the graphon section). Thus (4) and (5) are in essence free assumptions, given (3). In fact, convergence in cut metric (i.e. (5)) is not needed if we only want to study the limiting experiment. The main requirement of (5) is due to the fact that we want to characterize the limit distribution of the pseudo-likelihood estimator in the critical regime.

To understand the exact nature of the assumptions (2), (3), (4) and (5), it is instructive to consider the commonly studied case where $Q_n$ is a scaled adjacency matrix of a graph on $n$ vertices, defined as follows:

Let $G_n$ be a simple labeled graph on $n$ vertices, labeled by the set $[n]$. Abusing notation slightly, we also denote by $G_n$ the adjacency matrix of the graph. Then one takes $Q_n = \frac{1}{\bar{d}} G_n$, where $\bar{d} := \frac{1}{n} \sum_{i=1}^{n} d_i$ is the average degree of the graph $G_n$, and $(d_1, \ldots, d_n)$ is the labeled degree sequence of the graph $G_n$. This particular choice ensures that the resulting model is non trivial (see [1, Corollary 1.2]). For the above choice, the assumptions (2) and (3) reduce to the following:

$$d_i = \sum_{j=1}^{n} G_n(i, j) = \bar{d}, \text{ for all } i \in [n], \quad \text{and} \quad \bar{d} \geq \frac{n}{C_W}.$$  

The first assumption demands that the graph $G_n$ is regular, and the second assumption demands that the degree of $G_n$ grows linearly in $n$, i.e. the graph $G_n$ is dense. The fourth assumption demands the convergence of the graph $G_n$ in cut metric to the function $f$. In this case, the third assumption (4) follows from (5), as

$$\sum_{i,j=1}^{n} Q_n^2(i,j) = \frac{1}{\bar{d}^2} \sum_{i,j=1}^{n} G_n(i,j) = \frac{n}{\bar{d}} \rightarrow \frac{1}{\int_{[0,1]} f(x,y) dx dy}.$$  

Thus our results apply to Ising models on dense regular graphs converging in cut metric (i.e. under (2), (3) and (5)).

Below we give some concrete examples of matrix sequences $\{Q_n\}_{n \geq 1}$ which arise as scaled adjacency matrices of commonly studied graphs, for which our results apply.

- Suppose

$$Q_n(i,j) := \frac{1}{n} \text{ if } i \neq j.$$  

\[\text{(6)}\]
This is the adjacency matrix of the complete graph $G_n = K_n$ on $n$ vertices, i.e. $G_n(i,j) = 1$ for all $i, j \in [n]$ such that $i \neq j$. This is a regular graph of degree $d_i = n - 1$. In this case, the corresponding model in (1) is the well known Curie-Weiss model, given by

$$P_{\theta,CW}(X = x) = \exp \left( \frac{n\theta x^2}{2} - Z_n(\theta, CW) \right).$$

We note here that the above definition for $Q_n$ in (6) does not exactly satisfy (2), as

$$\sum_{j=1}^n Q_n(i,j) = \frac{n-1}{n} < 1.$$ 

Nevertheless, it is slightly convenient to specify $Q_n$ using (6), as the eventual formulas are cleaner. Also, it is not hard to show that the Ising models obtained by the two different specifications of $Q_n$ are close in total variation, so all results carry over effortlessly between the two models. The limiting graphon for $f_{nQ_n}$ in both these cases is the constant function $f = 1$, with only one non-zero eigenvalue $\lambda_1 = 1$.

- Suppose $n$ is even, and

$$Q_n(i,j) := \begin{cases} \frac{2}{n} & \text{if } 1 \leq i \leq \frac{n}{2} \text{ and } \frac{n}{2} + 1 \leq j \leq n, \\ \frac{2}{n} & \text{if } 1 \leq j \leq \frac{n}{2} \text{ and } \frac{n}{2} + 1 \leq i \leq n, \\ 0 & \text{otherwise.} \end{cases}$$

This $Q_n$ is the scaled adjacency matrix of the complete bipartite graph $G_n = K_{n/2,n/2}$, which is a regular graph of degree $d_i = \frac{n}{2}$. The limiting graphon for $f_{nQ_n}$ is the piecewise constant function $f$ given by

$$f(x,y) = \begin{cases} 2 & \text{if } 0 < x < 1/2 \text{ and } 1/2 < y < 1, \\ 2 & \text{if } 1/2 < x < 1 \text{ and } 0 < y < 1/2, \\ 0 & \text{otherwise.} \end{cases}$$

This function has two non zero eigenvalues, $\lambda_1 = 1$ and $\lambda_2 = -1$. In this case, we have $|\lambda_2| = |\lambda_1|$, and yet $\lambda_2 < \lambda_1$ (so our results apply).

- More generally, suppose $q \geq 2$ is a positive integer which divides $n$, and let

$$Q_n(i,j) := \frac{q}{n(q-1)} \left\lfloor \frac{i}{q} \right\rfloor \neq \left\lfloor \frac{j}{q} \right\rfloor$$

$$= 0 \text{ otherwise.}$$

Thus $Q_n$ is the scaled adjacency matrix of the complete $q$ partite graph $G_n = K_{n/q,n/q,\ldots,n/q}$, which is a regular graph of degree $\frac{n(q-1)}{q}$. More precisely, the vertex set of $G_n$ can be partitioned into $q$ classes of equal size, such that all cross edges between classes are present. The limiting graphon for $f_{nQ_n}$ is the piecewise constant function $f$ given by

$$f(x,y) = \frac{q}{q-1} \text{ if } \left\lfloor qx \right\rfloor \neq \left\lfloor qy \right\rfloor,$$

$$= 0 \text{ otherwise.}$$

This function has two non zero eigenvalues, $\lambda_1 = 1$ and $\lambda_2 = -\frac{1}{q-1}$. Again in this case we have $\lambda_2 < \lambda_1$, and so our results apply.
As in the example before, suppose \( q \geq 3 \) is a positive integer which divides \( n \). But now, let
\[
Q_n(i, j) := \frac{q}{2n} \text{ if } \left \lfloor \frac{i}{q} \right \rfloor - \left \lfloor \frac{j}{q} \right \rfloor = 1(\text{mod } q),
\]
\[= 0 \text{ otherwise.}
\]
In words, \( Q_n \) is the scaled adjacency matrix of a graph \( G_n \) whose vertex set can be partitioned into \( q \) classes of equal size, such that all edges between consecutive classes are present, i.e. class 2 vertices are all connected to class 1 vertices and class 3 vertices, and class \( q \) vertices are all connected to class \( q - 1 \) vertices and class 1 vertices. Thus \( G_n \) is a regular graph of degree \( \frac{2n}{q} \). The limiting graphon for \( f_nQ_n \) is the piecewise constant function \( f \) given by
\[
f(x, y) = \frac{q}{2} \text{ if } \left \lfloor qx \right \rfloor - \left \lfloor qy \right \rfloor = 1(\text{mod } q),
\]
\[= 0 \text{ otherwise.}
\]
This function has \( q \) non zero eigenvalues given by \( \lambda_a = \cos \left( \frac{2a \pi}{q} \right) \) for \( 0 \leq a \leq q - 1 \) (using eigenvalue formulas for circulant matrices, see [18, Chapter 3]). Again in this case we have \( \lambda_a < \lambda_1 \) for all \( a \in \{2, 3, \ldots, q\} \), and so our results apply. Note that if \( q \) is even, then the matrix \( Q_n \) is bipartite, and the smallest eigenvalue of \( f \) is \(-1\).

Suppose \( \{d_n\}_{n \geq 1} \) is a sequence of positive integers, such that \( \frac{d_n}{n} \to \eta \in (0, 1) \). Let \( G_n \) be a uniformly random \( d_n \)-regular graph, and let
\[
Q_n(i, j) = \frac{1}{d_n} \text{ if } (i, j) \text{ is an edge in } G_n,
\]
\[= 0 \text{ otherwise.}
\]
In this case the graph \( G_n \) is \( d_n \)-regular by construction. It follows from [11, Theorem 1.1] that \( G_n \) converges in cut metric to the constant function \( \eta \), and so \( nQ_n \) converges in cut metric to the constant function \( f \equiv 1 \). Thus the behavior here is the same as that of the Curie-Weiss model (introduced in (7)), as \( \lambda_1 = 1 \) is the only non zero eigenvalue here.

**Remark 2.2.** Even though we consider exactly regular graphs/matrices in this paper for the sake of clarity of arguments, we expect that our results will extend to the setting of approximately regular graphs/matrices, thereby allowing us to study Erdos-Rényi random graphs, stochastic block models, and Wigner matrices with non-negative entries, as done in [13].

### 2.3. Limits of Experiments

In this section we briefly introduce the notion of convergence of experiments. For more details we refer the reader to [26, Chapter 9].

Suppose that for each \( 1 \leq n \leq \infty \) we have a measure space \( (X_n, \mathcal{F}_n) \). Let \( \{\mathbb{P}_{h,n}, h \in H\} \) be a collection of probability measures on \( (X_n, \mathcal{F}_n) \). We say that \( \{\mathbb{P}_{h,n}, h \in H\} \) converges to \( \{\mathbb{P}_{h,\infty}, h \in H\} \) in the sense of limits of experiments, if for every finite subset \( I \) of \( H \) and \( h_0 \in H \) we have
\[
\left( \frac{d\mathbb{P}_{h,n}(X)}{d\mathbb{P}_{h_0,n}} \right)_{h \in I} \xrightarrow{d} \left( \frac{d\mathbb{P}_{h,\infty}(Y)}{d\mathbb{P}_{h_0,\infty}} \right)_{h \in I}.
\]
The RHS of (9) is also called the likelihood ratio process with base $h_0$. We will use the shorthand notation
\[ \mathbb{P}_{h,n} \xrightarrow{\text{Exp}} \mathbb{P}_{h,\infty} \]

to denote convergence of experiments.

In particular, if $H \subseteq \mathbb{R}$ is open, and $\mathbb{P}_{h,\infty} = N(h, \tau^2)$ for some $\tau > 0$, then we say the collection of experiments $\{\mathbb{P}_{h,n}, h \in H\}$ is locally asymptotically normal, or LAN. For examples of both LAN and non LAN experiments, see [26, Chapters 7,9]).

2.4. Estimation of $\theta$. In this paper we will focus on the behavior of two estimators for $\theta$, the maximum likelihood estimator, and the maximum pseudo-likelihood estimator.

- **Maximum Likelihood Estimator (MLE)**

  The maximum likelihood estimator $\hat{\theta}_{n}^{MLE}$ is defined as
  \[
  \arg \sup_{\theta \in \mathbb{R}} \left\{ \frac{\theta}{2} X^T Q_n X - Z_n(\theta, Q_n) \right\},
  \]
  provided the supremum is attained uniquely. Since the function in the above display is strictly concave in $\theta$, it follows that $\hat{\theta}_{n}^{MLE}$, if it exists, is the unique solution to the equation
  \[
  \frac{1}{2} X^T Q_n X = Z_n^\prime(\theta, Q_n) = \frac{\partial Z_n(\theta, Q_n)}{\partial \theta} \bigg|_{\theta = \hat{\theta}_{n}^{MLE}} = \frac{1}{2} \mathbb{E}_{\mathbb{P}_{\theta, Q_n}} X^T Q_n X
  \]
  in $\theta \in \mathbb{R}$. In the special case of the Curie-Weiss model, the asymptotics of $\hat{\theta}_{n}^{MLE}$ was studied in [12], where the authors demonstrated interesting phase transition properties in the limit distribution across different regimes of $\theta$. However, to the best of our knowledge, the behavior of $\hat{\theta}_{n}^{MLE}$ is not understood for almost any other graph sequence.

- **Maximum Pseudo-likelihood Estimator (MPLE)**

  Although the MLE is a natural estimator, from a computational perspective it is often difficult to evaluate, as the normalizing constant $Z_n(\theta, Q_n)$ is computationally intractable. To bypass this, Besag introduced the maximum pseudo-likelihood estimator ([3, 4]) for spatial interaction models. Below we define the maximum pseudo-likelihood estimator $\hat{\theta}_{n}^{MPLE}$.

  Given $X \sim \mathbb{P}_{\theta, Q_n}$, we have
  \[
  \mathbb{P}_{\theta, Q_n}(X_i = x_i | \text{for all } j \neq i) := \frac{\exp(\theta t_i x_i)}{\exp(\theta t_i) + \exp(-\theta t_i)},
  \]
  where $t_i = \sum_{j=1}^{n} Q_n(i, j) X_j$. Define the pseudo-likelihood as the product of the above one dimensional conditional distributions:
  \[
  PL_n(\theta) := \prod_{i=1}^{n} \mathbb{P}_{\theta, Q_n}(X_i = x_i | \text{for all } j \neq i) = \frac{\exp(\theta \sum_{i=1}^{n} X_i t_i)}{2^n \prod_{i=1}^{n} \cosh(\theta t_i)}.
  \]

  The maximum pseudo-likelihood estimator $\hat{\theta}_{n}^{MPLE}$ is defined as
  \[
  \arg \sup_{\theta \in \mathbb{R}} \log PL_n(\theta) = \arg \sup_{\theta \in \mathbb{R}} \left\{ \theta \sum_{i=1}^{n} X_i t_i - \sum_{i=1}^{n} \log \cosh(\theta t_i) \right\},
  \]
provided the supremum is attained uniquely. Since the function in the above display is strictly concave in \( \theta \), the MPLE, if it exists, satisfies the equation

\[
X^T Q_n X = \sum_{i=1}^{n} t_i \tanh(\theta t_i).
\]

The above equation (11) does not involve the intractable function \( Z_n(\theta, Q_n) \), and is much easier to compute. Thus computational complexity of the pseudo-likelihood estimator is much less as compared to the maximum likelihood estimator. The consistency of the pseudo-likelihood estimator for Ising models was established in [6, 10, 17], but the question of asymptotic distribution remained open.

Proposition 3.1 gives an exact characterization for the existence of the MLE and the MPLE, and shows that the conditions hold with probability tending to 1.

2.5. Hypothesis Testing for \( \theta \). Given \( X \sim \mathbb{P}_{\theta, Q_n} \) for some \( \theta > 0 \), suppose we want to test

\[
H_0 : \theta = \theta_0 \text{ vs } H_1 : \theta > \theta_n,
\]

for a positive real number sequence \( \{\theta_n\}_{n=1}^{\infty} \), at level \( \alpha \in (0, 1) \). Here \( \theta_n \) will be chosen (depending on \( \theta_0 \)) in such a manner, that we are in the contiguous regime, i.e. the limiting power of the most powerful test will be between \( (0, 1) \). We consider three natural tests for the above problem in this paper, which are introduced below:

- **Mean Square Test (MS-test)**
  Let \( \bar{X} \) denote the sample mean of \( X \), and let

\[
\psi_n(X) = \begin{cases} 
1 & \text{if } n\bar{X}^2 > K_n(\alpha) \\
0 & \text{otherwise}
\end{cases}
\]

where \( K_n(\alpha) \) is chosen such that \( \psi_n \) has level \( \alpha \). Let \( \beta_{MS} \) denote the limiting power of the above test (provided the limit exists).

- **Neyman Pearson Test (NP-test)** By Neyman Pearson Lemma, the UMP test for the above hypothesis testing problem is based on the sufficient statistics \( X^T Q_n X \), and is given by

\[
\psi_n(X) = \begin{cases} 
1 & \text{if } X^T Q_n X > K_n(\alpha) \\
0 & \text{otherwise}
\end{cases}
\]

where \( K_n(\alpha) \) is chosen such that the above test has level \( \alpha \). It follows from standard exponential family calculations that the above test is equivalent to rejecting for large values of \( \hat{\theta}_n^{MLE} \). Let \( \beta_{NP} \) denote the limiting power of the above test (provided the limit exists).

- **Pseudo-likelihood Test (PL-test)** With \( \hat{\theta}_n^{MLE} \) denoting the pseudo-likelihood estimator, define the pseudo-likelihood test by setting

\[
\psi_n(X) = \begin{cases} 
1 & \text{if } \hat{\theta}_n^{MLE} > K_n(\alpha) \\
0 & \text{otherwise}
\end{cases}
\]

where \( K_n(\alpha) \) is chosen such that the above test has level \( \alpha \). Let \( \beta_{PL} \) denote the limiting power of the above test (provided the limit exists).

**Remark 2.3.** Even though we consider a one sided testing problem in this paper, the same analysis applies to the two sided versions of the above testing problem, with obvious modifications.
2.6. Statement of main results. Before we state our main results, we require some technical definitions, most of which are motivated by the following proposition from [13, Lemma 1.1]:

**Proposition 2.1.** For any $\theta > 0$ consider the fixed point equation
\[
\tag{16} w(\theta, x) = 0, \quad \text{where} \quad w(\theta, x) := x - \tanh(\theta x).
\]
(a) If $\theta > 1$, then (16) has two non-zero roots $\pm m(\theta)$ in $x$, where $m(\theta) > 0$ and $\left. \frac{\partial w(\theta, x)}{\partial x} \right|_{x=m(\theta)} > 0$.
(b) If $\theta = 1$, then (16) has a unique root $m(1) = 0$, and $\left. \frac{\partial w(\theta, x)}{\partial x} \right|_{x=0} = 0$.
(c) If $\theta \in (0, 1)$, then (16) has a unique root $m(\theta) = 0$, and $\left. \frac{\partial w(\theta, x)}{\partial x} \right|_{x=0} > 0$.

**Definition 2.1.** Using Proposition 2.1, we define the function $\theta \mapsto m(\theta)$, where $m(\theta)$ is a non-negative root of the equation $w(\theta, x) = 0$ “chosen carefully” as in Proposition 2.1. Note that $m(\theta) = 0$ if $\theta \leq 1$, and $m(\theta) > 0$ if $\theta > 1$.

We also use the above proposition to partition the parameter space $[1, \infty)$ into two distinct domains:
\begin{itemize}
  \item Low Temperature Regime: $\Theta_1 := (1, \infty)$;
  \item Critical point: $\Theta_2 = 1$;
  \item High Temperature Regime: $\Theta_3 := (0, 1)$.
\end{itemize}

The nomenclature of these domains is inspired from statistical physics terminology (see for e.g. [16]). It follows from [6, Theorem 3.3 part (a)] that consistent estimation/testing for $\theta$ is not impossible in the model $\mathbb{P}_{\theta, Q_n}$ for $\theta \in (0, 1)$. This is also indicated by the fact that for $\theta \in (0, 1)$ the function $m(.)$ is constant. Thus the asymptotics in this regime is of less statistical interest, and we focus only on the regime $\theta \geq 1$ for the rest of this paper. For the interested reader, we refer to version 1 of this article which also treats the regime $\theta \in (0, 1)$ ([27]).

For any $\theta \in \Theta_1$, define the positive real $\sigma^2(\theta)$ by setting
\[
\sigma^2(\theta) := \frac{1 - m^2(\theta)}{1 - \theta(1 - m^2(\theta))}.
\]
Note that $\sigma^2(\theta)$ is well defined by Proposition 2.1, as $1 - (1 - m^2(\theta))\theta = \left. \frac{\partial w(\theta, x)}{\partial x} \right|_{x=m(\theta)} > 0$.

**Definition 2.2.** Given $h \in \mathbb{R}$ and $\theta_0 \geq 1$, define a positive sequence $\{\theta_n\}_{n \geq 1}$ (depending on $h, \theta_0$) by setting $\theta_n := \theta_0 + \frac{h}{\sqrt{n}}$. We will omit the dependence of $\theta_0, h$, since it will be clear from the context.

**Definition 2.3.** Given any continuous real valued random variable $\zeta$, let $\Psi_\zeta(.) : (0, 1) \mapsto \mathbb{R}$ denote the quantile function, i.e. the inverse cdf of $\zeta$, defined by
\[
\Psi_\zeta(p) := \inf \{ t \in \mathbb{R} : F_\zeta(t) \geq p \},
\]
where $F_\zeta$ is the cdf of $\zeta$. In particular if $\zeta \sim N(0, 1)$, then we will also use the notation $z_\alpha$ for $\Psi_\zeta(1 - \alpha)$, as is standard in statistics literature.

**Definition 2.4.** Let $\kappa := \gamma^2 - \|f\|_2^2$, where $\gamma, f$ are as in (4) and (5) respectively.

We now state the results for each of the domains separately.
2.6.1. *The low temperature regime* $\Theta_1$. Our first theorem describes the limits of experiments, asymptotic performance of estimators, and asymptotic performance of tests in low temperature regime $\Theta_1$.

**Theorem 2.2.** Suppose $X \sim \mathbb{P}_{\theta,Q_n}$ with $Q_n$ satisfying (2), (3), (4) and (5) for some $C_W, \kappa \in (0, \infty)$ and $f \in \mathcal{W}$. Then with $R(\theta_0) := m^2(\theta_0)\sigma^2(\theta_0)$, the following conclusions hold:

(a) We have

$$\begin{align*}
\left\{ \mathbb{P}_{\theta_0 + h n^{-1/2}, Q_n} \right\}_{h \in \mathbb{R}} \xrightarrow{\text{Exp}} \left\{ \mathcal{N}(h, R(\theta_0)^{-1}) \right\}_{h \in \mathbb{R}}.
\end{align*}$$

(b) The MLE $\hat{\theta}_n^{\text{MLE}}$ and the MPLE $\hat{\theta}_n^{\text{MPLE}}$ (as defined in section 2.4) exist with probability tending to 1, and have a common asymptotic distribution, given by

$$\sqrt{n}(\hat{\theta}_n^{\text{MLE}} - \theta_0) \xrightarrow{d} \mathcal{N}(0, R(\theta_0)^{-1}), \quad \sqrt{n}(\hat{\theta}_n^{\text{MPLE}} - \theta_0) \xrightarrow{d} \mathcal{N}(0, R(\theta_0)^{-1}).$$

(c) With $\theta_n = \theta_0 + \frac{h}{\sqrt{n}}$ (as in definition 2.2) for some $h > 0$, and $\beta_{MS}$, $\beta_{NP}$ and $\beta_{PL}$ as defined in section 2.5, we have

$$\beta_{NP} = \beta_{PL} = \beta_{MS} = \mathbb{P}(\mathcal{N}(0,1) > z_{\alpha} - h\sqrt{R(\theta_0)}),$$

where $z_{\alpha}$ represents the $(1 - \alpha)^{th}$ quantile for $\mathcal{N}(0,1)$.

**Remark 2.4.** The above theorem shows that for $\theta_0 > 1$, the family of Ising models is LAN at scale $n^{-1/2}$, which is what happens in classical statistics for iid models. It then follows by extension of classical arguments that $\hat{\theta}_n^{\text{MLE}}$ is asymptotically optimal. Perhaps surprisingly, part (b) above shows that $\hat{\theta}_n^{\text{MLE}}$ (which requires significantly less computational resources) is also asymptotically optimal. Carrying this through, it is shown in part (c) that the tests based on $\hat{\theta}_n^{\text{MLE}}$ and $\hat{\theta}_n^{\text{MPLE}}$ have the same asymptotic power. In fact, the much simpler test based on the sample mean $\bar{X}$ also has the same asymptotic power, computation of which does not even require the knowledge of the matrix $Q_n$. Thus in this regime it is possible to gain optimal asymptotic performance for tests of hypothesis without the knowledge of $Q_n$.

2.6.2. *The critical regime* $\Theta_2$. As demonstrated in our next result, the behavior is very different when $\theta_0 = 1$ (which is the critical point). To describe the limit experiment (which is no longer LAN), and the limiting behavior of estimators/tests we make the following definitions.

**Definition 2.5.** Let $\{\mathbb{H}_h(\cdot), h \in \mathbb{R}\}$ be a family of probability distributions on $\mathbb{R}$ parametrized by $h$, with density function

$$p_h(u) = \exp\left(-\frac{1}{12}u^4 + \frac{1}{2}hu^2 - F(h)\right).$$

Here

$$F(h) := \log \int_{\mathbb{R}} \exp\left(-\frac{1}{12}u^4 + \frac{1}{2}hu^2\right) du$$

is the log normalizing constant, which makes $p_h(\cdot)$ into a density. Let $U_h \sim \mathbb{H}_h$. 

Let \( W^* \sim N(0, 2\kappa) \) and \( \{ Y_j \}_{j \geq 2} \overset{iid}{\sim} \chi_1^2 \) be mutually independent. For \( \theta \geq 0 \), define two random variables \( S_\theta \) and \( T_\theta \) by setting

\[
(19) \quad S_\theta := (1 - m^2(\theta)) \left[ \sum_{j=2}^{\infty} \lambda_j \left( \frac{Y_j}{1 - \theta(1 - m^2(\theta))\lambda_j} - 1 \right) - 1 + (1 - m^2(\theta)) \theta \kappa + W^* \right],
\]

\[
(20) \quad T_\theta := (1 - m^2(\theta)) \left[ \sum_{j=2}^{\infty} \frac{\lambda_j^2 Y_j}{1 - \theta(1 - m^2(\theta))\lambda_j} + \kappa \right],
\]

where \( m(\theta) \) is as in definition 2.1. Here the infinite sums above converge in \( L_2 \) (see Lemma 4.1 in the supplementary file). Set

\[
(21) \quad V_h = \frac{1}{3} U_h^2 + \frac{S_1 - T_1}{U_h^2},
\]

where \( U_h \) is independent of \( (S_1, T_1) \), and \( (S_1, T_1) \) is as defined in (19) and (20)).

**Theorem 2.3.** Suppose \( X \sim \mathbb{P}_{\theta, Q_n} \) with \( Q_n \) satisfying (2), (3), (4) and (5) for some \( C_W, \kappa \in (0, \infty) \) and \( f \in \mathcal{W} \). Then the following conclusions hold:

(a) We have

\[
(22) \quad \{ \mathbb{P}_{\theta_n + h n^{-1/2}, Q_n}, h \in \mathbb{R} \} \overset{\text{Exp}}{\rightarrow} \{ \mathbb{H}_h \}, h \in \mathbb{R}.
\]

(b) The MLE \( \hat{\theta}_n^{\text{MLE}} \) and MPLE \( \hat{\theta}_n^{\text{MPLE}} \) as defined in section 2.4 exist with probability tending to 1, and satisfy

\[
(23) \quad \mathbb{P}_{1, Q_n} (\sqrt{n} (\hat{\theta}_n^{\text{MLE}} - 1) \leq h) \rightarrow \mathbb{P} (U_0^2 \leq \mathbb{E} U_0^2),
\]

\[
(24) \quad \sqrt{n} (\hat{\theta}_n^{\text{MPLE}} - 1) \overset{d, \mathbb{P}_{1, Q_n}}{\rightarrow} V_0,
\]

(c) With \( \theta_n = 1 + \frac{h}{\sqrt{n}} \) for some \( h > 0 \) (as in definition 2.2), and \( \beta_{MS}, \beta_{NP} \) and \( \beta_{PL} \) as defined in section 2.5, we have

\[
(25) \quad \beta_{NP} = \beta_{MS} = 2 \mathbb{P} (U_h > \Psi_{H_0} (1 - \alpha/2)),
\]

\[
(26) \quad \beta_{PL} = \mathbb{P} (V_h > \Psi_{V_0} (1 - \alpha)).
\]

**Remark 2.5.** Thus, similar to Theorem 2.2, the contiguous alternatives obtained in Theorem 2.3 in the critical regime is also of size \( O(\frac{1}{\sqrt{n}}) \). However, the limit experiment is no longer gaussian, and so we are outside the familiar LAN setting of classical settings. This is also reflected through the non-gaussian limiting distributions for MLE and MPLE. In terms of tests of hypothesis, part (c) shows that there is a discrepancy between the asymptotic powers of PL-Test and NP-test. The MS-Test continues to be asymptotically optimal, and thus provides a computationally efficient \( Q_n \) agnostic solution to our testing problem. We show in section 2.7 below that \( \beta_{MS} = \beta_{PL} \) for the Curie-Weiss model.

**2.7. Examples.** To illustrate our results, we will now apply our main theorems and simulation results on two concrete examples, the Curie-Weiss model (7), and the Ising model on the complete bi-partite graph \( K_{n/2,n/2} \) (defined in (8)). Table 2.7 compares the asymptotic distribution of the MLE and the MPLE under both the Curie-Weiss Model and the bipartite Ising model. In the low temperature regime \( \Theta_1 \), the asymptotic distribution of both the MLE and MPLE is \( N(0, R(\theta_0)^{-1}) \), for both the Curie-Weiss model and the bipartite Ising model.
Fig 1: The left panel shows the pdf of limiting distributions of MPLE and MLE under Biparite Ising model at critical point, and the right panel shows that under Curie-Weiss model at critical point.

(see Theorem 2.2 part (b)). In fact, the same universal limit continues to hold for any sequence of graphs $G_n$ satisfying (2), (3), (4) and (5). At the critical regime $\Theta_2$, the asymptotic distribution of the two estimators are not the same, for either the Curie-Weiss model (see figure 1a, or the bipartite Ising model (see figure 1b). The simulations in figures 1a and 1b use the limiting distribution obtained in Theorem 2.3 part (b). For the sake of completeness, we note that in the high temperature regime all estimators (and hence MLE and MPLE) are both inconsistent ([6, Theorem 3.3 part (a)]).

| Table 1: Asymptotic distribution of MLE vs MPLE |
|------------------------------------------------|
| Regimes of $\theta$ | Curie-Weiss Model | Bipartite Ising Model |
|----------------------|------------------|----------------------|
| Low Temperature      | normal, same limit | normal, same limit |
| Critical Point       | non-normal, different limit | non-normal, different limit |

In a similar manner, table 2.7 compares the asymptotic powers of the three tests (NP, MS and PL) for the Curie-Weiss Model and the bipartite Ising model. It follows from the main theorems that the three tests have the same asymptotic power if either $\theta \in \Theta_1$ or we are in the Curie-Weiss setting. At criticality the NP test and MS test have the same power, but the PL test has a lower power for the bipartite Ising model (see figure 2a).

| Table 2: Aymptotics powers of tests |
|-------------------------------------|
| Regimes of $\theta$ | Curie-Weiss Model | Bipartite Ising Model |
|----------------------|------------------|----------------------|
| Low Temperature      | $\beta_{NP} = \beta_{MS} = \beta_{PL}$ | $\beta_{NP} = \beta_{MS} = \beta_{PL}$ |
| Critical Point       | $\beta_{NP} = \beta_{MS} = \beta_{PL}$ | $\beta_{NP} = \beta_{MS} > \beta_{PL}$ |

2.8. Main Contributions & Future Scopes. In this paper we establish limits of experiments for a class of one parameter Ising models on dense regular matrices. We show that the limiting experiment is universal (i.e. does not depend on the graph sequence) and LAN in the low temperature regime, and is universal and non LAN in the critical regime. Using the tools developed, we analyze the performance of commonly studied estimators and tests of hypothesis, and compare their performance across different regimes. One surprising discovery is that the asymptotic performance of the MLE and the MPLE is the same in the low temperature regime, thus demonstrating that the extra computational burden of the MLE has
no asymptotic gain in terms of statistical accuracy. In terms of tests of hypothesis, there is a
more computationally efficient test (compared to tests based on either MLE or MPLE) based
on the sample mean, which matches the optimal power function in low and critical regimes.
Prior to this work, such detailed inferential result was largely non existent. We demonstrate
our results by applying them to Ising models on (i) complete graph, and (ii) complete bi-
partite graph.

Throughout this paper we focus on Ising models on dense regular matrices. It would be
interesting to develop inference for Ising models on matrices which are either non regular, or
non dense. Another direction of future interest is to consider the case when the Ising model
has a non-zero magnetic field, and study joint inference for both the inverse temperature
parameter, and the magnetic field. A more challenging problem is to extend these results to
cubic and other higher order interaction models, similar to the Exponential Random Graph
Models of social sciences.

2.9. Outline. The rest of this paper is organized as follows:
In section 3 below, we give the proofs of our the main theorems. The proofs of the theorems
in section 3 rely on two key lemmas about the Ising model (Lemma 3.1 and Lemma 3.2), the
proofs of which are deferred to the appendix (section 4).

3. Proofs of Main Theorems. We begin by stating the following lemmas, which we will
use to prove our main results.

Our first lemma computes the limiting distributions of various quantities under the Ising
model.

Lemma 3.1. Let $X \sim P_{\theta_n, Q_n}$, where $\theta_0 > 0, h \in \mathbb{R}$, and $\theta_n$ is as in definition 2.2. Assume
that the matrix $Q_n$ satisfies (2), (3), (4) and (5) for some $C_W, \kappa \in (0, \infty)$ and $f \in \mathcal{W}$. Also let $S_{\theta_0}, T_{\theta_0}$ be random variables as defined in (19) and (20) respectively. Then, setting
$B_n := Q_n - \frac{1}{n} 11^T$, the following conclusions hold:
(a) If $\theta_0 \in \Theta_1$, then conditional on the set $\bar{X} > 0$ we have
$$\left[\sqrt{n} (\bar{X} - m(\theta_n)), X^T B_n X, X^T B_n^2 X\right] \stackrel{d}{\to} [W_{\theta_0}, S_{\theta_0}, T_{\theta_0}].$$
where \( m(\cdot) \) is as in definition 2.1, and \( W_{\theta_0} \sim N(0, \sigma^2(\theta_0)) \) is independent of \((S_{\theta_0}, T_{\theta_0})\).

(b) If \( \theta_0 \in \Theta_2 \), then
\[
\left[ n^{1/4} \bar{X}, X^T B_n X, X^T B_n^2 X \right] \xrightarrow{d} [U_h, S_{\theta_0}, T_{\theta_0}],
\]
where \( U_h \sim \mathcal{H}_h \) (see (18)) is independent of \((S_{\theta_0}, T_{\theta_0})\).

Our second lemma gives very precise asymptotics for the normalizing constant of Ising models.

**Lemma 3.2.** Let \( \theta_0 > 0, h \in \mathbb{R} \), and let \( \theta_n \) be as in definition 2.2. Assume that the matrix \( Q_n \) satisfies (2), (3), (4) and (5) for some \( C_W, \kappa \in (0, \infty) \) and \( f \in \mathcal{W} \).

(a) If \( \theta_0 > 1 \) then we have
\[
\lim_{n \to \infty} \left\{ Z_n(\theta_n, Q_n) - Z_n(\theta_0, Q_n) - \frac{1}{2} \sqrt{n} h m^2(\theta_0) \right\} = R(\theta_0) h^2.
\]
(b) If \( \theta_0 = 1 \), then we have
\[
\lim_{n \to \infty} \{ Z_n(\theta_n, Q_n) - Z_n(\theta_0, Q_n) \} = F(h) - F(0),
\]
where \( F(h) \) is as defined in (18).

The proofs of Lemma 3.1 and Lemma 3.2 are given in the supplementary file.

Our third result gives an exact characterization of the existence of MLE and MPLE, and show that they exist with high probability.

**Proposition 3.1.** (a) Let
\[
a_n := \min_{x \in \{-1,1\}^n} x^T Q_n x \leq \max_{x \in \{-1,1\}^n} x^T Q_n x =: b_n.
\]
Then the MLE exists in \( \mathbb{R} \) iff
\[
a_n < X^T Q_n X < b_n.
\]
(b) In particular, the MLE exists with probability tending to 1 for all \( \theta_0 \in \Theta_1 \cup \Theta_2 \).
(c) Let
\[
S = S(X) := \{ i : t_i \neq 0 \}.
\]
Then the MPLE exists in \( \mathbb{R} \) iff neither of the following two events happen:
\begin{itemize}
  \item \( X_i = 1 \) for all \( i \in S \).
  \item \( X_i = -1 \) for all \( i \in S \).
\end{itemize}
(d) In particular, the MPLE exists with probability tending to 1 for all \( \theta_0 \in \Theta_1 \cup \Theta_2 \).

The proof of Proposition 3.1 is given after the proof of the main results.

Our final result is a calculus proposition which computes derivatives of the function \( m(\cdot) \) defined in 2.1.

**Proposition 3.2.** Let \( m(\cdot) : (1, \infty) \mapsto (0, 1) \) be as in definition 2.1. Then with \( \theta_n = \theta_0 + \frac{h}{\sqrt{n}} \) (as in definition 2.2) for some \( \theta_0 \in \Theta_1 \) and \( h > 0 \) we have
\[
\lim_{n \to \infty} \frac{m(\theta_n) - m(\theta_0)}{h/\sqrt{n}} = m(\theta_0) \sigma^2(\theta_0).
\]

The proof of Proposition 3.2 is deferred to the supplementary file.
3.1. Proof of Theorem 2.2. Before we begin, we use Lemma 3.1 parts (a) and (b) to conclude that if \((\theta_0, h) \in (\Theta_1 \cup \Theta_2, \mathbb{R})\) we have \(X^T B_n X = O_p(1)\) under \(\mathbb{P}_{\theta_0 + h n^{-1/2}}, Q_n\), and so
\[
(27) \quad X^T Q_n X = n\bar{X}^2 + X^T B_n X = n\bar{X}^2 + O_p(1).
\]

Part (a):
Let \(I = \{h_1, \ldots, h_k\}\) with \(h_1 < h_2 < \ldots < h_k\) for some positive integer \(k\). Then setting \(\theta_{n,i} := \theta_0 + h_i n^{-1/2}\) we have
\[
(28) \quad \log \frac{d\mathbb{P}_{\theta_{n,i}, Q_n}(X)}{d\mathbb{P}_{\theta_0, Q_n}(X)} = Z_n(\theta_0, Q_n) - Z_n(\theta_{n,i}, Q_n) + \frac{h_i \sqrt{n} m^2(\theta_0)}{2} + \frac{h_i}{2} \left( \frac{X^T Q_n X}{\sqrt{n}} - \sqrt{n} m^2(\theta_0) \right).
\]
By part (a) of Lemma 3.2 we have
\[
Z_n(\theta_{n,i}, Q_n) - Z_n(\theta_0, Q_n) - \frac{h_i \sqrt{n} m^2(\theta_0)}{2} \to \frac{R(\theta_0) h_i^2}{2}.
\]
Also, using (27), under \(\mathbb{P}_{\theta_0, Q_n}\) we have
\[
(29) \quad \frac{1}{\sqrt{n}} X^T Q_n X - \sqrt{n} m^2(\theta_0) = \sqrt{n} \left( \bar{X}^2 - m^2(\theta_0) \right) + O_p \left( \frac{1}{\sqrt{n}} \right) \overset{d}{\to} 2m(\theta_0) W_{\theta_0},
\]
where the last step we use part (a) of Lemma 3.1 to conclude that under \(\mathbb{P}_{\theta_0, Q_n}\) we have
\[
(30) \quad \sqrt{n} \left( \bar{X}^2 - m^2(\theta_0) \right) = \sqrt{n} \left( \bar{X} - m(\theta_0) \right) \left( \bar{X} + m(\theta_0) \right) \overset{d}{\to} 2m(\theta_0) W_{\theta_0},
\]
where \(W_{\theta_0} \sim N(0, \sigma^2(\theta_0))\). The above calculation gives that under \(\mathbb{P}_{\theta_0, Q_n}\) we have
\[
\left[ \log \frac{d\mathbb{P}_{\theta_{n,i}}(X)}{d\mathbb{P}_{\theta_0}(X)} \right]_{1 \leq i \leq k} \overset{d}{\to} \left[ - \frac{R(\theta_0) h_i^2}{2} + m(\theta_0) h_i W_{\theta_0} \right]_{1 \leq i \leq k}.
\]
Also note that if \(W \sim N(0, R(\theta_0)^{-1})\) then
\[
\left[ \log \frac{dN(h_i, R(\theta_0)^{-1})(W)}{dN(0, R(\theta_0)^{-1})}(W) \right]_{1 \leq i \leq k} = \left[ - \frac{R(\theta_0) h_i^2}{2} + R(\theta_0) W_{\theta_0} \right]_{1 \leq i \leq k} \overset{d}{=} \left[ - \frac{R(\theta_0) h_i^2}{2} + m(\theta_0) h_i W_{\theta_0} \right]_{1 \leq i \leq k}.
\]
It then follows from the last two displays that under \(\mathbb{P}_{\theta_0, Q_n}\) we have
\[
\left[ \log \frac{d\mathbb{P}_{\theta_{n,i}}(X)}{d\mathbb{P}_{\theta_0}(X)} \right]_{1 \leq i \leq k} \overset{d}{\to} \left[ \log \frac{dN(h_i, R(\theta_0)^{-1})}{dN(0, R(\theta_0)^{-1})}(W) \right]_{1 \leq i \leq k},
\]
which verifies convergence of experiments.

Part (b):
• MLE
Suppose \(X \sim \mathbb{P}_{\theta, Q_n}\). Using Proposition 3.1 part (b), it follows that the MLE exists with probability tending to 1. Also, the MLE \(\hat{\theta}_n^{MLE}\) satisfies the equation
\[
Z_n'(\hat{\theta}_n^{MLE}, Q_n) = \frac{1}{2} X^T Q_n X.
\]
Proceeding to find the limit distribution of \( \hat{\theta}_n^{MLE} \), for any \( h \in \mathbb{R} \) setting \( \theta_n := \theta_0 + h/\sqrt{n} \) we have

\[
\mathbb{P}_{\theta_0, Q_n}(\sqrt{n}(\hat{\theta}_n^{MLE} - \theta_0) \leq h) = \mathbb{P}_{\theta_0, Q_n}(X^T Q_n X - \sqrt{nm^2(\theta_0)} \leq 2Z'_n(\theta_n)).
\]

We now claim that

\[
\mathbb{P}_{\theta_0, Q_n}(\sqrt{n}(\hat{\theta}_n^{MLE} - \theta_0) \leq h) = \mathbb{P}_{\theta_0, Q_n}(\frac{1}{\sqrt{n}}X^T Q_n X - \sqrt{nm^2(\theta_0)} \leq 2Z'_n(\theta_n, Q_n) - \sqrt{nm^2(\theta_0)}).
\]

We proceed to prove (33), using (11), we have

\[
X^T Q_n X = \sum_{i=1}^{n} t_i \tanh(\hat{\theta}_n^{MPLE} t_i) = \sum_{i=1}^{n} t_i \tanh(\theta_0 t_i) + \sum_{i=1}^{n} t_i^2 \text{sech}^2(\xi_n t_i)(\hat{\theta}_n^{MPL} - \theta_0),
\]

where \( \xi_n \) lies between \( \theta_0 \) and \( \hat{\theta}_n^{MPL} \). The last display gives

\[
\sqrt{n}(\hat{\theta}_n^{MPL} - \theta_0) = \frac{1}{\sqrt{n}}(X^T Q_n X - \sum_{i=1}^{n} t_i \tanh(\theta_0 t_i)) \quad \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_i^2 \text{sech}^2(\xi_n t_i).
\]

For analyzing the RHS of (34), use part (a) to note that \( \mathbb{P}_{\theta_0, Q_n} \) we have

\[
\log \frac{d\mathbb{P}_{\theta_0, Q_n}}{d\mathbb{P}_{\theta_0, Q_n}}(X) \sim N\left(-\frac{h^2 R(\theta_0)}{2}, h^2 R(\theta_0)\right).
\]
Then, using Le Cam’s third lemma, it follows that the measures $\mathbb{P}_{\theta_n, Q_n}$ and $\mathbb{P}_{\theta_n, Q_n}$ are mutually contiguous. Along with [13, Lemma 2.1 part (a)], this gives that under $\mathbb{P}_{\theta_n, Q_n}$, we have

$$
\left( \sum_{i=1}^{n} \left[ t_i - m(\theta_0) \right]^2 | \bar{X} > 0 \right) = O_p(1).
$$

On the set $\bar{X} > 0$, under $\mathbb{P}_{\theta_n, Q_n}$ this gives

$$
\sum_{i=1}^{n} t_i \tanh(\theta_0 t_i)
$$

$$
= \sum_{i=1}^{n} \left[ m(\theta_0) \tanh (\theta_0 m(\theta_0)) + m(\theta_0) \left( 1 + \theta_0 (1 - m^2(\theta_0)) \right) \left( t_i - m(\theta_0) \right) \right] + O_p \left( \sum_{i=1}^{n} \left( t_i - m(\theta_0) \right)^2 \right)
$$

$$
= nm^2(\theta_0) + nm(\theta_0) \left( 1 + \theta_0 (1 - m^2(\theta_0)) \right) \left( \bar{X} - m(\theta_0) \right) + O_p(1).
$$

Since (27) gives $X^T Q_n X = n\bar{X}^2 + O_p(1)$, the above display implies that under $\mathbb{P}_{\theta_n, Q_n}$,

$$
X^T Q_n X - \sum_{i=1}^{n} t_i \tanh(\theta_0 t_i)
$$

$$
= n \left( \bar{X}^2 - m^2(\theta_0) \right) - nm(\theta_0) \left( 1 + \theta_0 (1 - m^2(\theta_0)) \right) \left( \bar{X} - m(\theta_0) \right) + O_p(1)
$$

$$
= n \left( \bar{X} - m(\theta_0) \right) \left[ \bar{X} - \theta_0 m(\theta_0) (1 - m^2(\theta_0)) \right] + O_p(1).
$$

Also, under $\mathbb{P}_{\theta_n, Q_n}$ we have

$$
\sqrt{n} \left( \bar{X} - m(\theta_0) \right) = \sqrt{n} \left( \bar{X} - m(\theta_n) \right) + \sqrt{n} \left( m(\theta_n) - m(\theta_0) \right) \overset{d}{\to} N \left( m(\theta_0) \sigma^2 (\theta_0) n, \sigma^2 \right),
$$

where the last step uses part (a) of Lemma 3.1 along with Proposition 3.2. The last two displays give that under $\mathbb{P}_{\theta_n, Q_n}$,

$$
\frac{1}{\sqrt{n}} \left( X^T Q_n X - \sum_{i=1}^{n} t_i \tanh(\theta_0 t_i) \right) \overset{d}{\to} \left( 1 - \theta_0 (1 - m^2(\theta_0)) \right) N \left( hm^2(\theta_0) \sigma^2 (\theta_0), m^2(\theta_0) \sigma^2 (\theta_0) \right).
$$

Also, using contiguity of the two measures $\mathbb{P}_{\theta_n, Q_n}$ and $\mathbb{P}_{\theta_n, Q_n}$ along with consistency of $\hat{\theta}_n^{MLE}$ gives $\xi_n \overset{P}{\to} 0$ under both measures, and so

$$
\frac{1}{n} \sum_{i=1}^{n} t_i^2 \text{sech}^2(\xi_n t_i) \overset{P}{\to} m^2(\theta_0) \text{sech}^2(\theta_0 m(\theta_0)) = m^2(\theta_0) \left( 1 - m^2(\theta_0) \right).
$$

Combining the last two displays along with (34) we get that under $\mathbb{P}_{\theta_n, Q_n}$,

$$
\sqrt{n} \left( \hat{\theta}_n^{MLE} - \theta_0 \right) \overset{d}{\to} \left( 1 - \theta_0 (1 - m^2(\theta_0)) \right) N \left( hm^2(\theta_0) \sigma^2 (\theta_0), m^2(\theta_0) \sigma^2 (\theta_0) \right)
$$

$$
= \frac{1}{m^2(\theta_0) \sigma^2 (\theta_0)} N \left( h m^2(\theta_0) \sigma^2 (\theta_0), m^2(\theta_0) \sigma^2 (\theta_0) \right),
$$

which is equivalent to (33).

**Part (c):**
• **MS Test**
  If $X \sim \mathbb{P}_{\theta_0, Q_n}$, then (30) gives
  $$K_n(\alpha) = nm^2(\theta_0) + 2z_\alpha \sqrt{nR(\theta_0)} + o(\sqrt{n}).$$
  Also, using (35), under $\mathbb{P}_{\theta_0, Q_n}$ we have
  $$\sqrt{n} \left( \bar{X}^2 - m^2(\theta_0) \right) \xrightarrow{d} 2m(\theta_0)N \left( m(\theta_0)\sigma^2(\theta_0)h, \sigma^2 \right) = N(2m^2(\theta_0)\sigma^2(\theta_0)h, 4m^2(\theta_0)\sigma^2(\theta_0)).$$
  Using the last two displays, we have
  $$\mathbb{P}_{\theta_0, Q_n}(n\bar{X}^2 > K_n(\alpha)) = \mathbb{P}_{\theta_0, Q_n} \left( \sqrt{n} \left( \bar{X}^2 - m^2(\theta_0) \right) > 2z_\alpha \sqrt{R(\theta_0)} \right) + o(1)$$
  $$= \mathbb{P} \left( N(2R(\theta_0)h, 4R(\theta_0)) > 2z_\alpha \sqrt{R(\theta_0)} \right) + o(1)$$
  $$= \mathbb{P} \left( N(h\sqrt{R(\theta_0)}, 1) > z_\alpha \right) + o(1).$$
  Thus we have $\beta_{MS} = 1 - \Phi(z_\alpha - h\sqrt{R(\theta_0)})$, as claimed.

• **NP Test**
  Note that
  $$X^T Q_n X = n\bar{X}^2 + O_p(1),$$
  using Lemma 3.1 part (a) under $\mathbb{P}_{\theta_0, Q_n}$ and $\mathbb{P}_{\theta, Q_n}$. Thus $X^T Q_n X$ has the same asymptotic distribution as $n\bar{X}^2$, under both null and the alternative. This gives $\beta_{NP} = \beta_{MS}$, as desired.

• **PL Test**
  Using the limit distribution of $\hat{\theta}_n^{MPLE}$ in part (b) (invoke (33) under $\mathbb{P}_{\theta_0, Q_n}$) gives
  $$K_n(\alpha) = \theta_0 + \frac{z_\alpha}{\sqrt{nR(\theta_0)}} + o\left( \frac{1}{\sqrt{n}} \right).$$
  Thus we have
  $$\mathbb{P}_{\theta_0, Q_n}(\hat{\theta}_n^{MPLE} > K_n(\alpha)) = \mathbb{P}_{\theta_0, Q_n} \left( \sqrt{n}(\hat{\theta}_n^{MPLE} - \theta_0) > \frac{z_\alpha}{\sqrt{R(\theta_0)}} \right) + o(1)$$
  $$= \mathbb{P} \left( N\left( h, \frac{1}{\sqrt{R(\theta_0)}} \right) > \frac{z_\alpha}{\sqrt{R(\theta_0)}} \right) + o(1),$$
  where the last step again uses (33) under $\mathbb{P}_{\theta_0, Q_n}$. The last term in the display above converges to
  $$\mathbb{P} \left( N(h\sqrt{R(\theta_0)}, 1) > z_\alpha \right) = 1 - \Phi(z_\alpha - h\sqrt{R(\theta_0)}),$$
  thus giving the same formula for $\beta_{PL}$.

3.2. **Proof of Theorem 2.3.**

**Part (a):**
As in the proof of Theorem 2.2, let $I := \{h_1, \ldots, h_k\}$ with $h_1 < h_2 < \ldots < h_k$ for some positive integer $k$, and let $\theta_{n,i} := \theta_0 + \frac{h_i}{\sqrt{n}}$ for $1 \leq i \leq k$. It thus suffices to analyze the terms in the RHS of (28). To this effect, use Lemma 3.2 part (b) to get
$$Z_n(\theta_{n,i}, Q_n) - Z_n(\theta_0, Q_n) \to F(h) - F(0).$$
Proceeding to analyze the second term in the RHS of (28), using (27), under $P_{\theta_0, Q_n}$ we have

$$\frac{X^T Q_n X}{\sqrt{n}} = \sqrt{n} \bar{X}^2 + O_p \left( \frac{1}{\sqrt{n}} \right) \xrightarrow{d} U_0^2,$$

where the last step uses part (b) of Lemma 3.1. Combining the last two displays along with (28), under $P_{\theta_0, Q_n}$ we have

$$\left[ \log \frac{dP_{\theta_n, Q_n}}{dP_{\theta_0, Q_n}} (X) \right] \xrightarrow{d} \left[ -F(h_i) + F(0) + \frac{h_i}{2} U_0^2 \right]_{1 \leq i \leq k}.$$

Also, we have

$$\log \frac{dH_{h_i}}{dH_0} (u) = \frac{1}{2} h_i u - F(h_i) + F(0).$$

Combining the above two displays, under $P_{\theta_0, Q_n}$ we have

$$\left[ \log \frac{dP_{\theta_n, Q_n}}{dP_{\theta_0, Q_n}} (X) \right] \xrightarrow{d} \left[ \frac{dH_{h_i}}{dH_0} (U_0) \right]_{1 \leq i \leq k},$$

where $U_0 \sim H_0$. This verifies the desired convergence of experiments.

**Part (b):**

- **MLE**
  
  Existence of MLE follows from Proposition 3.1 part (b). Proceeding to find the limiting distribution, for any $h \in \mathbb{R}$ setting $\theta_n := 1 + \frac{h}{\sqrt{n}}$, differentiating the second display in part (b) of Lemma 3.2 with respect to $h$ we get

$$\frac{2 Z_n' (\theta_n, Q_n)}{\sqrt{n}} = \mathbb{E} U_h^2 + o(1).$$

Consequently, using calculations similar to the proof of Theorem 2.2 part (b) we have

$$\mathbb{P}_{1, Q_n} \left( \sqrt{n} (\hat{\theta}_n^{MLE} - 1) \leq h \right) = \mathbb{P}_{1, Q_n} \left( X^T Q_n X \leq 2 Z_n' (\theta_n, Q_n) \right)$$

$$= \mathbb{P}_{1, Q_n} \left( \frac{1}{\sqrt{n}} X^T Q_n X \leq \mathbb{E} U_h^2 + o(1) \right)$$

$$= \mathbb{P} \left( U_0^2 \leq \mathbb{E} U_h^2 + o(1) \right),$$

which derives the limiting distribution of the MLE.

- **MLE**
  
  Existence of the MLE follows from Proposition 3.1 part (d). Proceeding to show consistency, use (11) to note that on the set $\hat{\theta}_n^{MLE} > 1 + \delta$ under $P_{1, Q_n}$ we have

$$X^T Q_n X = \sum_{i=1}^{n} t_i \tanh(\hat{\theta}_n^{MLE} t_i)$$

$$\geq \sum_{i=1}^{n} t_i \tanh((1 + \delta) t_i)$$

$$= n \bar{X} \tanh((1 + \delta) \bar{X}) + O_p \left( \sum_{i=1}^{n} (t_i - \bar{X})^2 \right).$$
Using Lemma 3.1 part (b), under $\mathbb{P}_{1,Q_n}$, we get
\[ X^T B_n X = O_p(1), \quad \sum_{i=1}^{n} (t_i - \bar{x})^2 = X^T B_n^2 X = O_p(1). \]

Combining the last two displays, under $\mathbb{P}_{1,Q_n}$, we have
\[ \sqrt{n}X^2 \geq \sqrt{n}X \tanh((1 + \delta)X) + O_p\left(\frac{1}{\sqrt{n}}\right). \]

But this is a contradiction, as the LHS of the above display converges in distribution $U_0^2$ under $\mathbb{P}_{1,Q_n}$ (by Lemma 3.1 part (b)), and the RHS converges in distribution to $(1 + \delta)U_0^2$ (which is larger). Thus we have shown that for any $\delta > 0$ we have $\mathbb{P}_{1,Q_n}(\hat{\theta}_n^{MLE} \geq 1 + \delta) \to 0$. A similar proof gives $\mathbb{P}_{1,Q_n}(\hat{\theta}_n^{MLE} \leq 1 - \delta) \to 0$, and so $\hat{\theta}_n^{MLE} \overset{p}{\to} 1$.

Proceeding to find the limiting distribution of $\hat{\theta}_n^{MLE}$, setting $\theta_n := 1 + \frac{h}{\sqrt{n}}$ for some $h \in \mathbb{R}$ we work under the measure $\mathbb{P}_{\theta_n,Q_n}$, and claim that
\[ (37) \quad \sqrt{n}(\hat{\theta}_n^{MLE} - 1) \overset{d}{\to} V_h. \]

As before, the desired limiting distribution for $\sqrt{n}(\hat{\theta}_n^{MLE} - 1)$ under $\mathbb{P}_{1,Q_n}$ then follows on taking $h = 0$.

For proving (37), use part (a) to conclude that the measures $\mathbb{P}_{1,Q_n}$ and $\mathbb{P}_{\theta_n,Q_n}$ are mutually contiguous (as in the proof of Theorem 2.2 part (b)), and so $\hat{\theta}_n^{MLE} \overset{p}{\to} 1$ under $\mathbb{P}_{\theta_n,Q_n}$ as well. Further, using (11), under $\mathbb{P}_{\theta_n,Q_n}$, we have
\[
\begin{align*}
\sum_{i=1}^{n} t_i \tanh(\hat{\theta}_n^{MLE} t_i) &= \sum_{i=1}^{n} t_i \left[ \tanh(t_i) + (\hat{\theta}_n^{MLE} - 1)t_i \sech^2(t_i) + O_p\left((\hat{\theta}_n^{MLE} - 1)^2|t_i|^3\right) \right].
\end{align*}
\]

Under $\mathbb{P}_{\theta_n,Q_n}$, this gives
\[ (38) \quad \hat{\theta}_n^{MLE} - 1 = \frac{X^T Q_n X - \sum_{i=1}^{n} t_i \tanh(t_i)}{\sum_{i=1}^{n} t_i^2 \sech^2(t_i) + O_p((\hat{\theta}_n^{MLE} - 1)^2 \sum_{i=1}^{n} t_i^4)}. \]

For analyzing the numerator in (38), under $\mathbb{P}_{\theta_n,Q_n}$ we have
\[
\begin{align*}
X^T Q_n X - \sum_{i=1}^{n} t_i \tanh(t_i) &= n\bar{x}^2 + X^T B_n X - n\bar{x} \tanh(\bar{x}) - \sum_{i=1}^{n} (t_i - \bar{x})^2 + O_p\left(\sum_{i=1}^{n} |t_i - \bar{x}|^3\right) \\
&= \frac{n\bar{x}^4}{3} + X^T B_n X - X^T B_n^2 X + O_p\left(n \sqrt{\frac{\log n}{n}}\right) + O_p\left(n^{1 - \frac{\epsilon}{4}}\right) \\
&= \frac{n\bar{x}^4}{3} + X^T B_n X - X^T B_n^2 X + o_p(1),
\end{align*}
\]
where in the last but one step we use Lemma 3.1 part (b) and [13, Lemma 2.4], along with mutual contiguity shown above, to get that under $P_{\theta_n,Q_n}$ we have

$$|\bar{X}| = O_p(n^{-1/4}), \quad \max_{i \in [n]} |t_i - \bar{X}| = O_p\left(\sqrt{\frac{\log n}{n}}\right).$$

For the denominator in (38), under $P_{\theta_n,Q_n}$ we have

$$\sum_{i=1}^{n} t_i^2 \text{sech}^2(t_i) + O_p\left(\left(\hat{\theta}_n^{\text{MLE}} - 1\right) \sum_{i=1}^{n} t_i^4\right)$$

$$= n\bar{X}^2 + \sum_{i=1}^{n} (t_i - \bar{X})^2 + O_p\left(\sum_{i=1}^{n} (t_i - \bar{X})^4 + n\bar{X}^4\right)$$

$$= n\bar{X}^2 + O_p(1),$$

where the last step uses Lemma 3.1 part (a), and mutual contiguity. Combining the above two displays along with (38), under $P_{\theta_n,Q_n}$ we get

$$\sqrt{n}(\hat{\theta}_n^{\text{MLE}} - 1) \xrightarrow{d} \frac{n\bar{X}^4 + \bar{X}^T B_n X - \bar{X}^T B_n^2 X + o_p(1)}{\sqrt{n}\bar{X}^2 + o_p(1)}$$

$$\xrightarrow{d} \frac{U_h^2}{\frac{1}{3} - S_1 - T_1} \frac{U_h}{U_h^2},$$

where we again use part (b) of Lemma 3.1. Recalling the formula of $V_h$ we have verified (37), and this completes part (b).

**Part (c):**

**MS Test**

Using symmetry of the distribution of $\bar{X}$ we have

$$\alpha = \mathbb{P}_{1,Q_n}(n\bar{X}^2 > K_n(\alpha)) = 2\mathbb{P}_{1,Q_n}(\bar{X} > \sqrt{K_n(\alpha)}).$$

Using this, along with the limit distribution $n^{1/4}\bar{X} \xrightarrow{d} U_0$ under $P_{1,Q_n}$ (see part (b) of Lemma 3.1) gives

$$\sqrt{K_n(\alpha)} = n^{1/4}\Psi_{E_0}(1 - \alpha/2) + o(n^{1/4}).$$

Then, setting $\theta_n = 1 + \frac{h}{\sqrt{n}}$, the asymptotic power is given by

$$2\mathbb{P}_{\theta_n}(\sqrt{n}\bar{X} > K_n(\alpha)) = 2\mathbb{P}_{\theta_n,Q_n}\left(n^{1/4}\bar{X} > \Psi_{E_0}(1 - \alpha/2) + o(1)\right)$$

$$= 2\mathbb{P}\left(U_h > \Psi_{E_0(1 - \alpha/2)}\right) + o(1),$$

where the last line again uses part (b) of Lemma 3.1. Thus we have $\beta_{MS} = 2\mathbb{P}(U_h > \Psi_{E_0}(1 - \alpha/2))$, as desired.

**NP Test**

As in the proof of Theorem 2.2, we have $X^T Q_n X = n\bar{X}^2 + O_p(1)$ under both $P_{1,Q_n}$ and $P_{\theta_n,Q_n}$, using (27) and mutual contiguity. Thus we get $\beta_{NP} = \beta_{MS}$ as before.
**PL Test**

Using (37), under \( P_{1,Q_n} \) we have

\[
\sqrt{n}(\hat{\theta}_n^{MLE} - 1) \xrightarrow{d} V_0,
\]

which gives \( K_n(\alpha) = 1 + \frac{\Psi_{V_0}(\alpha)}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right) \).

Then the asymptotic power is given by

\[
P_{\theta_n,Q_n}(\hat{\theta}_n > K_n(\alpha)) = 1 + \Psi_{V_0}(\alpha) + o(1),
\]

where the last step again uses (37). This shows that \( \beta_{PL} = \mathbb{P}(V_h > \Psi_{V_0}(\alpha)) \), as desired.

### 3.3. Proof of Proposition 3.1.

(a) By definition of \( a_n, b_n \) we have

\[
\lim_{\theta \to -\infty} \mathbb{E}_\theta X'Q_nX = a_n, \quad \lim_{\theta \to \infty} \mathbb{E}_\theta X'Q_nX = b_n.
\]

Since the function

\[
\theta \mapsto \frac{\theta}{2} X'Q_nX - Z_n(\theta, Q_n)
\]

is strictly concave, it follows that there exists a unique MLE in \( \mathbb{R} \) iff the equation

\[
X'Q_nX = \mathbb{E}_\theta X'Q_nX
\]

has a real solution, which holds iff \( a_n < X'TQ_nX < b_n \), as desired.

(b) This is immediate from part (a), and on noting that \( X'TQ_nX \) has a continuous limiting distribution in all regimes, as shown in the proofs above (in particular, see (29) and (36) for domains \( \Theta_1 \) and \( \Theta_2 \) respectively).

(c) Using (11), the existence of MPLE is equivalent to the existence of a real valued root of the equation

\[
X'TQ_nX = \sum_{i=1}^{n} t_i \tanh(\theta t_i).
\]

Taking limits as \( \theta \to \pm \infty \), we get

\[
\lim_{\theta \to -\infty} \sum_{i=1}^{n} t_i \tanh(\theta t_i) = -\sum_{i=1}^{n} |t_i|, \quad \lim_{\theta \to \infty} \sum_{i=1}^{n} t_i \tanh(\theta t_i) = \sum_{i=1}^{n} |t_i|.
\]

Thus the existence of MPLE holds iff

\[
-\sum_{i=1}^{n} |t_i| < X'TQ_nX < \sum_{i=1}^{n} |t_i|.
\]

Suppose \( X'TQ_nX = \sum_{i=1}^{n} t_i \). This happens iff \( X_i = 1 \) for all \( i \in S(X) \). Similarly we have

\[
X'TQ_nX = -\sum_{i=1}^{n} t_i \iff X_i = -1 \text{ for all } i \in S(X).
\]

The conclusion of part (c) follows from this.

(d) By symmetry, we only show that

\[
\mathbb{P}_{\theta_0,Q_n}(X_i = 1, i \in S) \to 0.
\]
To this end, note that by mutual contiguity it suffices to show the result under the Curie-Weiss model. To this effect, we first claim that for any positive sequence \( \{ \varepsilon_n \}_{n \geq 1} \) converging to 0 and constant \( C > 0 \) free of \( n \), we have

\[
\lim_{n \to \infty} \sup_{a \in \mathbb{R}^n / \{0\} : \|a\|_\infty \leq \varepsilon_n \|a\|_2} \mathbb{P}\left( \sum_{i=1}^{n} a_i \xi_i = 0 \right) = 0,
\]

where \( \{ \xi_i \}_{1 \leq i \leq n} \) are iid random variables such that \( \operatorname{Var}(\xi_1) \neq 0 \), \( \mathbb{E}|\xi_1|^3 \leq C \).

Given this claim, we first complete the proof of part (d). Let \( \phi_n \) be the auxiliary variable introduced in Proposition. Then we have

\[
\mathbb{P}_{\theta_0, \text{CW}}(t_i = 0 \text{ for some } i, 1 \leq i \leq n | \phi_n) = \mathbb{E}\mathbb{P}_{\theta_0, \text{CW}}(t_i = 0 \text{ for some } i, 1 \leq i \leq n | \phi_n).
\]

Given \( \phi_n \), \( t_i \) is a weighted sum of iid random variables, with \( \mathbb{E}(|X_1|^3 | \phi_n) \leq 1 \). Also, we have

\[
|Q_n(i,j)| \leq C_w \sqrt{n} \sum_{j=1}^{n} Q_n(i,j)^2 = \frac{1}{\sqrt{d_n}},
\]

and so we can take \( C = 1, \varepsilon_n = \frac{C'}{\sqrt{n}} \) for some suitable constant \( C' \) free of \( n \). Thus we have

\[
\mathbb{P}_{\theta_0, \text{CW}}(t_i = 0 \text{ for some } i, 1 \leq i \leq n | \phi_n) \xrightarrow{p} 0 \Rightarrow \lim_{n \to \infty} \mathbb{P}_{\theta_0, \text{CW}}(t_i = 0 \text{ for some } i) = 0.
\]

To complete the proof, it suffices to show that \( \mathbb{P}_{\theta_0, \text{CW}}(\bar{X} = 1) \to 0 \). But this is immediate from the weak law of \( \bar{X} \) derived in Lemma 3.1 (and mutual contiguity of \( \mathbb{P}_{\theta_0, Q_n} \) and \( \mathbb{P}_{\theta_0, \text{CW}} \)).

It thus remains to verify the claim. But this follows on setting \( \mu_n := \mathbb{E}\xi_1, \tau_n^2 = \operatorname{Var}(\xi_1) \) and noting that

\[
\mathbb{P}\left( \sum_{i=1}^{n} a_i \xi_i = 0 \right) = \mathbb{P}\left( \frac{\sum_{i=1}^{n} a_i (\xi_i - \mu_n)}{\tau_n \|a\|_2} = -\frac{\mu_n \sum_{i=1}^{n} a_i}{\tau_n \|a\|_2} \right),
\]

and the fact that

\[
\frac{\sum_{i=1}^{n} a_i (\xi_i - \mu_n)}{\tau_n \|a\|_2} \xrightarrow{d} N(0, 1)
\]

by the Lyapunov CLT.

Acknowledgments. We thank Nabarun Deb and Rajarshi Mukherjee for helpful comments throughout this project. We also thank Richard Nickl for suggesting this problem. The presentation of the paper greatly benefitted from the suggestions of an anonymous referee, and the Associate Editor.

Funding. The second author was supported in part by NSF (DMS-2113414).

REFERENCES
[1] Basak, A. and Mukherjee, S. (2017). Universality of the mean-field for the Potts model. Probability Theory and Related Fields 168 557–600.
[2] Berthet, Q., Rigollet, P. and Srivastava, P. (2019). Exact recovery in the Ising blockmodel. The Annals of Statistics 47 1805–1834.
[3] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society: Series B (Methodological) 36 192–225.
Lemma 3.2. In section 4.3 we use the two auxiliary results from section 4.2 to verify Lemma 3.1 and general results. In section 4.2 we use the results from 4.1 to prove two auxiliary lemmas. Finally in section 4.3 we use the two auxiliary results from section 4.2 to verify Lemma 3.1 and Lemma 3.2.

4. Appendix. The appendix is organized as follows: In section 4.1 we prove some general results. In section 4.2 we use the results from 4.1 to prove two auxiliary lemmas. Finally in section 4.3 we use the two auxiliary results from section 4.2 to verify Lemma 3.1 and Lemma 3.2.

4.1. Proofs of Independent Results.

4.1.1. Limit distribution for IID quadratic forms. We first state a proposition connecting eigenvalues of the matrix $Q_n$ and eigenvalues of the limiting graphon $f$. 

[4] Besag, J. (1975). Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician)* 24 179–195.
[5] Bhattacharya, B. B., Diaconis, P. and Mukherjee, S. (2017). Universal limit theorems in graph coloring problems with connections to extremal combinatorics. *The Annals of Applied Probability* 27 337–394.
[6] Bhattacharya, B. B. and Mukherjee, S. (2018). Inference in Ising models. *Bernoulli* 24 493–525.
[7] Bhattacharya, B. B., Das, S., Mukherjee, S. and Mukherjee, S. (2022). Asymptotic Distribution of Random Quadratic Forms. *arXiv preprint arXiv:2203.02850*.
[8] 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. *Advances in Mathematics* 219 1801–1851.
[9] 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. *Annals of Mathematics* 151–219.
[10] Chatterjee, S. (2007). Estimation in spin glasses: A first step. *The Annals of Statistics* 35 1931–1946.
[11] Chatterjee, S., Diaconis, P. and Sly, A. (2011). Random graphs with a given degree sequence. *Annals of Applied Probability* 21 1400–1435.
[12] Comets, F. and Gidas, B. (1991). Asymptotics of maximum likelihood estimators for the Curie-Weiss model. *The Annals of Statistics* 557–578.
[13] Deb, N. and Mukherjee, S. (2023+). Fluctuations in mean-fields Ising models. *to appear in Annals of Applied Probability*.
[14] Deb, N., Mukherjee, R., Mukherjee, S. and Yuan, M. (2023+). Detecting Structured Signals in Ising Models. *to appear in Annals of Applied Probability*.
[15] Dembo, A. and Montanari, A. (2010). Ising models on locally tree-like graphs. *The Annals of Applied Probability* 20 565–592.
[16] Ellis, R. S. and Newman, C. M. (1978). The statistics of Curie-Weiss models. *Journal of Statistical Physics* 19 149–161.
[17] Ghosal, P. and Mukherjee, S. (2020). Joint estimation of parameters in Ising model. *The Annals of Statistics* 48 785–810.
[18] Gray, R. M. et al. (2006). Toeplitz and circulant matrices: A review. *Foundations and Trends® in Communications and Information Theory* 2 155–239.
[19] Harris, A. B. (1974). Effect of random defects on the critical behaviour of Ising models. *Journal of Physics C: Solid State Physics* 7 1671.
[20] Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik* 31 253–258.
[21] Kabluchko, Z., Löwe, M. and Schubert, K. (2019). Fluctuations of the magnetization for Ising models on dense Erdős-Renyi random graphs. *Journal of Statistical Physics* 177 78–94.
[22] Le Cam, L. et al. (1972). Limits of experiments. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* 1 245–261. University of California Press Berkeley-Los Angeles.
[23] Lovász, L. (2012). *Large networks and graph limits* 60. American Mathematical Soc.
[24] Mukherjee, R., Mukherjee, S. and Yuan, M. (2018). Global testing against sparse alternatives under Ising models. *The Annals of Statistics* 46 2062–2093.
[25] Onsager, L. (1944). Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Physical Review* 65 117.
[26] van der Vaart, A. W. (2000). *Asymptotic statistics* 3. Cambridge university press.
[27] Xu, Y. and Mukherjee, S. (2022). Ising Models on Dense Regular Graphs. *arXiv preprint arXiv:2210.13178v1*.
PROPOSITION 4.1. Let \( \{Q_n\}_{n=1}^{\infty} \) be a sequence of matrices satisfying (2), (3), (4) and (5) for some \( C_W, \kappa \in (0, \infty) \) and \( f \in \mathcal{W} \). Let \( \{\lambda_{j,n}\}_{j=1}^{n} \) denote the eigenvalues of \( Q_n \) arranged in decreasing order of absolute value, and let \( \{\lambda_j\}_{j \geq 1} \) be the eigenvalues of the operator \( T_f \) as defined in section 2.2. Then the following conclusions hold:

(a) \[
\sum_{j=1}^{\infty} \lambda_j^2 = \int_{[0,1]^2} f(x,y)^2 \, dx \, dy = \|f\|_2^2 < \infty;
\]

(b) For any \( j \in \mathbb{N} \),
\[
\lim_{n \to \infty} \lambda_{j,n} = \lambda_j.
\]

(c) For any \( i \geq 3 \) we have
\[
\lim_{n \to \infty} \sum_{j=2}^{n} \lambda_{j,n}^i = \sum_{j=2}^{\infty} \lambda_j^i.
\]

PROOF. The proof of Proposition 4.1 part (a) follows [23, Chapter 7.5], whereas part (b) and (c) follow from [23, Theorem 11.54].

Utilizing Proposition 4.1, we now characterize the limit distribution of quadratic forms of IID random variables, which is a crucial result in our analysis, and is of possible independent interest. For interested readers, we refer to [5, Theorem 1.4] and [7, Theorem 1.4] for results with a similar flavor.

LEMMA 4.1. Suppose \( Z := (Z_i)_{1 \leq i \leq n} \) are IID random variables with mean 0 and variance \( \tau_n^2 \) which converges to \( \tau^2 \in (0, \infty) \). Assume that the matrix \( Q_n \) satisfies (2), (3), (4) and (5) for some \( C_W, \kappa \in (0, \infty) \) and \( f \in \mathcal{W} \). Then with \( B_n = Q_n - \frac{1}{n}11^T \) we have
\[
[\sqrt{n}Z, Z^T B_n Z, Z^T B_n^2 Z] \xrightarrow{d} \left[ \tau W_0, \tau^2 \left( \sum_{j=2}^{\infty} \lambda_j (Y_j - 1) - 1 + W^* \right), \tau^2 \left( \sum_{j=2}^{\infty} \lambda_j^2 Y_j + \kappa \right) \right],
\]
where
\[
W_0 \sim N(0,1), \quad W^* \sim N(0,2\kappa), \quad \{Y_j\}_{j \geq 2} \text{ iid } \sim \chi^2_1
\]
are mutually independent. Here the infinite sums in the limiting distributions converge in \( L_2 \).

PROOF. The proof of Lemma 4.1 will be completed, once we can show the following two steps:

(a) Suppose \( \{R_i\}_{1 \leq i \leq n} \) are IID \( N(0,1) \). Then the desired conclusion holds.

(b) For any positive integers \( a, b, c \) we have
\[
\mathbb{E}\left( \sqrt{n}Z \right)^a (Z^T B_n Z)^b (Z^T B_n^2 Z)^c - \mathbb{E}\left( \sqrt{n}R \right)^a (R^T B_n R)^b (R^T B_n^2 R)^c \to 0.
\]

PROOF OF (A). Let
\[
Q_n = P^T A P = \sum_{i=1}^{n} \lambda_{i,n} P_i P_i^T
\]
be the spectral decomposition of $Q_n$, where the eigenvalues $\{\lambda_{i,n}\}_{1 \leq i \leq n}$ are arranged in decreasing order of absolute values. Thus we have $\lambda_{1,n} = 1$, and $p_1 = \frac{1}{\sqrt{n}}$. Then setting $\bar{R} := PR$ we have

$$\begin{bmatrix} \sqrt{n}\bar{R}, R^T B_n R, R^T B_n^2 R \end{bmatrix} = \begin{bmatrix} \bar{R}_1, \sum_{i=2}^{n} \lambda_{i,n} \bar{R}_i^2, \sum_{i=2}^{n} \lambda_{i,n}^2 \bar{R}_i^2 \end{bmatrix}.$$  

Since $\bar{R} \overset{d}{=} R$, it suffices to find the limiting distribution of

$$\begin{bmatrix} R_1, \sum_{i=2}^{n} \lambda_{i,n} R_i^2, \sum_{i=2}^{n} \lambda_{i,n}^2 R_i^2 \end{bmatrix}.$$  

Clearly, $R_1$ is independent of the other two random variables, and has a $N(0,1)$ distribution. It thus suffices to focus on the joint distribution of the other two random variables. To this effect, for any $s, t$ with $\max(|s|, |t|) \leq \frac{1}{8}$ we have

$$\log \mathbb{E} \exp \left\{ s \sum_{j=2}^{n} \lambda_{j,n} R_j^2 + t \sum_{j=2}^{n} \lambda_{j,n}^2 R_j^2 \right\}$$

$$= -\frac{1}{2} \sum_{j=2}^{n} \log \left[ 1 - 2(\lambda_{j,n}s + \lambda_{j,n}^2 t) \right]$$

$$= \frac{1}{2} \sum_{j=2}^{n} \sum_{i=1}^{\infty} \frac{2^i (\lambda_{j,n}s + \lambda_{j,n}^2 t)^i}{i}$$

$$= \left[ -s + t \sum_{j=2}^{n} \lambda_{j,n}^2 \right] + \frac{1}{2} \sum_{j=2}^{n} \sum_{i=2}^{\infty} \frac{2^i (\lambda_{j,n}s + \lambda_{j,n}^2 t)^i}{i}$$

$$= \left[ -s + t \sum_{j=2}^{n} \lambda_{j,n}^2 \right] + \frac{1}{2} \sum_{i=2}^{\infty} \sum_{j=2}^{n} \frac{2^i (\lambda_{j,n}s + \lambda_{j,n}^2 t)^i}{i},$$

where the last line uses Fubini’s theorem, along with the trivial bound

$$2^i (\lambda_{j,n}s + \lambda_{j,n}^2 t)^i \leq 4^i |\lambda_{j,n}|^i 8^{-i} \leq 2^{-i} \lambda_{j,n}^2.$$  

(42)

For every fixed $i \geq 2$, uses Proposition 4.1 part (b) we have

$$\sum_{j=2}^{n} (\lambda_{j,n}s + \lambda_{j,n}^2 t)^i \to \sum_{j=2}^{\infty} (\lambda_j s + \lambda_j^2 t)^i.$$  

Also, using (42) we have

$$\sum_{j=2}^{n} \frac{2^i (\lambda_{j,n}s + \lambda_{j,n}^2 t)^i}{i} \leq 2^{-i} \sum_{j=2}^{n} \lambda_{j,n}^2,$$

where

$$\sum_{j=2}^{n} \lambda_{j,n}^2 \to \sum_{j=2}^{\infty} \lambda_j^2 < \infty.$$
by Proposition 4.1 part (c). Combining the last three displays along with dominated convergence theorem, the RHS of (41) converges to
\[
\left[ -s + t \sum_{j=2}^{\infty} \lambda_j^2 \right] + \frac{1}{2} \sum_{j=2}^{\infty} \sum_{i=2}^{\infty} \frac{2^i (\lambda_j s + \lambda_j^2 t)^i}{i}. \]

This is the log moment generating function of
\[
\left( \sum_{j=2}^{\infty} \lambda_j (Y_j - 1) - 1 + W^* \sum_{j=2}^{\infty} \lambda_j^2 Y_j + \kappa \right).
\]

The random variables in the RHS above converge in $L_2$, as
\[
\begin{align*}
E \left[ \sum_{j=k+1}^{\infty} \lambda_j (Y_j - 1) \right]^2 &= 2 \sum_{j=k+1}^{\infty} \lambda_j^2 \to 0, \\
E \left[ \sum_{j=k+1}^{\infty} \lambda_j^2 Y_j \right]^2 &\leq 3 \sum_{j=k+1}^{\infty} \lambda_j^4 + \left( \sum_{j=k+1}^{\infty} \lambda_j^2 \right)^2 \to 0.
\end{align*}
\]

The convergence in the above display uses Proposition 4.1 part (a). The proof of part (a) is complete.

PROOF OF (B). To begin, use (3) to note that $|B_n(i,j)| \leq C_W \frac{2}{n}$, and
\[
|B_n^2(i,j)| \leq \sum_{k=1}^{n} |B_n(i,k)B_n(k,j)| \leq C_W \frac{2}{n}.
\]

Set $r := a + b + c$, and let $S(\ell, 2r)$ denote the set of all positive integer solutions to the equation $\sum_{i=1}^{\ell} \alpha_i = 2r$. Then we have
\[
\left| E(\sqrt{n}Z)^a (Z^T B_n Z)^b (Z^T B_n^2 Z)^c - E(\sqrt{n}R)^a (R^T B_n R)^b (R^T B_n^2 R)^c \right| \leq n^{-r} C_W^{b+2e} \sum_{\ell=1}^{2r} \sum_{\alpha \in S(\ell,2r)} \left| E \prod_{i=1}^{\ell} Z_i^{\alpha_i} - E \prod_{i=1}^{\ell} R_i^{\alpha_i} \right|.
\]

To bound the RHS of (43), we consider the following cases based on $\alpha$

- There exists $i \in [\ell]$ such that $\alpha_i = 1$

  In this case we have
  \[
  E \prod_{i=1}^{\ell} Z_i^{\alpha_i} = E \prod_{i=1}^{\ell} R_i^{\alpha_i} = 0.
  \]

- $\ell > r$

  In this case we claim that there exists $i \in [\ell]$ such that $\alpha_i = 1$. Thus this is a sub case of the above case.

  Suppose not. Then we have
  \[
  2r = \sum_{i=1}^{\ell} \alpha_i \geq 2\ell,
  \]
  which is a contradiction.
• $\ell = r, \alpha_i \geq 2$ for all $i \in [\ell]$

In this case we must have $\alpha_i = 2$ for all $i$. If not, then we must have

$$2r = \sum_{i=1}^{\ell} \alpha_i > 2\ell,$$

a contradiction. Since $E Z_i^2 = E R_i^2 = 1$, we have

$$E \prod_{i=1}^{\ell} Z_i^{\alpha_i} = E \prod_{i=1}^{\ell} R_i^{\alpha_i} = 1.$$

Combining the cases above, the RHS of (43) is bounded by

$$n^{-r} C W_{b+2c} \sum_{\ell=1}^{r-1} n^\ell \sum_{\alpha \in S(\ell, 2r)} \left| E \prod_{i=1}^{\ell} Z_i^{\alpha_i} - E \prod_{i=1}^{\ell} R_i^{\alpha_i} \right| = O \left( \frac{1}{n} \right),$$

and so the proof of part (b) is complete.

4.1.2. Curie-Weiss model. As it turns out, our proof technique relies on a very precise understanding of what happens under the Curie-Weiss model $P_{\theta, CW}$, defined in (7), and recalled here below:

$$P_{\theta, CW}(X = x) = \exp \left( \frac{n\theta \bar{x}^2}{2} - Z_n(\theta, CW) \right).$$

The following proposition expresses the Curie-Weiss model as a mixture of IID laws. The same decomposition was also utilized previously in the literature (see [13, 24]. We omit the proof.

PROPOSITION 4.2. [24, Lemma 3], [13, Proposition 4.1] Given $X \sim P_{\theta, CW}$, let $\phi_n$ be a real valued random variable defined by

$$\phi_n \sim N \left( \bar{X}, \frac{1}{n\theta} \right).$$

Then the following conclusions hold:

(a) Given $\phi_n$, the random variables $(X_1, \ldots, X_n)$ are IID, with

$$P_{\theta, CW}(X_j = 1|\phi_n) = \frac{\exp (n\theta \phi_n)}{\exp (n\theta \phi_n) + \exp (-n\theta \phi_n)}$$

(b) The marginal density of $\phi_n$ has a density with respect to Lebesgue measure, which is proportional to

$$f_{\theta, n}(\phi_n) = \exp \left( -\frac{1}{2}n\theta \phi_n^2 + n \log \cosh(\theta \phi_n) \right).$$

DEFINITION 4.1. Let $F_{n, \theta}$ denote the distribution of $\phi_n$, as defined in Proposition 4.2.

We begin by proving a lemma about the distribution $F_{n, \theta}$.

LEMMA 4.2. Fix $\theta_0 > 0, h \in \mathbb{R}$, and let $\theta_n = \theta_0 + \frac{h}{\sqrt{n}}$ be as in definition 2.2. Let $\phi_n \sim F_{n, \theta_n}$, where $F_{n, \theta}$ is as in definition 4.1.
(a) If $\theta_0 \in \Theta_3$, then

$$\sqrt{n}\phi_n \to N\left(0, \frac{1}{\theta_0 - \theta_0^2}\right)$$

in distribution, and in moments.

(b) If $\theta_0 \in \Theta_2$, then

$$n^{1/4}\phi_n \to H_h$$

in distribution, and in moments, where $H_h$ is as defined in (18).

(c) If $\theta_0 \in \Theta_1$, then conditional on $\phi_n > 0$ we have

$$\sqrt{n}(\phi_n - m(\theta_n)) \to N\left(0, \frac{1}{\theta_0 - (1 - m^2(\theta_0))\theta_0^2}\right)$$

in distribution, and in moments, where $m(.)$ as in definition 2.1.

**Proof.** (a) **High Temperature Regime** $\Theta_3$.

Use Proposition 4.2 to note that $\phi_n$ has a density proportional to

$$f_{\theta_n}(\phi) = \exp\{-nq_{\theta_n}(\phi)\}, \quad q_\theta(\phi) = \frac{1}{2}\theta\phi^2 - \log\cosh(\theta\phi).$$

Differentiating twice we get

$$\theta_n \geq q''_{\theta_n}(\phi) \geq \theta_n - \theta_n^2,$$

and so if $\theta_0 \in \Theta_3$, there exists finite positive constants $c_1, c_2$ depending only on $\theta_0, h$ (and free of $n$), such that for all $n$ large enough we have

$$c_1 \leq q''_{\theta_n}(\phi) \leq c_2.$$

Consequently, for any $\phi$ we have

$$\frac{c_1}{2}\phi^2 \leq q_{\theta_n}(\phi) \leq \frac{c_2}{2}\phi^2,$$

and so for any $K > 0$ we have

$$\mathbb{P}(\sqrt{n}\phi_n > K) \leq \frac{2}{\int_{-\infty}^{\infty} e^{-nq_{\theta_n}(\phi)} d\phi} \leq \frac{2}{\int_{-\infty}^{\infty} e^{-c_1\phi^2/2} d\phi} \leq \frac{2}{\int_{-\infty}^{\infty} e^{-c_2\phi^2/2} d\phi} \leq \frac{2\sqrt{c_1}}{c_2} \mathbb{P}(N(0, 1/c_1) > K).$$

Thus we have $\sqrt{n}\phi_n = O_p(1)$, and further all moments of $\sqrt{n}\phi_n$ are bounded. Finally, since $q''_{\theta_n}(\phi) \to \theta_0 - \theta_0^2$, it follows from standard calculus that for any $a, b$ fixed, standard calculus gives

$$\sqrt{n}\int_a^{b/\sqrt{n}} e^{-nq_{\theta_n}(\phi)} d\phi \to \int_a^b e^{-\frac{\theta_0 - \theta_0^2}{2}t^2} dt.$$

Combining the above calculations, it follows that

$$\sqrt{n}\phi_n \to N\left(0, \frac{1}{\theta_0 - \theta_0^2}\right),$$

in distribution and in moments.
30

(b) **Critical Regime** $\Theta_2$.

In this case we have

\[ q'_{\theta_n}(0) = q'''_{\theta_n}(0) = 0, \quad \sqrt{n}q''_{\theta_n}(0) = \sqrt{n}(\theta_n - \theta_0^2) \to -h, \quad c'_1 \leq \inf_{|\phi| \leq 2} q'''_{\theta_n}(\phi) \leq \sup_{|\phi| \leq 2} q'''_{\theta_n}(\phi) \leq c'_2, \]

for some constants $c'_1, c'_2$ depending only on $h$. Also, using Proposition 4.2 we have

\[ \mathbb{P}(|\phi_n| > 2) \leq \mathbb{P}(|N(0, 1)| > \sqrt{n}\theta_n), \]

which is exponentially small in $n$. The above two displays together give $n^{1/4}\phi_n = O_p(1)$.

Finally, fixing $a, b$, straightforward calculus gives

\[ \int_a^{b/n^{1/4}} e^{-nq_{\theta_n}(\phi)} d\phi \to \int_a^b e^{\frac{b}{2}\phi^2 - \phi^4/12} d\phi, \]

where we use the fact that $q'''_{\theta_n}(0) \to 2$. Combining, the desired limiting distribution follows. Uniform integrability also follows from the estimates on $q_{\theta_n}(\cdot)$.

(c) **Low Temperature Regime** $\Theta_1$.

In this case we have $\mathbb{P}(\phi_n > 0) = \mathbb{P}(\phi_n < 0) = \frac{1}{2}$ by symmetry. Restricting on the positive half without loss of generality, note that the function $q_{\theta}(\phi)$ has a unique minimizer in $\phi$ on $(0, \infty)$, at the point $m(\theta)$. From Proposition 4.2 we have $q''(m(\theta)) > 0$, and so the function $\Psi : [0, 2] \times [0, 1]$ defined by

\[ \Psi(x, \theta) := \begin{cases} \frac{q_{\theta}(x) - q_{\theta}(m(\theta))}{(x - m(\theta))^2} & \text{if } x \neq m(\theta), \\ \frac{q''_{\theta}(m(\theta))}{2} & \text{if } x = m(\theta), \end{cases} \]

is strictly positive and continuous, and so there exists finite positive constants $c_1, c_2$ depending on $\theta_0, h$, such that for all $\phi > 0$ we have

\[ \frac{c_1}{2} \phi^2 \leq q_{\theta_n}(\phi) - q_{\theta_n}(m(\theta_n)) \leq \frac{c_2}{2} \phi^2. \]

From this, a similar calculation as in part (a) of this lemma applies, on noting that

\[ q''_{\theta_n}(m(\theta_n)) \xrightarrow{p} \theta_0 - \theta_0^2 \left(1 - m^2(\theta_0)\right). \]

4.1.3. **Two analysis results.** In this section, we first verify Proposition 3.2, which was used to prove the main results, and will be used in the sequel as well.

**Proof of 3.2.** We prove the more general result

\[ \lim_{h \to 0} \frac{m(\theta_0 + h) - m(\theta_0)}{h} = m(\theta_0)\sigma^2(\theta_0). \]

The desired conclusion then follows on replacing $h$ by $\frac{h}{\sqrt{n}}$, and letting $n \to \infty$. Recall that $m(\theta)$ satisfies the equation $w(\theta, m) = 0$ in $m$, where

\[ w(\theta, m) := m - \tanh(\theta m). \]
Differentiating with respect to $\theta$ we get
\[
\frac{\partial w(\theta, m)}{\partial \theta} = 1 - \theta \text{sech}^2(\theta m).
\]

By Proposition 3.1, we have $\theta(1 - m^2(\theta)) < 1$, and so the above derivative is always positive. By Implicit Function Theorem, it follows that the function $\theta \mapsto m(\theta)$ is differentiable. On differentiating the equation
\[
m(\theta) = \tanh(\theta m(\theta))
\]
with respect to $\theta$, we get
\[
m'(\theta) = \text{sech}^2(\theta m(\theta))[m(\theta) + \theta m'(\theta)] = (1 - m^2(\theta))[m(\theta) + \theta m'(\theta)].
\]
Solving for $m'(\theta)$ gives
\[
m'(\theta) = \frac{m(\theta)(1 - m^2(\theta))}{1 - \theta(1 - m^2(\theta))} = m(\theta)\sigma^2(\theta),
\]
as desired.

Finally, we prove a uniform convergence result for monotone functions.

**Proposition 4.3.** Suppose $g_n(\cdot)$ is a sequence of functions defined on a compact interval $[a, b]$. Assume the following:

- There exists a function $g(\cdot)$ on $[a, b]$, such that $g_n(\cdot)$ converges to $g(\cdot)$ pointwise.
- The function $g_n(\cdot)$ is non-decreasing, for every $n \geq 1$.
- The function $g(\cdot)$ is continuous.

Then $g_n$ converges to $g$ uniformly on $[a, b]$.

**Proof.** Let $\{t_n\}_{n \geq 1}$ be a real sequence in $[a, b]$ converging to $t_\infty$. We need to show that $g_n(t_n)$ converges to $g_\infty(t_\infty)$. To this effect, fixing $\delta > 0$ arbitrary, for all $n$ large we have $|t_n - t_\infty| < \delta$. Using the monotonicity of $\{g_n\}_{1 \leq n \leq \infty}$ gives
\[
g_n(t_n) - g_\infty(t_\infty) \leq g_n(t - \delta) - g_\infty(t_\delta).
\]
Taking limits as $n \to \infty$ gives
\[
\limsup_{n \to \infty}\{g_n(t_n) - g_\infty(t_\infty)\} \leq g_\infty(t_\infty + \delta) = g_\infty(\delta).
\]
Since $\delta$ is arbitrary, letting $\delta \downarrow 0$ and using the continuity of $g_\infty(\cdot)$ gives
\[
\limsup_{n \to \infty}\{g_n(t_n) - g_\infty(t_\infty)\} \leq 0.
\]
A similar proof gives
\[
\liminf_{n \to \infty}\{g_n(t_n) - g_\infty(t_\infty)\} \geq 0.
\]
The proof is complete by combining the last two displays. \qed
4.2. Two auxiliary lemmas. We now utilize the results of section 4.1 to state and prove two auxiliary lemmas under the Curie-Weiss model, which we will use to verify Lemma 3.1 and Lemma 3.2.

**Lemma 4.3.** For any $h \in \mathbb{R}$ and $\theta_0 > 0$, let $\theta_n = \theta_0 + \frac{h}{\sqrt{n}}$ be as in definition 2.2. Let $X \sim \mathbb{P}_{\theta_n, CW}$, where $Q_n$ is a sequence of matrices which satisfy (2), (3), (4) and (5). Set $B_n = Q_n - \frac{1}{n}11^T$ as before. Also, let $W_{\theta_n} \sim N(0, \sigma^2(\theta_0)), U_h \sim \mathbb{H}_h$ (see (18)) be independent of $(S_0, T_0)$ (see (19) and (20) respectively). Then the following conclusions hold under $\mathbb{P}_{\theta_n, CW}$:

(a) **Low Temperature Regime:** $\theta_0 \in \Theta_1$.

   (i) With $m(.)$ as in definition 2.1, we have
   $$\left( \sqrt{n}(\bar{X} - m(\theta_n)), X^T B_n X, X^T B_n^2 X \right) \overset{d}{\to} (W_{\theta_n}, (1 - m^2(\theta_0))S_0, (1 - m^2(\theta_0))T_0).$$

   (ii) Further we have
   $$\lim_{n \to \infty} \left\{ Z_n(\theta_n, CW) - Z_n(\theta_0, CW) - \frac{\sqrt{n}m^2(\theta_0)}{2} \right\} = \frac{R(\theta_0)h^2}{2},$$
   where $R(\theta_0)$ is as defined in Theorem 2.2.

(b) **Critical Point:** $\theta_0 \in \Theta_2$.

   (i) We have
   $$\left( n^{1/4}\bar{X}, X^T B_n X, X^T B_n^2 X \right) \overset{d}{\to} (U_h, S_0, T_0).$$

   (ii) Further, we have
   $$\lim_{n \to \infty} \left\{ Z_n(1 + h_n) - Z_n(1) \right\} = F(h) - F(0),$$
   where $F(.)$ is as defined in (18).

**Proof.** (a) Before we begin the proof, we point out that Lemma 4.2 part (c) implies that by symmetry,
$$\phi_n \overset{d}{\to} \frac{1}{2}(\delta_{m(\theta_0)} + \delta_{-m(\theta_0)}),$$
and Proposition 4.2 implies that
$$|\phi_n - \bar{X}| \overset{p}{\to} 0.$$  

The last two displays together imply
$$\mathbb{P}_{n, \theta_n}(\phi_n < 0|\bar{X} > 0) \to 0,$$
and so without loss of generality we can interchange between the conditioning events $\bar{X} > 0$ and $\phi_n > 0$. Also, conditional on $\phi_n > 0$ we have $\phi_n \overset{p}{\to} m(\theta_0)$, from Lemma 4.2 part (c).

(i) Using Proposition 4.2, conditioning on $\phi_n$, the random variables $(X_1, \ldots, X_n)$ are IID, with
$$\mathbb{E}(X_1|\phi_n) = \tanh(\theta_n \phi_n) =: \mu_n, \quad Var(X_1|\phi_n) = \text{sech}^2(\theta_n \phi_n).$$
Noting that $B_n1 = 0$, setting $\mu_n := \mu_n1$ one can write
\[
\begin{align*}
&\left[\sqrt{n}(X - m(\theta_n)), X^TB_nX, X^TB_n^2X\right] \\
&= \left[\sqrt{n}(\bar{X} - \mu_n), (X - \mu_n)^TB_n(X - \mu_n), (X - \mu_n)^TB_n^2(X - \mu_n)\right] \\
&\quad + \left[\sqrt{n}(\mu_n - m(\theta_n)), 0, 0\right].
\end{align*}
\]
(48)
By Lemma 4.1, conditioning on $\phi_n$, on the event $\phi_n > 0$ we get
\[
\begin{align*}
&\left[\sqrt{n}(\bar{X} - \mu_n), (X - \mu_n)^TB_n(X - \mu_n), (X - \mu_n)^TB_n^2(X - \mu_n)\right] \\
&\quad \xrightarrow{d} \left[\tau W_0, \tau^2 S_0, \tau^2 T_0\right],
\end{align*}
\]
(49)
where
\[
sech(\theta_n\phi_n) \xrightarrow{p} sech(\theta_0m(\theta_0)) = \sqrt{1 - m^2(\theta_0)} =: \tau.
\]
Proceeding to analyze the second term in the RHS of (48), a one term Taylor’s expansion then gives that
\[
\sqrt{n}(\mu_n - m(\theta_n)) = \sqrt{n}\left(\tanh(\theta_n\phi_n) - \tanh(\theta_n m(\theta_n))\right)
\]
\[
= \sqrt{n}(\phi_n - m(\theta_n))\theta_n sech^2(\theta_n\xi_n),
\]
where $\xi_n$ lies between $\phi_n$ and $m(\theta_n)$, and hence converges to $m(\theta_0)$ in probability. Using the above display, conditional on $\phi_n > 0$ we have
\[
\begin{align*}
&\sqrt{n}(\mu_n - m(\theta_n)) \xrightarrow{d} \theta_0 sech^2(\theta_0 m(\theta_0))N\left(0, \frac{1}{\theta_0 - (1 - m^2(\theta_0))\theta_0^2}\right) \\
&\quad = N\left(0, \frac{\theta_0(1 - m^2(\theta_0))^2}{1 - (1 - m^2(\theta_0))\theta_0}\right).
\end{align*}
\]
Combining the last display along with (48) and (49), the conclusion of part (a) follows, on noting that
\[
1 - m^2(\theta_0) + \frac{\theta_0(1 - m^2(\theta_0))^2}{1 - (1 - m^2(\theta_0))\theta_0} = \frac{1 - m^2(\theta_0)}{1 - \theta_0(1 - m^2(\theta_0))} = \sigma^2(\theta_0).
\]
(ii) To begin, use Proposition 4.2 to get
\[
\phi_n = \bar{X} + \frac{W_0}{\sqrt{n}\theta_n},
\]
where $W_0 \sim N(0, 1)$ is independent of $X$. This gives
\[
\sqrt{n}(\bar{X} - m(\theta_n)) = \sqrt{n}(\phi_n - m(\theta_n)) - \frac{W_0}{\sqrt{n}\theta_n}.
\]
(51)
which along with part (c) of Lemma 4.2 shows that conditional on $\bar{X} > 0$ we have
\[
\sqrt{n}(\bar{X} - m(\theta_n)) \xrightarrow{d} N\left(0, \frac{1 - m^2(\theta_0)}{1 - \theta_0(1 - m^2(\theta_0))}\right) = N(0, \sigma^2(\theta_0)).
\]
Thus in turn implies that unconditionally, we have
\[
\sqrt{n}(\bar{X}^2 - m^2(\theta_n)) \xrightarrow{d} N(0, 4\sigma^2(\theta_0)m^2(\theta_0)).
\]
Using Proposition 3.2, we then have
\[ \frac{\sqrt{n}}{2} \left( \overline{X}^2 - m^2(\theta_0) \right) \overset{d}{\to} N \left( m^2(\theta_0)\sigma^2(\theta_0)h, m^2(\theta_0)\sigma^2(\theta_0) \right), \]
which gives
\[ \frac{Z_n'(\theta_0 + h/\sqrt{n}, CW) - \sqrt{n}m^2(\theta_0)}{\sqrt{n}} = \frac{\sqrt{n}m^2(\theta_0)}{2} \left( \mathbb{E}_{\theta_n,CW} \overline{X}^2 - m^2(\theta_0) \right) \]
\[ = \mathbb{E}_{\theta_0,CW} \left( m^2(\theta_0)\sigma^2(\theta_0)h, m^2(\theta_0)\sigma^2(\theta_0) \right)^2 = m^2(\theta_0)\sigma^2(\theta_0)h = R(\theta_0)h. \]

In the line above, we have used the fact that \( \sqrt{n}(\overline{X}^2 - m^2(\theta_0)) \) is uniformly integrable. But this follows on using (51), along with the fact that \( \sqrt{n}(\phi_n - m(\theta_n)) \) is uniformly integrable (from Lemma 4.2 part (c)).

The convergence in the above display holds for all \( h \) fixed. Integrating both sides over the interval \([0, h]\) we get
\[ Z_n(\theta_0 + h/\sqrt{n}, CW) - Z_n(\theta_0, CW) - \frac{\sqrt{n}hm^2(\theta_0)}{2} \to R(\theta_0)h^2 / 2, \]
as desired. In the last convergence above, we use the fact that the function \( h \mapsto Z_n'(\theta_0 + h/\sqrt{n}, CW) \) is monotone, and hence converges uniformly over compact sets, by Proposition 4.3. This completes the proof of part (a).

(b) (i) Again using calculations similar to (48), we get
\[ \left[ n^{1/4} \overline{X}, X^T B_n X, X^T B_n^2 X \right] \]
\[ = n^{1/4} \left( \overline{X} - \mu_n, (X - \mu_n)^T B_n (X - \mu_n), (X - \mu_n)^T B_n^2 (X - \mu_n) \right) + [n^{1/4} \mu_n, 0, 0]. \]
Conditioning on \( \phi_n \), using Proposition 4.2 and Lemma 4.1 we have
\[ \left[ n^{1/4} (X - \mu_n), (X - \mu_n)^T B_n (X - \mu_n), (X - \mu_n)^T B_n^2 (X - \mu_n) \right] \overset{d}{\to} [0, S_0, T_0], \]
where we use the fact
\[ Var(X_1 | \phi) = \text{sech}(\theta_n \phi_n) \overset{P}{\to} 1. \]
Also, an application of delta theorem along with Lemma 4.2 part (b) gives
\[ n^{1/4} \mu_n = n^{1/4} \tanh(\phi_n) \overset{d}{\to} U_{1,h}. \]
Combining the above, it follows that
\[ \left[ n^{1/4} \overline{X}, X^T B_n X, X^T B_n^2 X \right] \overset{d}{\to} [U_h, S_0, T_0] \]
as desired.
(ii) Using (50) we get
\[ n^{1/4} \overline{X} = n^{1/4} \phi_n - n^{1/4} \frac{W_0}{\sqrt{n\theta_0}} \overset{d}{\to} U_h, \]
where we have used Lemma 4.2 part (b). This in turn gives,
\[ \sqrt{n}X^2 \xrightarrow{d} U^2, \]
and so
\[ Z_n'(\theta_n) = \frac{\sqrt{n}}{2} \mathbb{E}_{P_{\theta_n,CW}} X^2 \xrightarrow{1} \mathbb{E} U^2 = F'(h). \]

Again in the last step we use the fact that \( \sqrt{n}X^2 \) is uniformly integrable, which follows from (50) and Lemma 4.2 part (b). The desired conclusion again follows on integrating the above display over \([0,h]\), on noting that the above convergence is uniform on compact sets, by Proposition 4.3.

**Lemma 4.4.** Suppose the matrix \( Q_n \) satisfies (2), (3), (4) and (5) for some \( C_W, \kappa \in (0, \infty) \) and \( f \in W \). Let \( \theta_0 > 0, h \in \mathbb{R}, \) and \( \theta_n = \theta_0 + \frac{h}{\sqrt{n}} \) be as in definition 2.2. Then the following conclusions hold for \( \theta \in \Theta_1 \cup \Theta_2 \).

(a) \[ \lim_{n \to \infty} Z_n(\theta_n, Q_n) - Z_n(\theta_n, CW) = C(\theta_0), \]
where
\[ C(\theta_0) := -\frac{1}{2} \theta (1 - m^2) + \frac{\kappa}{4} \theta_0^2 (1 - m^2)^2 + \frac{1}{2} \sum_{i=2}^{\infty} \left[ \log \left( 1 - \theta_0 (1 - m^2) \lambda_i \right) + \lambda_i \theta_0 (1 - m^2) \right]. \]

(b) The probability measures \( P_{\theta_n,Q_n} \) and \( P_{\theta_n,CW} \) are mutually contiguous.

**Proof.** (a) To begin, note that
\[ \exp \left( Z_n(\theta_n, Q_n) - Z_n(\theta_n, CW) \right) = \frac{\sum_{\mathbf{x} \in \{-1,1\}^n} e^{\frac{\theta_n}{2} \mathbf{x}^T Q_n \mathbf{x}}}{\sum_{\mathbf{x} \in \{-1,1\}^n} e^{\frac{\theta_n}{2} n \mathbf{x}^2}} \]
\[ = \mathbb{E}_{\theta_n,CW} e^{\frac{\theta_n}{2} \mathbf{x}^T B_n \mathbf{x}}. \]

It follows from Lemma 4.3 that if \( \mathbf{X} \sim P_{\theta_n,CW}, \) then for all \( \theta_0 \in \Theta_0 \) we have
\[ \mathbf{X}^T B_n \mathbf{X} \xrightarrow{d} (1 - m^2(\theta_0)) S_0 \Rightarrow e^{\frac{\theta_n}{2} \mathbf{x}^T B_n \mathbf{x} \xrightarrow{d} e^{\frac{\theta_0 (1 - m^2(\theta_0))}{2} S_0}. \]

Assume now that there exists \( \delta > 0 \) such that
\[ E_{\theta_n,CW} e^{(1+\delta)\frac{\theta_n}{2} \mathbf{x}^T B_n \mathbf{x} < \infty.} \]

Uniform integrability then gives
\[ E_{\theta_n,CW} e^{\frac{\theta_n}{2} \mathbf{x}^T B_n \mathbf{x}} \to E e^{\frac{\theta_0 (1 - m^2(\theta_0))}{2} S_0}, \]
which equals \( C(\theta_0) \) using the formula for \( S_0 \) (see (19)).

It thus remains to verify (52). To this effect, note that
\[ E_{\theta_n,CW} e^{\frac{\theta_n}{2} \mathbf{x}^T B_n \mathbf{x}} = E_{\theta_n,CW} \left( e^{\frac{\theta_n}{2} (\mathbf{x} - \mu_n)^T B_n (\mathbf{x} - \mu_n) | \phi_n} \right), \]
where $\mathbf{m}_n = \mathbf{m}_n \mathbf{1}$ with $\mathbf{m}_n = \tanh(\theta_n \phi_n)$, as in the proof of Lemma 4.3. Invoking Lemma 4.2, we have that given $\phi_n$ the random variables $(X_1, \ldots, X_n)$ are IID with mean $\mu_n$. Also, setting

$$s_\mu := \frac{2\mu}{\log(1+\mu)-\log(1-\mu)} \quad \text{if } \mu \neq 0,$$

$$s_\mu := 1 \quad \text{if } \mu = 0$$

we have that $s_\mu$ is a strictly positive continuous even function, with $s_{\mu_n} \xrightarrow{p} s_{m(\theta_0)} = \frac{1}{B_{\theta_0}}$ for all $\theta_0 \in \Theta_1 \cup \Theta_2$, where the last equality uses the fact that $m(\theta_0) = \tanh(\theta_0 m(\theta_0))$. Since

$$\limsup_{n \to \infty} \lambda_1(B_n) = \lambda_2 < 1$$

by (5), there exists $\delta > 0$ such that on the set $|\mu_n - m(\theta_0)| > \delta$ we have

$$\limsup_{n \to \infty} \theta_n \lambda_1(B_n)s_{\mu_n} < 1.$$

Thus using [13, Proposition 4.1] with

$$N = n, \quad D_N(i, j) = \theta_n B_n(i, j), \quad c_i = 0,$$

we get the existence of a constant $C$ free of $n$ such that on the set $|\mu_n - m(\theta_0)| > \delta$ we have

$$\log \mathbb{E}_{\theta_n, CW}(e^{\frac{\theta_n}{2} (X - \mu_n)^T B_n (X - \mu_n)} | \phi_n) \leq C.$$

To complete the proof of (52), it suffices to show that

$$\limsup_{n \to \infty} \mathbb{E}_{\theta_n, CW}(e^{\frac{\theta_n}{2} X^T B_n X} 1\{|\mu_n - m(\theta_0)| > \delta\}) < \infty,$$

for some $\delta > 0$. But this follows from [13, Lemma 4.2] on using

$$N = n, \quad V_N = \theta_n X^T B_n X.$$

We note that the proof of [13, Lemma 4.2] goes through verbatim if $\theta = \theta_n$ depends on $n$, even though the lemma is stated for a fixed $\theta_0$.

(b) The likelihood ratio between $\mathbb{P}_{\theta_n, Q_n}$ and $\mathbb{P}_{\theta_n}$ is given by

$$\frac{d\mathbb{P}_{\theta_n, Q_n}}{d\mathbb{P}_{\theta_n, CW}}(X) = \exp \left( \frac{\theta_n}{2} X^T B_n X - Z_n(\theta_n, Q_n) + Z_n(\theta_n, CW) \right).$$

Using Lemma 4.3, for all $\theta_0 \in \Theta$ we have

$$X^T B_n X \xrightarrow{d} (1 - m^2(\theta_0)) S_0.$$

Also, using part (a) we have

$$Z_n(\theta_n, Q_n) - Z_n(\theta_n, CW) \to \log \mathbb{E}_e^{\frac{\theta_n (1 - m^2(\theta_0))}{2} S_0}.$$

Combining, if $X \sim \mathbb{P}_{\theta_n, CW}$, then we have

$$\frac{d\mathbb{P}_{\theta_n, Q_n}}{d\mathbb{P}_{\theta_n, CW}}(X) \xrightarrow{d} \frac{e^{\frac{\theta_n (1 - m^2(\theta_0))}{2} S_0}}{e^{\frac{\theta_n (1 - m^2(\theta_0))}{2} S_0}}.$$

Since the limiting random variable in the above display is strictly positive and has mean 1, mutual contiguity follows by Le-Cam’s first lemma.
4.3. Proof of Lemma 3.1 and Lemma 3.2. Using Lemma 4.3 and Lemma 4.4, we now prove Lemma 3.1 and Lemma 3.2, which were stated in the main draft, and were used to prove our main results.

PROOF OF LEMMA 3.1. To begin, note that in all regimes of \( \theta \), the following hold:

- The log likelihood ratio
  \[
  \log \frac{dP_{\theta, Q_n}}{dP_{\theta, CW}}(X) = \frac{\theta_n}{2} X^T B_n X - Z_n(\theta_n, Q_n) + Z_n(\theta_n, CW)
  \]
  is a function of \( X^T B_n X \).
- The asymptotic non degenerate limiting distribution of \( \tilde{X} \) is jointly independent of \( X^T B_n X \) and \( X^T B_n^2 X \) (this follows from Lemma 4.3).
- The two measures \( P_{\theta, Q_n} \) and \( P_{\theta, CW} \) are mutually contiguous (this follows from Lemma 4.4 part (b)).

It thus follows from Le-Cam’s third lemma that the asymptotic non degenerate distribution of \( \tilde{X} \) under \( P_{\theta, Q_n} \) and \( P_{\theta, CW} \) are the same, and is asymptotically independent of the joint distribution of \( (X^T B_n X, X^T B_n^2 X) \) under both models.

To complete the proof of Lemma 3.1, it then suffices to show that for \( \theta_0 \in \Theta_1 \cup \Theta_2 \), under \( P_{\theta_0, Q_n} \) we have

\[
(X^T B_n X, X^T B_n^2 X) \overset{d}{\to} (S_{\theta_0}, T_{\theta_0}).
\]

To show this, first note that under \( P_{\theta_0, CW} \) we have

\[
\begin{align*}
&\left[ X^T B_n X, X^T B_n^2 X, \log \frac{dP_{\theta_0, Q_n}}{dP_{\theta_0, CW}}(X) \right] \\
&= \left[ X^T B_n X, X^T B_n^2 X, \frac{\theta_0}{2} X^T B_n X - Z_n(\theta_0, Q_n) + Z_n(\theta_0, CW) \right] \\
&\overset{d}{\to} \left[ (1 - m^2(\theta_0)) S_0, (1 - m^2(\theta_0)) T_0, \frac{\theta_0(1 - m^2(\theta_0))}{2} S_0 - C(\theta_0) \right],
\end{align*}
\]

and we have used (53) in the first step, and Lemma 4.3 and Lemma 4.4 part (a) in the second step (and \( C(\theta_0) \) is defined in Lemma 4.4 part (a)). Using mutual contiguity, it follows that under \( P_{\theta_0, Q_n} \), we have

\[
(X^T B_n X, X^T B_n^2 X) \overset{d}{\to} (S', T'),
\]

where \( (S', T') \) is a bi-variate random vector with characteristic function

\[
\mathbb{E} e^{i(sS + tT')}
\]

\[
= \mathbb{E} \exp \left\{ (1 - m^2(\theta_0))isS_0 + tT_0 + \frac{\theta_0(1 - m^2(\theta_0))}{2} S_0 - C(\theta_0) \right\}
\]

\[
= e^{-C(\theta_0)} \mathbb{E} \exp \left\{ (1 - m^2(\theta_0))(is + \frac{\theta_0}{2})S_0 + itT_0 \right\}
\]

\[
= e^{-C(\theta_0)} \mathbb{E} \exp \left\{ (1 - m^2(\theta_0))(is + \frac{\theta_0}{2}) \left( \sum_{j=2}^{\infty} \lambda_j (Y_j - 1) - 1 + W^* \right) + it(1 - m^2(\theta_0)) \left( \sum_{j=2}^{\infty} \lambda_j^2 Y_j + \kappa \right) \right\}
\]
\[ e^{-C(\theta_0)} \exp \left\{ -(1 - m^2(\theta_0)) \left( is + \frac{\theta_0}{2} \right) + i\kappa (1 - m^2(\theta_0)) \right\} \mathbb{E} \exp \left\{ (1 - m^2(\theta_0)) \left( is + \frac{\theta_0}{2} \right) W^* \right\} \]

\[ \prod_{j=2}^{\infty} \exp \left\{ -(1 - m^2(\theta_0)) \left( is + \frac{\theta_0}{2} \right) \lambda_j \right\} \mathbb{E} \exp \left\{ (1 - m^2(\theta_0)) \left( is + \frac{\theta_0}{2} \right) \lambda_j Y_j + (1 - m^2(\theta_0)) t \lambda_j^2 Y_j \right\} \]

\[ e^{-C(\theta_0)} \exp \left\{ (1 - m^2(\theta_0)) \left( it\kappa - is - \frac{\theta_0}{2} \right) + (1 - m^2(\theta_0)) \left( is + \frac{\theta_0}{2} \right)^2 \kappa \right\} \]

\[ \prod_{j=2}^{\infty} \frac{\exp \left\{ -(1 - m^2(\theta_0)) \left( is + \frac{\theta_0}{2} \right) \lambda_j \right\}}{\sqrt{1 - (1 - m^2(\theta_0)) (\theta_0 \lambda_j + 2i\lambda_j s + 2i\lambda_j^2 t)}} \]

\[ = e^{(1 - m^2(\theta_0)) (it\kappa - is) + (1 - m^2(\theta_0))^2 (is + \kappa)^2} \prod_{j=2}^{\infty} \frac{e^{- (1 - m^2(\theta_0)) \left( is + \frac{\theta_0}{2} \right) \lambda_j}}{1 - (1 - m^2(\theta_0)) \left( 1 - 2i(1 - m^2(\theta_0)) \frac{\lambda_j s + \lambda_j^2 t}{1 - \theta_0 (1 - m^2(\theta_0)) \lambda_j} \right)} \]

where in the last step we use the formula for \( C(\theta_0) \) from Lemma 4.4 part (a). The last display above can be checked to be the joint characteristic function of

\[ (1 - m^2(\theta_0)) \sum_{j=2}^{\infty} \lambda_j \left( \frac{Y_j}{1 - \theta_0 (1 - m^2(\theta_0)) \lambda_j} - 1 \right) - 1 + (1 - m^2(\theta_0)) \theta_0 \kappa + W^*, \]

\[ \sum_{j=2}^{\infty} \frac{\lambda_j^2 Y_j}{1 - \theta_0 (1 - m^2(\theta_0)) \lambda_j} + \kappa \],

and so the proof of (54) is complete.