# Inference in ERGMs and Ising Models. 

Yuanzhe Xu

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy under the Executive Committee of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY
© 2023

Yuanzhe Xu

All Rights Reserved

# Abstract <br> Inference in ERGMs and Ising Models. 

Yuanzhe Xu

Discrete exponential families have drawn a lot of attention in probability, statistics, and machine learning, both classically and in the recent literature. This thesis studies in depth two discrete exponential families of concrete interest, (i) Exponential Random Graph Models (ERGMs) and (ii) Ising Models. In the ERGM setting, this thesis consider a "degree corrected" version of standard ERGMs, and in the Ising model setting, this thesis focus on Ising models on dense regular graphs, both from the point of view of statistical inference.

The first part of the thesis studies the problem of testing for sparse signals present on the vertices of ERGMs. It proposes computably efficient tests for a wide class of ERGMs. Focusing on the two star ERGM, it shows that the tests studied are "asymptotically efficient" in all parameter regimes except one, which is referred to as "critical point". In the critical regime, it is shown that improved detection is possible. This shows that compared to the standard belief, in this setting dependence is actually beneficial to the inference problem. The main proof idea for analyzing the two star ERGM is a correlations estimate between degrees under local alternatives, which is possibly of independent interest.

In the second part of the thesis, we derive the limit of experiments for a class of one parameter Ising models on dense regular graphs. In particular, we show that the limiting experiment is Gaussian in the "low temperature" regime, non Gaussian in the "critical" regime, and an infinite collection of Gaussians in the "high temperature" regime. We also derive the limiting distributions of commonlt studied estimators, and study limiting power for tests of hypothesis against contiguous alternatives (whose scaling changes across the regimes). 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).

## Table of Contents

Acknowledgments ..... iv
Chapter 1: Introduction ..... 1
Chapter 2: Signal Detection in Degree Corrected ERGMs ..... 5
2.1 Introduction and Problem Setup ..... 5
2.1.1 Degree Corrected ERGM ..... 7
2.1.2 Hypothesis Testing Problem for $\boldsymbol{\beta}$ ..... 9
2.2 Main Results ..... 11
2.2.1 General Degree Corrected ERGMs ..... 11
2.2.2 Degree Corrected Two-star ERGM ..... 13
2.2.3 Main Contributions and Future Scope ..... 16
2.2.4 Outline ..... 19
2.3 Proofs of Main Theorems ..... 19
2.3.1 Proof of Theorem 2.2.1 ..... 21
2.3.2 Proof of Theorem 2.2.2 ..... 22
2.3.3 Proof of part (b) and (d) of Theorem 2.2.3 and Theorem 2.2.4 ..... 23
2.3.4 Proof of Theorem 2.2.5 part (b) ..... 23
2.3.5 Proof of part (a) and (c) of Theorems 2.2.3 ..... 28
2.3.6 Proof of Theorem 2.2.5 Part (a) ..... 31
2.3.7 Proof of part (a) and (c) of Theorem 2.2.4 ..... 32
2.3.8 Proof of Lemma 2.3.5 ..... 35
2.4 Proofs of Auxiliary Variable Lemmas ..... 42
2.4.1 Proof of Proposition 2.3.1 ..... 42
2.4.2 Proof of Lemma 2.3.2 ..... 43
2.4.3 Proof of Lemma 2.3.6 ..... 49
2.4.4 Proof of Lemma 2.3.7 ..... 53
Chapter 3: Ising models on dense regular graphs ..... 56
3.1 Introduction ..... 56
3.2 Main results ..... 57
3.2.1 Formal set up ..... 57
3.2.2 Graphon Convergence ..... 58
3.2.3 Limits of Experiments ..... 61
3.2.4 Estimation of $\theta$ ..... 62
3.2.5 Hypothesis Testing for $\theta$ ..... 63
3.2.6 Statement of main results ..... 65
3.2.7 Examples ..... 71
3.2.8 Main Contributions \& Future Scopes ..... 73
3.2.9 Outline ..... 75
3.3 Proofs of Main Theorems ..... 75
3.3.1 Proof of Theorem 3.2.2 ..... 78
3.3.2 Proof of Theorem 3.2.3 ..... 84
3.3.3 Proof of Theorem 3.2.4 ..... 89
3.3.4 Proof of Proposition 3.3.1 ..... 95
3.4 Acknowledgements ..... 97
3.5 Appendix A:Proofs of main lemmas ..... 97
3.5.1 Proof of Lemma 3.3.1 ..... 99
3.5.2 Proof of Lemma 3.3.2 ..... 102
3.6 Appendix B: Proofs of lemmas on Curie-Weiss models ..... 102
3.6.1 Proof of Lemma 3.5.1 ..... 104
3.6.2 Proof of Lemma 3.5.2 ..... 110
3.7 Appendix C: Proofs of supporting lemmas ..... 112
3.7.1 Proof of Proposition 3.3.2 ..... 112
3.7.2 Proof of Proposition 3.6.2 ..... 113
3.7.3 Proof of Lemma 3.6.1 ..... 114
3.7.4 Proof of Lemma 3.6.2 ..... 116
References ..... 122

## Acknowledgements

Working with Prof. $\sim$ Sumit Mukherjee is always a wonderful experience. Countless times, Sumit offers innovative ideas in formulating questions and insightful feedback on technical details. His enthusiasm and perseverance in research persistently encouraged me to sharpen my thinking and level up my work to a higher quality. Furthermore, his passion in life persistently rubs off on me, inspiring me with hope and confidence. I am exceptionally grateful for his generous and continuous support throughout my Ph.D. life!

I am deeply indebted to Prof. $\sim$ Yang Feng, who supported me in setting up my Ph.D. study, patiently guided my early stages of research, and has been an invaluable resource for me these years. Discussing with Yang and attending his group meeting diversified my exposure to various research fields, which benefited me building up a statistical mindset.

Great appreciation for Prof.~Zhiliang Ying and Prof. $\sim$ Ming Yuan, for being so warmhearted and helpful these years. Exploiting or maybe overusing their time with or without appointments, my interactions with Zhiliang and Ming always award me great support and encouragement. Without these help, this dissertation may not come into place today.

I am also sincerely thankful to Prof. $\sim$ Rajarshi Mukherjee, Prof. $\sim$ Philip Protter and Prof. $\sim$ Arian Maleki for the helpful discussions on my projects, which benefit this thesis and my understanding in related questions. Special thanks to Prof. $\sim$ Richard Davis and Prof. $\sim$ Tian Zhen, for being fantastic leaders of our department! Besides, I want to give huge admiration to Dood Kalicharan, Anthony Cruz, and Thomson Batidzirai, for taking care of our lives and allowing us to concentrate on research in the department.

Thank you to my cohort: Shun Xu, Charles margossian, Nick Galbraith, Reed Palmer, Elliott Gordon-Rodriguez, Chi Wing Chu and Naburun Deb, for sharing the pains in courseworks, sport
streams at "bar 1020" and happy hours at Craftsman. Thank you to all Ph.D fellows in the department, and all my friends in US or oversea. It is hard to name everyone here as there is a long list, who share my feelings, encourage me to step out of my comfort zones, help me out during tough times, and hang out with me to rest my mind outside work. Thank you for composing my enjoyable and memorable Ph.D. life.

With immense gratitude and pride, I want to acknowledge my parents, for constantly being present, for supporting all my decisions, for their never-ending love and for all that they have done for me!

## Chapter 1: Introduction

Exponential families of probability measures have a long history in Statistics and Machine Learning, and have also been the focus of attention in several other disciplines, including Social Science, Biology, Neuroscience, and Mathematical Physics, to name a few. In this thesis, we focus on two well-known classes of exponential families, which are described below:

## 1. Exponential Random Graph Models

Network models have received significant attention in Statistics and Machine Learning, motivated by problems arising in several disciplines. Possibly the most simple (but also the most well studied) network model is the Erdős-Rényi model (see [24]), where the number of edges is the sufficient statistic. For this model, the edges of the graph are IID Bernoulli random variables. Another commonly studied model of network analysis is the stochastic block model (SBM, see [29]), in which the edge probabilities depend on the community structure. A major drawback of these models is the assumption of independence between the edges, which may be not realistic.

Exponential Random Graph Models (ERGMs in short) alleviate this, by assuming a dependence structure between the edges of the graph. ERGMs originated in the Social Science Literature, and have since then received a lot of attention in Probability and Statistics. Essentially, the joint distribution of the random network is specified as a finite parameter exponential family, over the space of simple labeled graphs. Typical choices of sufficient statistics for ERGMs are subgraph counts. Possibly the most simple ERGM (outside of the Erdős-Rényi model) is the two-star model, first studied in [43]. This model has two sufficient statistics, the number of edges, and the number of two stars. A two-star structure means a path of length 2 , which consist of 3 vertices and 2 edges. Below we define the two-star
model formally, by introducing some notation that we will carry throughout the rest of the thesis:

For a simple labeled graph $G$ with vertex set $[n]:=\{1,2, \ldots, n\}$, by a slight (but standard) abuse of notation, let $G$ also denote the adjacency matrix $\left(G_{i j}\right)_{i, j=1}^{n}$, defined by

$$
G_{i j}= \begin{cases}1 & \text { If an edge is present between vertices } i \text { and } j \text { in the graph } G \\ 0 & \text { If no edge is present between vertices } i \text { and } j \text { in the graph } G\end{cases}
$$

Given parameters $\omega_{1}>0$ and $\omega_{2} \in \mathbb{R}$, the two-star model has the following p.m.f.:

$$
\begin{equation*}
\mathbb{P}_{n}(G)=\frac{1}{Z_{n}\left(\omega_{1}, \omega_{2}\right)} \exp \left(\left(\omega_{2}+\frac{\omega_{1}}{n-1}\right) E(G)+\frac{\omega_{1}}{n-1} T(G)\right) \tag{1.1}
\end{equation*}
$$

where

$$
E(G)=\sum_{i<j} G_{i j}, \quad T(G)=\sum_{i=1}^{n} \sum_{j \neq k} G_{i j} G_{i k} .
$$

Going beyond the two-star model, one can substitute $T(G)$ by more general sub-graph counts, such as cycles, triangles or cliques. In Chapter 2, we will study a class of ERGMs which generalize (1.1) by allowing for general sub-graph counts, but also allowing for degree heterogeneity, something which is not present in (1.1).

The reader may have noticed that classically in a graphical model, the nodes of the graph are random, whereas the edges of the graph are random in a network model. When referring to two-star ERGM, this difference is only superficial, and one can express a two-star model as a graphical model on the line graph of the complete graph (see [41, Section 1.2]).

## 2. Ising Models.

Ising Models originated in statistical physics for studying Ferro-magnetism and has since then found its way across several disciplines, including Statistics and Machine Learning. In


Figure 1.1: In a network model, the edges be- Figure 1.2: In a graphical model, vertices are bitween vertices are binary random variables, whereas nary random variables, whereas the edges denote the vertices are fixed. conditional dependence.
an Ising model, the dependence mechanism is controlled by a graph/matrix. To describe the model, suppose that $Q_{n}=\left(Q_{i j}\right)_{i, j=1}^{n}$ is a symmetric $n \times n$ matrix with 0 on the diagonal, and let $\mu=\left(\mu_{1}, \mu_{2}, \ldots, \mu_{n}\right)^{T}$ be a vector of real parameters, commonly referred as 'external magnetic field' in Statistical Physics. Finally, let $\theta$ be a real valued scalar parameter, commonly referred to as "inverse temperature" (and hence assumed to be non-negative). Armed with these definitions, we introduce the Ising model as the following p.m.f. on $\{-1,1\}^{n}$ :

$$
\begin{equation*}
\mathbb{P}_{\theta, \mu, Q_{n}}(X)=\frac{1}{Z\left(\theta, \mu, Q_{n}\right)} \exp \left(\frac{\theta}{2} \mathbf{X}^{T} Q_{n} \mathbf{X}+\mu^{T} \mathbf{X}\right) \tag{1.2}
\end{equation*}
$$

The quantity $Z_{n}\left(\theta, \mu, Q_{n}\right)$ is the normalizing constant, which makes $\mathbb{P}_{\theta, \mu, Q_{n}}$ into a probability distribution. Very often the matrix $Q_{n}$ is assumed to be a (scaled) adjacency matrix of an undirected graph. We refer to Chapter 3 for more on this connection.

The question of statistical inference for both the models described above is of interest and is the main focus of this thesis. It is also a topic that is of recent interest. In particular, [15] developed a large deviation theory to study a wide class of ERGMs. Focusing on the two-star model, [41] investigates the limiting distributions for the number of edges, across the different phases of parameter configurations. Focusing on Ising Models, [7,14] study consistency of maximum pseudo-
likelihood estimators.In another line of work, [18] and [40] investigate the effect of dependence in signal detection for external magnetic field $\mu$ under Ising Models in a minimax framework. Under the same framework, [39] study sharp threshold in testing sparse signals under $\beta$-model.

However, several interesting questions remain unresolved. One natural question is the signal detection problem in general ERGMs. This is the focus of chapter 2, where we introduce a class of general degree corrected Exponential Random Graphical Models, and set up a similar minimax framework for the hypothesis testing problem as in [39, 40]. We adopt centered versions of two classic test statistics for testing and demonstrate their optimality. To match the sharp testing threshold, we include the lower bounds under degree corrected two-star ERGMs for the entire parameter space.

Another topic of interest is inference on "inverse temperature" $\theta$ in Ising Models on a dense regular graphs. Addressing this, we show that the Ising Models converges in the sense of limits of experiment, (c.f. [48, Chapter 9] for discussion and more examples):

Given a parameter space $\Theta$, a family of distribution $\mathbb{P}_{\theta}(\cdot)$ is called to be the limits of experiments of $\mathbb{P}_{n, \theta}(\cdot)$ if:

$$
\left(\frac{d \mathbb{P}_{n, \theta}}{d \mathbb{P}_{n, \theta_{0}}}(\mathbf{X})\right)_{\theta \in I} \xrightarrow{d, \mathbb{P}_{n, \theta_{0}}}\left(\frac{d \mathbb{P}_{\theta}}{d \mathbb{P}_{\theta_{0}}}(Y)\right)_{\theta \in I},
$$

for every finite subset $I \subset \Theta$. As a consequence, we derive the best possible power function for tests involving $\theta$. We study three natural tests and investigate whether they reach the optimal threshold, in each of the parameter domains (high temperature, criticality, low temperature). We show that the optimality of tests depends crucially on the parameter domain and the graph spectrum. All of this is the content of chapter 3. Our focus is restricted to a certain subclass of Ising Models, which we will refer to as dense Regular Ising Models.

Chapter 2 and Chapter 3 compose the main body of this thesis. As indicated above, chapter 2 covers signal detection in degree corrected ERGMs, and chapter 3 is dedicated to inference in dense Regular Ising Models. Some of the technical lemmas may be of independent interest, and suggest a scope for more theoretical studies in this field.

# Chapter 2: Signal Detection in Degree Corrected ERGMs 

### 2.1 Introduction and Problem Setup

Studying network models has a long and rich history in Statistics, with applications across various disciplines such as Social Science, Biology, Neuroscience, Climatology, and Ecology, to name a few. One of the most well known network models is the Exponential Random Graph Model (often abbreviated as ERGM). ERGMs originated in the Social Science Literature (c.f. [1, $25,30,45,49,50$ ] and the references there-in), and have since then received considerable attention in Statistics and Probability (c.f. [15, 16, 27, 39, 40, 46, 47] and references there-in). ERGMs represent exponential families of distributions the space of simple labeled graphs with a finite dimensional sufficient statistics, which are usually taken to be subgraph counts. The simplest class of examples under this framework consists of the one parameter ERGMs, which admits a one dimensional sufficient statistic. Below we start by introducing such a one parameter ERGM:

Letting $\mathcal{G}_{n}$ denote the set of all simple labeled graphs $G$ with vertex set $[n]:=\{1,2, \ldots, n\}$, we consider the following probability mass function on $\mathcal{G}_{n}$ :

$$
\begin{equation*}
\mathbb{P}_{n, \theta}(G):=\frac{1}{Z_{n}(\theta, H)} \exp \left\{\theta \frac{N(H, G)}{n^{\zeta-2}}\right\} \tag{2.1}
\end{equation*}
$$

Here
(i) $H$ is a graph of fixed size (such as an edge, triangle, cycle, star, etc.),
(ii) $N(H, G)$ is the number of copies of the graph $H$ in the graph $G$,
(iii) $\zeta$ is the number of vertices in the graph $H$,
(iv) $\theta$ is a real valued parameter,
(v) $Z_{n}(\theta, H)$ is the normalizing constant.

In particular if the graph $H$ is an edge, the model in (2.1) is an Erdős-Rényi model, where the edges of the graph $G$ are i.i.d from suitable a Bernoulli distribution. For any other choice of $H$, the model in (2.1) is not an Erdős-Rényi model since one allows nontrivial dependence between the edges. An ERGM can thus be thought of as a natural generalization of the Erdős-Rényi model, which allows for growing degrees of dependence between edges by through the indexing subraph $H$. It is natural to allow for this dependence while modeling networks, to incorporate features like "friends of friends are more likely to be friends". However, one drawback of ERGMs (or at least the model introduced in (2.1)) is that the edges of the random graph are still jointly exchangeable, in the sense that permuting the vertices of $G$ does not change the distribution of the graph $G$. Consequently the degree sequence $\left(d_{1}(G), \ldots, d_{n}(G)\right)^{1}$ marginally have the same distribution for each $i \in[n]$. This may not be desirable for modeling networks where there are a few vertices of very high degree (see [6]), compared to the remaining vertices. Such a feature is often present in social networks, where the vertex corresponding to a popular/famous person has a very high degree compared to the remaining vertices.

One model which captures degree homogeneity is the $\beta$-model of social networks (c.f. $[8,13$, $16,39,44]$ and references there-in). The $\beta$-model is defined by the following p.m.f. on $\mathcal{G}_{n}$ :

$$
\begin{equation*}
\mathbb{P}_{n, \boldsymbol{\beta}}(G):=\frac{1}{Z_{n}(\boldsymbol{\beta})} \exp \left\{\sum_{i=1}^{n} \beta_{i} d_{i}(G)\right\} . \tag{2.2}
\end{equation*}
$$

Here
(i) $\left(d_{1}(G), \ldots, d_{n}(G)\right)$ is the degree sequence of the graph $G$.
(ii) $\boldsymbol{\beta}=\left(\beta_{1}, \ldots, \beta_{n}\right)^{T} \in \mathbb{R}^{n}$ is a vector valued parameter,
(iii) $Z_{n}(\boldsymbol{\beta})$ is the normalizing constant.

In this model, for each vertex $i \in[n]$ there is a real valued parameter $\beta_{i}$ which controls the ${ }^{1} d_{i}(G)=\sum_{j} G_{i j}$ with $\left\{G_{i j}\right\}_{i, j \in[n]}$ being the adjacency matrix of $G$.
effect of the $i^{t h}$ vertex, and consequently the typical size of the degree $d_{i}(G)$. This allows for heterogeneity among the degrees. A large value of $\beta_{i}$ results in a large value of the degree of the $i^{\text {th }}$ vertex, and vice versa. One drawback of the $\beta$-model (2.2) is that the edges of the graph $G$ are no longer dependent. This is not immediate from (2.2), but is not hard to check (see for e.g. [16]). Thus although the $\beta$-model allows for degree heterogeneity, it does not involve dependence between the edges.

A natural way to retain both the dependence between edges and the heterogeneity of the degrees is to consider an exponential family which has both the terms $\theta N(H, G)$ and $\sum_{i=1}^{n} \beta_{i} d_{i}(G)$ in the exponent. Indeed, dependence between edges is present because of the term $\theta N(H, G)$, and degree heterogeneity is present because of the term $\sum_{i=1}^{n} \beta_{i} d_{i}(G)$. Such a model, which we introduce formally below, can be thought of as a degree corrected ERGM.

### 2.1.1 Degree Corrected ERGM

As before, let $\mathcal{G}_{n}$ denote the set of all simple labelled graphs $G$ with vertex set $[n]:=\{1,2, \ldots, n\}$, Given a graph $G \in \mathcal{G}_{n}$, by slight abuse of notation we use $G$ to also denote the adjacency matrix of $G$, defined as follows:

$$
G_{i j}= \begin{cases}1 & \text { If an edge is present between vertices } i \text { and } j \text { in } G,  \tag{2.3}\\ 0 & \text { If no edge is present between vertices } i \text { and } j \text { in } G .\end{cases}
$$

Thus, we encode presence or absence of edges by $\{0,1\}$. By convention, set $G_{i i}:=0$, and note that $G$ is a symmetric $n \times n$ matrix with 0 on the diagonal, and $\{0,1\}$ entries on the off-diagonals. Let $\left(d_{1}, d_{2}, \cdots, d_{n}\right)$ denote the labeled degree sequence of the graph $G$, defined by

$$
d_{i}:=\sum_{j=1}^{n} G_{i j}, 1 \leq i \leq n .
$$

Let $H$ be a fixed connected subgraph with $\zeta:=|V(H)| \geq 2$ (i.e. $H$ is not an isolated vertex). Assume that the vertices of $H$ are labeled as $[\zeta]=\{1,2, \ldots, \zeta\}$. Let $I_{n}$ denote the the set of all

1-1 maps from [ $n$ ] to [ $\zeta$ ]. For any $G \in \mathcal{G}_{n}$, let $N(H, G)$ denote the number of copies of $H$ in $G_{n}$, defined by

$$
N(H, G)=\sum_{l \in \mathcal{I}_{n}} \prod_{(i, j) \in E(H)} G_{\iota(i), \iota(j)},
$$

where $E(H):=\{(a, b) \in V(H):(a, b)$ is an edge in $H\}$ is the edge set of $H$. As for illustration, the expression of $N\left(H, G_{n}\right)$ when $H$ is an edge, triangle, and two star (to be denoted by $K_{2}, K_{3}, K_{1,2}$ respectively) are respectively given by:

$$
\begin{aligned}
& N\left(K_{2}, G\right)=\sum_{i \neq j} G_{i j}=2 \sum_{i<j} G_{i j}=\sum_{i=1}^{n} d_{i}, \\
& N\left(K_{3}, G\right)=\sum_{i \neq j \neq k} G_{i j} G_{j k} G_{k i}=6 \sum_{i<j<k} G_{i j} G_{j k} G_{k i}, \\
& N\left(K_{1,2}, G\right)=\sum_{i \neq j \neq k} G_{i j} G_{i k}=2 \sum_{i=1}^{n} \sum_{j<k} G_{i j} G_{i k}=2 \sum_{i=1}^{n}\binom{d_{i}}{2} .
\end{aligned}
$$

Given a parameter $\theta>0$ and vector $\boldsymbol{\beta}=\left(\beta_{1}, \beta_{2}, \ldots, \beta_{n}\right) \in \mathbb{R}^{n}$, we subsequently define a probability mass function on $\mathcal{G}_{n}$ by setting

$$
\begin{equation*}
\mathbb{P}_{n, \theta, \boldsymbol{\beta}}(G):=\frac{1}{Z_{n}(\boldsymbol{\beta}, \theta, H)} \exp \left\{\frac{\theta}{n^{\zeta-2}} N(H, G)+\sum_{i=1}^{n} \beta_{i} d_{i}\right\} . \tag{2.4}
\end{equation*}
$$

where as usual $Z_{n}(\boldsymbol{\beta}, \theta, H)$ is the normalizing constant. The scaling $n^{\zeta-2}$ ensures that the resulting model is non-trivial as $n \rightarrow \infty$ (c.f. [15]). If $\beta_{i}=\beta_{0}$ for some $\beta_{0} \in \mathbb{R}$ free of $i$, then the model in (2.4) is an Exponential Random Graph Model with two sufficient statistics $N(H, G)$ and $E(G)$, where $E(G)=\frac{1}{2} N\left(K_{2}, G\right)$ is the number of edges in the graph $G$. In this case the random graph $G$ represents a bivariate exchangeable array. More precisely, for any permutation $\pi \in S_{n}$ the graph $G_{\pi}$ defined by $G_{\pi}(i, j):=G_{\pi(i), \pi(j)}$ has the same distribution as $G$, i.e. $G_{\pi} \stackrel{D}{=} G$. The vector of parameters $\boldsymbol{\beta}$, therefore, measures the individual effects of each vertex, and for a general vector $\boldsymbol{\beta}$ a random graph $G$ from the model (2.4) is no longer exchangeable. For $\theta>0$, the term $N(H, G)$ ensures that there is positive dependence among the edges in $G$, in the sense that conditional on presence of an edge, any other edge is more likely to be present. If $\theta=0$, the model (2.4) reduces
to the $\beta$-model as in (2.2), in which all edges $G_{i j}$ are independent, with

$$
\mathbb{P}_{n, 0, \beta}\left(G_{i j}=1\right)=\frac{e^{\beta_{i}+\beta_{j}}}{1+e^{\beta_{i}+\beta_{j}}}
$$

Thus the model in (2.4) combines the features of the $\beta$-model and traditional ERGMs. We will use the term degree corrected ERGM to refer to the model (2.4).

### 2.1.2 Hypothesis Testing Problem for $\boldsymbol{\beta}$

Given the model (2.4), a natural question is to carry out inference regarding the vector $\boldsymbol{\beta}$. In the setting where $\theta=0$, the problem of estimation of $\boldsymbol{\beta}$ using the MLE $\hat{\boldsymbol{\beta}}_{M L}$ was studied in [16], where the authors gave bounds on $\left\|\hat{\boldsymbol{\beta}}_{M L}-\boldsymbol{\beta}\right\|_{\infty}$. The question of testing of the grand null hypothesis $\boldsymbol{\beta}=\mathbf{0}$ versus non negative sparse alternatives was studied in [39], where the authors show that the optimal test depends on the sparsity level and strength of the signal. Since both these papers assumed $\theta=0$, the edges of the graph $G$ were independent, which was used significantly in the proofs of the results. A natural question is whether one can extend these results in the presence of dependence between edges, i.e. when $\theta>0$. In this chapter, we study the question of testing the grand null hypothesis $\boldsymbol{\beta}=\beta_{0} \mathbf{1}$ against sparse one sided alternatives. Essentially we want to test the null hypothesis that all nodes in the network are equally popular (have the same $\beta_{i}$ ), versus the alternative hypothesis that there is a small hub of nodes which are more popular (have a higher value of $\beta_{i}$ ) compared to the baseline popularity $\beta_{0}$ of the remaining nodes. Here $\beta_{0} \in \mathbb{R}$ is a real valued parameter which is assumed to be known. In section 2.2 .3 we briefly discuss what can go wrong if the parameter $\beta_{0}$ is not assumed to be known. Below we formally introduce the testing problem discussed above.

Let $\beta_{0} \in \mathbb{R}$ be known. Let $G$ be a graph drawn from the probability distribution (2.4), and for
a known $\theta>0$ and given $\beta_{0} \in \mathbb{R}$ we consider the following hypothesis testing problem:

$$
\begin{equation*}
\mathcal{H}_{0}: \boldsymbol{\beta}=\beta_{0} \mathbf{1} \quad \text { vs } \quad \mathcal{H}_{1}: \boldsymbol{\beta} \in \Xi(s, A) . \tag{2.5}
\end{equation*}
$$

Here under the null hypothesis we have $\beta_{i}=\beta_{0}$ for all $i \in[n]$ and we denote this null probability measure as $\mathbb{P}_{n, \theta, \beta_{0}}$. The set of vectors $\Xi(s, A)$ in the alternative hypothesis $H_{1}$ is defined as

$$
\begin{equation*}
\Xi(s, A):=\left\{\boldsymbol{\beta}=\beta_{0} \mathbf{1}+\boldsymbol{\mu}:|\operatorname{supp} \boldsymbol{\mu}| \geq s, \text { and } \min _{i \in \operatorname{supp} \mu} \mu_{i} \geq A\right\} . \tag{2.6}
\end{equation*}
$$

In words, under the alternative hypothesis there is a sparse set $S$ of size $s$, such that $\beta_{i} \geq \beta_{0}+A$ if $i \in S$, and $\beta_{i}=\beta_{0}$ if $i \notin S$. Our main goal of this chapter is to study the effect of the nuisance parameter $\theta$ on the hypothesis testing problem (2.5). For studying the proposed hypothesis testing problem, here we adopt an asymptotic minimax framework similar to [39, 40], which is introduced below (see also [11, 31, 32, 33]).

Given a non randomized test function $T_{n}: \mathcal{G}_{n} \mapsto\{0,1\}$, define the risk of test $T_{n}(G)$ as the sum of type I and type II errors, as follows:

$$
\begin{equation*}
R\left(T_{n}, \Xi(s, A), \boldsymbol{\beta}\right):=\mathbb{P}_{n, \theta, \beta_{0}}\left(T_{n}(G)=1\right)+\sup _{\boldsymbol{\beta} \in \Xi(s, A)} \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(T_{n}(G)=0\right) \tag{2.7}
\end{equation*}
$$

Given a sequence of test functions $\left\{T_{n}\right\}_{n \geq 1}$ for the testing problem (2.5), we call $\left\{T_{n}\right\}_{n \geq 1}$ as
(i) Asymptotically Powerful, if

$$
\begin{equation*}
\lim _{n \rightarrow \infty} R\left(T_{n}, \Xi(s, A), \boldsymbol{\beta}\right)=0 \tag{2.8}
\end{equation*}
$$

(ii) Asymptotically not Powerful, if

$$
\begin{equation*}
\liminf _{n \rightarrow \infty} R\left(T_{n}, \Xi(s, A), \boldsymbol{\beta}\right)>0 \tag{2.9}
\end{equation*}
$$

(iii) Asymptotically Powerless, if

$$
\begin{equation*}
\lim _{n \rightarrow \infty} R\left(T_{n}, \Xi(s, A), \boldsymbol{\beta}\right)=1 \tag{2.10}
\end{equation*}
$$

By definition, both type I and type II errors converge to 0 for asymptotically powerful tests. Also, if a sequence of tests is asymptotically powerless, then it is also asymptotically not powerful, and so (iii) is a stronger notion than (ii).

### 2.2 Main Results

In this section we present and discuss our main results. To that end, we first consider general degree corrected ERGMs and analyze the performance of two natural tests. We then focus on a particular degree corrected ERGM, where the graph $H$ is a two star. In this setting we show that the general tests studied above attains the "optimal detection boundary" for all configurations $\left(\theta, \beta_{0}\right)$ barring a specific point, which we refer to as the critical point/configuration. At this point, using a slightly different test from the ones studied under the general ERGM framework, we are able to detect much lower signals, compared to the independent case $(\theta=0)$.

### 2.2.1 General Degree Corrected ERGMs

In this section, we discuss the hypothesis testing problem (2.5) in the setting of general degree corrected ERGMs as in (2.4). Specifically, we will show how signal density and strength $(s, A)$ coordinate to determine the threshold for testing efficiency. Two natural test statistics for this problem are the sum of degrees $\sum_{i=1}^{n} d_{i}(G)$, and the maximum degree $\max _{i \in[n]} d_{i}(G)$. However, because of the presence of dependence, it is very difficult to calibrate the cut-off for these statistics, as they depend on the parameter $\theta$ in a non-trivial way. To counter this, we use conditionally centered versions of the sum of degrees, and the maximum degree, similar to what was done in [40].

Our first theorem studies the performance of a test based on the conditionally centered sum of
degrees. For stating the result we require a few notations.

Definition 2.2.1. Let $\mathcal{E}:=\{(i, j): 1 \leq i<j \leq n\}$ be the set of all edges in the complete graph $K_{n}$. For any $e=(i, j) \in \mathcal{E}$, let $N_{e}(H, G)$ denote the number of copies of $H$ in the graph $G$ which contains the edge e, and let $N_{e, f}(H, G)$ denote the number of copies of $H$ in the graph $G$ which contains both the edges e, $f$.

Setting $\psi(x):=\frac{e^{x}}{1+e^{x}}$ for $x \in \mathbb{R}$, for any $e=(i, j) \in \mathcal{E}$ we have

$$
\begin{equation*}
\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left(G_{e} \mid G_{f}: f \neq e\right)=\psi\left(\theta t_{e}(H, G)+\beta_{i}+\beta_{j}\right), \tag{2.11}
\end{equation*}
$$

where $t_{e}(H, G):=\frac{N_{e}(H, G)}{n^{\zeta-2}}$.
Since our results are asymptotic in nature, below we introduce some standard notations, to be used in the remainder of the chapter.

Definition 2.2.2. Given two sequence of real numbers $\left\{a_{n}\right\}_{n \geq 1}$ and $\left\{b_{n}\right\}_{n \geq 1}$, we use the notation $a_{n}=O\left(b_{n}\right)$ or $a_{n} \lesssim b_{n}$ to imply the existence of a positive finite constant $c$ free of $n$, such that $a_{n} \leq c b_{n}$. We use the notation $a_{n} \gg b_{n}\left(a_{n} \ll b_{n}\right)$ to imply $\lim _{n \rightarrow \infty} \frac{a_{n}}{b_{n}}=\infty\left(\lim _{n \rightarrow \infty} \frac{a_{n}}{b_{n}}=0\right.$ respectively).

Theorem 2.2.1. With $G$ from the model (2.4), consider the hypothesis testing problem described in (2.5). If $s A \rightarrow \infty$, then for any sequence $L_{n}$ such that $n \ll L_{n} \ll n s A$ the conditionally centered sum of degrees test $T_{n}(G)$ given by

$$
\begin{aligned}
T_{n}(G) & =1 \text { if } \sum_{e \in \mathcal{E}}\left[G_{e}-\mathbb{E}_{n, \theta, \boldsymbol{\beta}_{0}}\left(G_{e} \mid G_{f}: f \neq e\right)\right]>L_{n} \\
& =0 \text { otherwise }
\end{aligned}
$$

is asymptotically powerful.

In settings where the signal size $s$ is small, a test based on the conditionally centered maximum of degrees can sometimes detect lower signals. The performance of this test is studied in our
second result.

Theorem 2.2.2. With $G$ from the model (2.4), consider the hypothesis testing problem described in (2.5). Then there exists constants $\kappa, C$ such that if $A \geq \kappa \sqrt{\frac{\log n}{n}}$ and $L_{n}=C \sqrt{n \log n}$, then the conditionally centered maximum degree test defined by

$$
T_{n}(G)= \begin{cases}1 & \text { If } \max _{i \in[n]} \sum_{e \ni i}\left[G_{e}-\mathbb{E}_{n, \theta, \boldsymbol{\beta}_{0}}\left(G_{e} \mid G_{f}: f \neq e\right)\right]>L_{n}  \tag{2.12}\\ 0 & \text { Otherwise }\end{cases}
$$

is asymptotically powerful.

Comparing Theorem 2.2.1 and 2.2.2 yields that the conditionally centered maximum degree test is better (has a lower detection boundary) for sparser alternative ( $s \ll \sqrt{\frac{n}{\log n}}$ ), and the conditionally centered sum of degrees test is better for denser alternatives $\left(s \gg \sqrt{\frac{n}{\log n}}\right)$. This is similar to the findings of [39], where it was shown that optimal rate detection is obtained by the sum of degrees if $s=n^{b}$ with $b>1 / 2$ (see [39, Theorem 3.1]), and by the maximum degree test if $b<1 / 2$ (see [39, Theorem 3.3]).

### 2.2.2 Degree Corrected Two-star ERGM

In Theorems 2.2.1 and 2.2.2, there is no effect of the nuisance parameter $\theta$ on the detection rate of the tests. To demonstrate that the best possible detection rate can change depending on the value of $\theta$, we study in detail the degree corrected two star ERGM, The two star is the graph $K_{1,2}$, which is a path of length 3. For notational and computational convenience, for the Degree Corrected Two-star ERGM our edge variables take values in $\{-1,1\}$ instead of $\{0,1\}$. More precisely, given a graph $G \in \mathcal{G}_{n}$, our adjacency matrix $Y$ is now defined as follows:

$$
\begin{aligned}
Y_{i j} & =+1 \text { if }(i, j) \text { is an edge in } G, \\
& =-1 \text { if }(i, j) \text { is not an edge in } G .
\end{aligned}
$$

As before, we set $Y_{i i}=0$ by convention. Thus $Y$ is a symmetric matrix with $\{-1,1\}$ entries, and 0 on the diagonal. Let $\left(k_{1}, k_{2}, \cdots, k_{n}\right)$ denote the labeled "degree sequence" of the graph $Y$, i,e,

$$
k_{i}:=\sum_{j=1}^{n} Y_{i j}, 1 \leq i \leq n .
$$

The following display introduces the degree corrected two star ERGM as a p.m.f. on $\{-1,1\}\left(\begin{array}{c}\binom{n}{2}\end{array}\right.$

$$
\begin{equation*}
\mathbb{P}_{n, \theta, \boldsymbol{\beta}}(Y)=\frac{1}{Z_{n}(\boldsymbol{\beta}, \theta)} \exp \left\{\frac{\theta}{n-1} \widetilde{N}\left(K_{1,2}, G_{n}\right)+\frac{1}{2} \sum_{i=1}^{n} \beta_{i} k_{i}\right\} \tag{2.13}
\end{equation*}
$$

where

$$
\widetilde{N}\left(K_{1,2}, G_{n}\right):=\sum_{i=1}^{n} \sum_{j<k} Y_{i j} Y_{i k}=\frac{1}{2} \sum_{i=1}^{n} k_{i}^{2}-\frac{n(n-1)}{2} .
$$

Having observed $Y$, consider the same hypothesis testing problem (2.5) as above. For the sake of clarity of presentation, in this section we parametrize the signal size $s$ and signal strength $A$ by $n^{b}$ and $n^{t}$ respectively, where $b \in(0,1)$ and $t<0$. The detection boundary for this problem shows a phase transition depending on the nuisance parameter $\theta$. Stating this requires the following partitioning of the parameter space for $\left(\theta, \beta_{0}\right)$ :

Definition 2.2.3. - Let $\Theta_{1}=\Theta_{11} \cup \Theta_{12}$, where $\Theta_{11}:=(0,1 / 2) \times\{0\}$, and $\Theta_{12}=\left\{\left(\theta, \beta_{0}\right)\right.$ : $\left.\theta>0, \beta_{0} \neq 0\right\}$.

- Let $\Theta_{2}:=(1 / 2, \infty) \times\{0\}$.
- Let $\Theta_{3}:=(1 / 2,0)$, usually referred as critical point. Note that $\Theta_{1} \cup \Theta_{2} \cup \Theta_{3}=(0, \infty) \times \mathbb{R}$.

Our first result describes the detection boundary for the degree corrected two star ERGM if $\left(\theta, \beta_{0}\right) \in \Theta_{1}$.

Theorem 2.2.3. Let $Y$ be an observation from from (2.13), and assume $\left(\theta, \beta_{0}\right) \in \Theta_{1}$. Consider the hypothesis testing problem described in (2.5) with $s=n^{b}$ and $A=n^{t}$ for $b \in(0,1)$ and $t<0$.
(a) If $b \geq \frac{1}{2}$ and $b+t<0$, all tests are asymptotically powerless.
(b) If $b \geq \frac{1}{2}$ and $b+t>0$, then the conditionally centered sum test of Theorem 2.2.1 is asymptotically powerful.
(c) If $b<\frac{1}{2}$ and $t+\frac{1}{2} \leq 0$ then all tests are asymptotically powerless.
(d) If $b<\frac{1}{2}$ and $t+\frac{1}{2}>0$ then the conditionally centered max test of Theorem 2.2.2 is asymptotically powerful.

Our second result describes the detection boundary for the degree corrected two star ERGM if $\left(\theta, \beta_{0}\right) \in \Theta_{2}$.

Theorem 2.2.4. Let $Y$ be an observation from from (2.13), and assume $\left(\theta, \beta_{0}\right) \in \Theta_{2}$. Consider the hypothesis testing problem described in (2.5) with $s=n^{b}$ and $A=n^{t}$ for $b \in(0,1)$ and $t<0$.
(a) If $b \geq \frac{1}{2}$ and $b+t<0$, all tests are asymptotically not powerful.
(b) If $b \geq \frac{1}{2}$ and $b+t>0$, then the conditionally centered sum test of Theorem 2.2.1 is asymptotically powerful.
(c) If $b<\frac{1}{2}$ and $t+\frac{1}{2} \leq 0$ then all tests are asymptotically not powerful.
(d) If $b<\frac{1}{2}$ and $t+\frac{1}{2}>0$ then the conditionally centered max test of Theorem 2.2.2 is asymptotically powerful.

Note that at a qualitative level, the detection boundary in the regimes $\Theta_{1}$ and $\Theta_{2}$ are the same. The only difference is that below the detection boundary, in domain $\Theta_{1}$ Theorem 2.2.3 shows that all tests are powerless, and in domain $\Theta_{2}$ Theorem 2.2 . shows that all tests are asymptotically not powerful. On the other hand, something fundamentally different happens in the critical domain $\Theta_{3}$, which corresponds to the choice $\left(\theta, \beta_{0}\right)=(1 / 2,0)$. In this case the optimal testing threshold is significantly lower than the other regimes, and does not depend on whether $b<1 / 2$ or $b>1 / 2$. Moreover, this improved performance does not follow from either Theorem 2.2.1 or 2.2.2. In this case a test based on the unconditional sum of degrees attains the optimal detection boundary, for all values of $(s, A)$. This is explained in our final result below.

Theorem 2.2.5. Let $Y$ be an observation from from (2.13), and assume $\left(\theta, \beta_{0}\right)=\left(\frac{1}{2}, 0\right)$. Consider the hypothesis testing problem described in (2.5), with $s=n^{b}$ and $A=n^{t}$ for some $b \in(0,1)$ and $t<0$.
(a) If $b+t+\frac{1}{2}<0$, then all tests are asymptotically not powerful.
(b) If $b+t+\frac{1}{2}>0$, then the total degree test $T_{n}($.$) defined by$

$$
\begin{aligned}
T_{n}(G) & =1 \text { if } \sum_{i=1}^{n} k_{i}>L_{n}, \\
& =0 \text { otherwise }
\end{aligned}
$$

is asymptotically powerful for some sequence $L_{n}$ satisfying $L_{n} \gg n^{3 / 2}$.

This demonstrates that the much weaker criterion $b+t+\frac{1}{2}>0$ is enough for detection at criticality, whereas away from criticality we need stronger conditions on $b, t$. Similar phenomenon of improved detection at criticality have been observed for Ising models [18, 39, 40]. Given that the two star ERGM can be viewed as an Ising model, it is thus not surprising that this continues to hold here. A summary of the detection boundary for the degree corrected two star ERGM is given in figure 2.1 below.

### 2.2.3 Main Contributions and Future Scope

In this chapter we introduce the degree corrected ERGM, which combines traditional ERGMs with the $\boldsymbol{\beta}$-model and thereby allowing for not degree heterogeneity but also dependence between the edges. In this setting, we study the performance of two tests, based on conditionally centered sum of degrees, and conditionally centered maximum degree. The detection rate of these two tests match the performance of the corresponding tests based on the unconditionally centered sum of degree and unconditional maximum degree, respectively, in the independent case $(\theta=0)$. To explore the sharpness of these general tests, we subsequently study the degree corrected two star ERGM in detail. Here we show that in all parameter configurations other than $\left(\theta, \beta_{0}\right)=(1 / 2,0)$,


Figure 2.1: In this figure, we plot $(b, t)$ along $X$ and $Y$ axis respectively, where $s=n^{b}$ is the size of the signal set, and $A=n^{t}$ is the magnitude of the signal. The range of $b$ is $(0,1)$, and the range of $t$ is $(-\infty, 0)$. The deep blue portion of the plot represents the pairs $(b, t)$ where detection is possible in all regimes $\Theta_{1} \cup \Theta_{2} \cup \Theta_{3}$. The light blue portion of the plot represents the pairs $(b, t)$ where detection is possible $\Theta_{3}$, but not for $\Theta_{1} \cup \Theta_{2}$. Finally, the grey portion of the plot represents the pairs $(b, t)$ where detection is impossible in all regimes $\Theta_{1} \cup \Theta_{2} \cup \Theta_{3}$. Also note that in $\Theta_{1} \cup \Theta_{2}$ the optimal test depends on whether $b<1 / 2$ or $b>1 / 2$, whereas in $\Theta_{3}$ the optimal test does not depend on $b$.
the optimal detection boundary is attained by one of the conditionally centered tests. At the critical configuration $\left(\theta, \beta_{0}\right)=(1 / 2,0)$, we show that the optimal detection rate is significantly improved, and this optimal rate is attained by a test based on the unconditionally centered sum of degrees.

Throughout this chapter we assume that the parameters $\left(\theta, \beta_{0}\right)$ are known. If $\left(\theta, \beta_{0}\right)$ is unknown, it may be possible to estimate $\left(\theta, \beta_{0}\right)$ if the signal $(s, A)$ is small, by ignoring the signals altogether and estimating the parameters via the null model MLE/pseudo-likelihood. However such a strategy is hopeless for all values of $(s, A)$, without the knowledge of $\left(\theta, \beta_{0}\right)$. Indeed, consider the following extreme configuration when $s=n, A=\infty$, in which case the graph $G$ equals $K_{n}$ with probability 1 for any value of $\beta_{0}$. On the other hand, if $s=A=0$, but $\theta=\infty$, the observed graph is again $K_{n}$ with probability 1 for any value of $\beta_{0}$. Thus having observed $G$, it is impossible to decide whether signal is present or absent, if we are not told the value of $\theta$. It remains to be seen to what extent a partial knowledge of $\left(\theta, \beta_{0}\right)$ can help in our testing problem.

The analysis of the conditionally centered sum and maximum of degrees for general (degree corrected) ERGMs is achieved using concentration results based on the method of exchangeable pairs ([12]). Focusing on the degree corrected two star ERGM, to verify the improved detection rate at criticality, we introduce a continuous auxiliary variable $\phi \in \mathbb{R}^{n}$ (similar to [40]), and show that a suitable function of $\phi$ is stochastically much larger under the alternative than under the null hypothesis. Using this, we show that the unconditional sum of degrees is stochastically much larger under the alternative, which gives the improved detection at criticality. The lower bound argument uses the second moment method, which reduces to bounding the correlation between the degrees under the alternative. In the regimes $\Theta_{1}$ and $\Theta_{3}$, using GHS inequality ([37]) we can bound the correlations between the edges under the alternative by the correlation under the null, for which bounds are available from [41], using exchangeability of the null model. In the regime $\Theta_{2}$ we need to do a conditional second moment argument restricted to the set where the degrees are large. In the absence of a conditional GHS inequality, we have to directly bound the conditional correlations between the edges under the alternative. To do this, we make crucial use of the auxiliary variable $\phi$ and set up a recursive equation involving the correlations between degrees of the graph. This recursion leads to a uniform bound on the correlations which is also a tight upper bound (in terms of rate), and suffices for the second moment argument. It is of interest to see if one can set up similar recursive equations to bound correlation between edges in general (degree corrected) ERGMs, in presence/absence of auxiliary variables.

In this chapter we focus on the optimal detection rates while studying the detection boundary. A natural follow up question is to study existence of sharp constants (depending on $\theta, \beta_{0}$ ) which controls the detection boundary for the degree corrected two star ERGMs. Similar to [39], we expect a sharp phase transition (i.e. existence of a constant which determines the optimal detection boundary) in the regime $b<1 / 2$, when $\left(\theta, \beta_{0}\right) \neq(1 / 2,0)$. We believe that to attain optimal detection constants, one needs to study a conditionally centered version of the Higher Criticism Test in the regime $1 / 4<b<1 / 2$, wheres the maximum test should suffice in the regime $b<1 / 2$. Going beyond the two star case, it is of interest to find optimal detection rates, both away from, and
at, "criticality", for general degree corrected ERGMs. A major challenge in carrying out the lower bound argument beyond the two star case is the absence of tight correlation bounds for general ERGMs, both under the null and alternative hypotheses.

### 2.2.4 Outline

The outline of this Chapter is as follows. In section 2.3 we verify results Theorems 2.2.1 and 2.2.2. In section 2.3 .5 we verify Theorems 2.2 .3 and 2.2.4. The proofs of the results of section 2.3.5 uses some supporting lemmas, the proofs of which is deferred to section 3.7.

### 2.3 Proofs of Main Theorems

We will need the following concentration bound for conditionally centered linear statistics for proving the results of this section. The proof of this lemma is similar to [19, Lemma 2.1] and [40, Lemma 1].

Lemma 2.3.1. Let $G$ be a random graph from the model (2.4). Then for any arbitrary collection of positive numbers $\left\{c_{e}\right\}_{e \in \mathcal{E}}$ and any $x>0$ we have

$$
\begin{equation*}
\mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\left|\sum_{e \in \mathcal{E}} c_{e}\left(G_{e}-\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left(G_{e} \mid G_{f}: f \neq e\right)\right)\right|>x\right) \leq 2 \exp \left\{-\frac{x^{2}}{\lambda \sum_{e \in \mathcal{E}} c_{e}^{2}}\right\} \tag{2.14}
\end{equation*}
$$

where $\lambda=\lambda(\theta, H)$ is a constant depending only on $\theta>0$ and the subgraph $H$.

Proof. Produce an exchangeable pair ( $G, G^{\prime}$ ) in the following way:
Pick a random vertex pair $I$ of the uniformly from the set $\mathcal{E}$ with cardinality $N=\binom{n}{2}$. If $I=e$, replace the random variable $G_{e}$ by $G_{e}^{\prime}$ a pick from the conditional distribution given $\left\{G_{f}, f \neq e\right\}$. Let this new graph be denoted by $G^{\prime}$. It is easy to verify that ( $G, G^{\prime}$ ) is indeed an exchangeable
pair. Setting $J(G):=\sum_{e \in \mathcal{E}} c_{e} G_{e}$, note that

$$
\begin{aligned}
h(G):=\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left(J(G)-J\left(G^{\prime}\right) \mid G\right) & =\frac{1}{N} \sum_{e \in \mathcal{E}} c_{e}\left(G_{e}-\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left(G_{e} \mid G_{f}: f \neq e\right)\right) \\
& =\frac{1}{N} J(G)-\frac{1}{N} \sum_{e \in \mathcal{E}} c_{e} \frac{\exp \left\{\frac{\theta}{n^{\zeta-2}} N_{e}(H, G)+\beta_{e}\right\}}{1+\exp \left\{\frac{\theta}{n^{\zeta-2}} N_{e}(H, G)+\beta_{e}\right\}},
\end{aligned}
$$

where $N_{e}(H, G)$ is the number of copies of $H$ in the graph $G$, which contains the edge $e$. Using the fact that the derivative of the function $\psi(x)=\frac{e^{x}}{1+e^{x}}$ is bounded by $\frac{1}{4}$, this gives

$$
\begin{aligned}
\left|h(G)-h\left(G^{\prime}\right)\right| & \leq \frac{\left|c_{I}\right|}{N}+\frac{|\theta|}{4 N n^{\zeta-2}} \sum_{e \in \mathcal{E}}\left|c_{e}\right|\left|N_{e}(H, G)-N_{e}\left(H, G^{\prime}\right)\right| \\
& \leq \frac{\left|c_{I}\right|}{N}+\frac{|\theta|}{4 N n^{\zeta-2}} \sum_{e \in \mathcal{E}}\left|c_{e}\right| N_{e, I}\left(H, K_{n}\right),
\end{aligned}
$$

where $N_{e, f}\left(H, K_{n}\right)$ is the number of copies of $H$ in the complete graph $K_{n}$ passing through both the edges $e$ and $f$. Consequently, we have

$$
\begin{aligned}
& \left|\mathbb{E}_{n, \theta, \beta}\left(\left(h(G)-h\left(G^{\prime}\right)\right)\left(J(G)-J\left(G^{\prime}\right)\right) \mid G\right)\right| \\
& \leq \frac{1}{N} \sum_{f \in \mathcal{E}}\left|c_{f}\right|\left[\frac{\left|c_{f}\right|}{N}+\frac{|\theta|}{4 N n^{\zeta-2}} \sum_{e \in \mathcal{E}}\left|c_{e}\right| N_{e, f}\left(H, K_{n}\right)\right] \\
= & \frac{1}{N^{2}} \sum_{f \in \mathcal{E}} c_{f}^{2}+\frac{|\theta|}{4 N^{2} n^{\zeta-2}} \sum_{e, f \in \mathcal{E}} N_{e, f}\left(H, K_{n}\right)\left|c_{e}\right|\left|c_{f}\right| \\
= & \frac{1}{N^{2}} \sum_{e, f \in \mathcal{E}} B_{N}(e, f)\left|c_{e} \|\left|c_{f}\right|,\right.
\end{aligned}
$$

where $B_{N}$ is a $N \times N$ symmetric matrix defined by:

$$
B_{N}(e, f):= \begin{cases}1 & \text { if } e=f \\ \frac{|\theta|}{4 n^{\zeta-2}} N_{e, f}\left(H, K_{n}\right) & \text { if } e \neq f\end{cases}
$$

Now for any $e \neq f$ we have

$$
\begin{aligned}
N_{e, f}\left(H, K_{n}\right) & \lesssim n^{\zeta-4} \text { if } e \text { and } f \text { have no vertex in common, } \\
& \lesssim n^{\zeta-3} \text { if } e \text { and } f \text { have one vertex in common. }
\end{aligned}
$$

This gives

$$
\max _{e \in \mathcal{E}} \sum_{f \in \mathcal{E}} B_{N}(e, f) \lesssim 1+n^{2} \frac{1}{n^{\zeta-2}} n^{\zeta-4}+n \frac{1}{n^{\zeta-2}} n^{\zeta-3} \lesssim 1,
$$

which in turn implies that the operator norm of the matrix $B_{N}$ is $O(1)$, and consequently,

$$
\left|\mathbb{E}_{n, \theta, \beta}\left(\left(h(G)-h\left(G^{\prime}\right)\right)\left(J(G)-J\left(G^{\prime}\right)\right) \mid G\right)\right| \lesssim \frac{1}{N^{2}} \sum_{e \in \mathcal{E}} c_{e}^{2} \lesssim \frac{1}{n^{4}} \sum_{e \in \mathcal{E}} c_{e}^{2}
$$

Then by Stein's Method for concentration inequalities as in [12, Theorem 1.5], the conclusion of the lemma follows.

### 2.3.1 Proof of Theorem 2.2.1

To begin, using Lemma 2.3.1 with $c_{e}=1$ for all $e \in \mathcal{E}$ gives the existence of a constant $\lambda$ (depending only on $\theta, H$ ) such that

$$
\begin{equation*}
\mathbb{P}_{\mathcal{H}_{0}}\left(\left|\sum_{e \in \mathcal{E}}\left(G_{e}-\mathbb{E}_{n, \boldsymbol{\beta}_{0}, \theta}\left(G_{e} \mid G_{f}: f \neq e\right)\right)\right|>L_{n}\right) \leq 2 \exp \left\{-\frac{L_{n}^{2}}{\lambda\binom{n}{2}}\right\} \longrightarrow 0, \tag{2.15}
\end{equation*}
$$

where the last limit uses $L_{n} \gg n$. This shows that type I error converges to 0 .
It thus remains to show that type II error converges to 0 . To this effect, note that $t_{e}(H, G) \leq$ $t_{e}\left(H, K_{n}\right)$ which is bounded, and so therefore there exist a constant $\delta>0$ such that

$$
\begin{align*}
\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left(G_{e} \mid G_{f}: f \neq e\right)-\mathbb{E}_{n, \boldsymbol{\beta}_{0}, \theta}\left(G_{e} \mid G_{f}: f \neq e\right) & =\psi\left(\theta t_{e}(H, G)+\beta_{i}+\beta_{j}\right)-\psi\left(\theta t_{e}(H, G)+2 \beta_{0}\right) \\
& \geq \delta \min \left\{\beta_{i}+\beta_{j}-2 \beta_{0}, 1\right\} \tag{2.16}
\end{align*}
$$

Adding this gives

$$
\sum_{e \in \mathcal{E}}\left(\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left(G_{e} \mid G_{f}: f \neq e\right)-\mathbb{E}_{n, \boldsymbol{\beta}_{0}, \theta}\left(G_{e} \mid G_{f}: f \neq e\right)\right) \geq \delta n s A
$$

Since $L_{n} \ll n s A$, for all $n$ large we have

$$
\begin{aligned}
& \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\sum_{e \in \mathcal{E}}\left(G_{e}-\mathbb{E}_{n, \boldsymbol{\beta}_{0}, \theta}\left(G_{e} \mid G_{f}: f \neq e\right)\right) \leq L_{n}\right) \\
\leq & \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\left|\sum_{e \in \mathcal{E}}\left(G_{e}-\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left(G_{e} \mid G_{f}: f \neq e\right)\right)\right| \geq L_{n}\right) \leq 2 \exp \left\{-\frac{L_{n}^{2}}{\lambda n^{2}}\right\} .
\end{aligned}
$$

where we again invoke Lemma 2.3.1 in the last line above. This gives

$$
\sup _{\beta \in \Xi(s, A)} \mathbb{P}_{n, \beta, \theta}\left(\left|\sum_{e \in \mathcal{E}}\left(G_{e}-\mathbb{E}_{n, \boldsymbol{\beta}_{0}, \theta}\left(G_{e} \mid G_{f}: f \neq e\right)\right)\right| \leq L_{n}\right) \leq 2 \exp \left\{-\frac{L_{n}^{2}}{\lambda\binom{n}{2}}\right\},
$$

which converges to 0 as $L_{n} \gg n$. This completes the proof of the theorem.

### 2.3.2 Proof of Theorem 2.2.2

As in the previous theorem, it suffices to show that both type I and type II errors converge to 0 . For estimating the type I error, using a union bound gives

$$
\begin{align*}
& \mathbb{P}_{\mathcal{H}_{0}}\left(\max _{1 \leq i \leq n} \mid \sum_{e \ni i}\left(G_{e}-\mathbb{E}_{n, \boldsymbol{\beta}_{0}, \theta}\left(G_{e} \mid G_{f}: f \neq e\right) \mid>C \sqrt{n \log n}\right)\right. \\
& \leq \sum_{i=1}^{n} \mathbb{P}_{\mathcal{H}_{0}}\left(\left\lvert\, \sum_{e \ni i}\left(G_{e}-\mathbb{E}_{n, \boldsymbol{\beta}_{0}, \theta}\left(G_{e} \mid G_{f}: f \neq e\right) \mid>C \sqrt{n \log n}\right) \leq n \exp \left\{-\frac{C^{2} n \log n}{\lambda(n-1)}\right\}\right.,\right. \tag{2.17}
\end{align*}
$$

where the last inequality uses Lemma 2.3 .1 with $c_{e}=1$ if $e \ni i$, and 0 otherwise. For the choice $C>\sqrt{\lambda}$ the RHS above converges to 0 , and so Type I error converges to 0 .

For estimating the Type II error, fix vertex $i$ such that $\beta_{i} \geq A$. Then using (2.16) gives

$$
\sum_{e \ni i}\left(\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left(G_{e} \mid G_{f}: f \neq e\right)-\mathbb{E}_{n, \boldsymbol{\beta}_{0}, \theta}\left(G_{e} \mid G_{f}: f \neq e\right)\right) \geq \delta n \min \{A, 1\}
$$

Since $A \geq \kappa \sqrt{\frac{\log n}{n}}$, for all $n$ large we have

$$
\delta n \min \{A, 1\} \geq \delta \kappa \sqrt{n \log n} \geq 2 C \sqrt{\log n}
$$

for the choice $\kappa=\frac{2 C}{\delta}$. This gives

$$
\begin{aligned}
& \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\sum_{e \ni i}\left(G_{e}-\mathbb{E}_{n, \theta, \beta_{0} 1}\left(G_{e} \mid G_{f}: f \neq e\right) \leq C \sqrt{n \log n}\right)\right. \\
& \leq \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\left\lvert\, \sum_{e \ni i}\left(G_{e}-\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left(G_{e} \mid G_{f}: f \neq e\right) \mid \geq C \sqrt{n \log n}\right) \leq 2 \exp \left\{-\frac{C^{2} n \log n}{\lambda(n-1)}\right\}\right.,\right.
\end{aligned}
$$

where the last inequality again uses Lemma 2.3.1. Thus we have shown

$$
\sup _{\beta \in \Xi(s, A)} \mathbb{P}_{n, \beta, \theta}\left(\left|\sum_{e \ni i}\left(G_{e}-\mathbb{E}_{n, \boldsymbol{\beta}_{0}, \theta}\left(G_{e} \mid G_{f}: f \neq e\right)\right)\right| \leq C \sqrt{n \log n}\right) \leq 2 \exp \left\{-\frac{C^{2} n \log n}{\lambda(n-1)}\right\},
$$

which converges to 0 as before for the choice $C>\sqrt{\lambda}$.

### 2.3.3 Proof of part (b) and (d) of Theorem 2.2.3 and Theorem 2.2.4

Part (b) follows by a direct application of Theorem 2.2.1, on noting that $s A=n^{b+t} \rightarrow \infty$ if $b+t>0$. Similarly, part (d) follows by a direct application of Theorem 2.2.2, on noting that $A=n^{t} \gg \sqrt{\frac{\log n}{n}}$ if $t>-\frac{1}{2}$. Both Theorem 2.2.1 and Theorem 2.2.2 were proved for $\{0,1\}$ valued random variables, but essentially the same proof goes through for $\{-1,1\}$ valued random variables.

### 2.3.4 Proof of Theorem 2.2.5 part (b)

To prove Theorem 2.2.5 part (b) (as well as parts (a) and (c) of Theorem 2.2.4 later), we express the two star model as a mixture of $\beta$ models by introducing auxiliary variables, as done in [41, 43].

Suppose $Y$ be a random graph from degree corrected two-star model (2.13). Conditional on $Y$, let $\left(\phi_{1}, \cdots, \phi_{n}\right)$ be mutually independent components, with

$$
\begin{equation*}
\phi_{i} \sim N\left(\frac{k_{i}}{n-1}, \frac{1}{\theta(n-1)}\right) . \tag{2.18}
\end{equation*}
$$

The joint distribution of $(\phi, Y)$ is computed in the following Proposition. The proof of this is deferred to the appendix (section 3.7).

Proposition 2.3.1. (a) Given $\phi$, the random variables $(Y)_{1 \leq i<j \leq n}$ are mutually independent, with

$$
\mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(Y_{i j}=1 \mid \phi\right)=\frac{e^{\theta\left(\phi_{i}+\phi_{j}\right)+\frac{1}{2}\left(\beta_{i}+\beta_{j}\right)}}{e^{\theta\left(\phi_{i}+\phi_{j}\right)+\frac{1}{2}\left(\beta_{i}+\beta_{j}\right)}+e^{-\theta\left(\phi_{i}+\phi_{j}\right)-\frac{1}{2}\left(\beta_{i}+\beta_{j}\right)}} .
$$

(b) The marginal density of $\phi$ (w.r.t. Lebesgue measure) is proportional to

$$
\begin{equation*}
f_{n, \theta, \boldsymbol{\beta}}(\phi):=\exp \left\{-\sum_{i<j} p_{i j}\left(\phi_{i}, \phi_{j}\right)\right\}, \tag{2.19}
\end{equation*}
$$

where $p_{i j}(x, y)$ equals

$$
\begin{align*}
& \frac{\theta}{2}\left(x^{2}+y^{2}\right)-\log \cosh \left[\theta(x+y)+\frac{1}{2}\left(\beta_{i}+\beta_{j}\right)\right]  \tag{2.20}\\
= & \frac{\theta}{4}(x-y)^{2}+q\left(\frac{x+y}{2}\right)+\log \cosh (\theta(x+y))-\log \cosh \left(\theta(x+y)+\frac{1}{2}\left(\beta_{i}+\beta_{j}\right)\right),
\end{align*}
$$

with

$$
\begin{equation*}
q(x):=\theta x^{2}-\log \cosh (2 \theta x) . \tag{2.21}
\end{equation*}
$$

We now state the following lemma, which is the analogue of [41, Lemma 4.1]. The proof of these lemmas are deferred to the appendix (3.7).

Lemma 2.3.2. Suppose $\theta=1 / 2$, and $\boldsymbol{\beta} \in\left[0, n^{-1 / 2}\right]^{n}$. Then for any positive integer $\ell \in \mathbb{N}$, there
exist a constant $C$ depending only on $\ell, \theta$ such that

$$
\max _{1 \leq i \leq n} \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left|\phi_{i}-\bar{\phi}\right|^{l} \leq C n^{-l / 2}
$$

Proof of Theorem 2.2.5 part (b). We begin by claiming the existence of a sequence of positive reals $K_{n} \rightarrow \infty$ such that

$$
\begin{equation*}
\lim _{n \rightarrow \infty} \sup _{\boldsymbol{\beta} \in \Xi(s, A)} \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\tanh (\bar{\phi}) \leq n^{-1 / 4} K_{n}\right)=0 . \tag{2.22}
\end{equation*}
$$

Given (2.22), we first finish the proof of the theorem. Note that

$$
\begin{align*}
\sum_{i<j}\left[Y_{i j}-\tanh (\bar{\phi})\right] & =\sum_{i<j}\left[\tanh \left(\frac{\phi_{i}+\phi_{j}}{2}+\frac{\beta_{i}+\beta_{j}}{2}\right)-\tanh (\bar{\phi})\right] \\
& \geq \sum_{i<j}\left[\tanh \left(\frac{\phi_{i}+\phi_{j}}{2}\right)-\tanh (\bar{\phi})\right]  \tag{2.23}\\
& \gtrsim-\sum_{i<j}\left(\frac{\phi_{i}+\phi_{j}}{2}-\bar{\phi}\right)^{2} \gtrsim-n \sum_{i=1}^{n}\left(\phi_{i}-\bar{\phi}\right)^{2} .
\end{align*}
$$

Using (2.22) and Lemma 2.3.2 along with the above display we have

$$
\lim _{n \rightarrow \infty} \sup _{\boldsymbol{\beta} \in \Xi(s, A)} \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\sum_{i<j} Y_{i j} \leq n^{3 / 2} K_{n}\right)=0
$$

and so Type II error converges to 0 . Since

$$
\mathbb{P}_{\mathcal{H}_{0}}\left(\sum_{i<j} Y_{i j}>n^{\frac{3}{2}} K_{n}^{\frac{1}{2}}\right) \rightarrow 0
$$

using [41, Theorem 1.1], Type I error converges to 0 as well. This shows that the test which rejects for large values of $\sum_{i<j} Y_{i j}$ is asymptotically powerful.

It thus remains to verify (2.22). To this end, assume without loss of generality that

$$
\begin{aligned}
\beta_{i} & =A \text { if } 1 \leq i \leq s \\
& =0 \text { if } s+1 \leq i \leq n,
\end{aligned}
$$

where $A=n^{t}$. Also if $b+t+1 / 2>0$, replacing $t$ by $t^{\prime}:=\min (t,-1 / 2)$ we have

$$
b+t^{\prime}+1 / 2=\min \left(b+t+1 / 2, b-\frac{1}{2}+\frac{1}{2}\right)=\min (b+t+1 / 2, b)>0 .
$$

Since the distribution of $\bar{\phi}$ is stochastically increasing in $A$, without loss of generality by replacing $t$ by $t^{\prime}$ if necessary we can assume $t \leq-\frac{1}{2}$, which gives $A \leq n^{-1 / 2}$. Using Taylor's series expansion twice, we have

$$
\begin{aligned}
& \log \cosh \left(\frac{\phi_{i}+\phi_{j}}{2}+\frac{\beta_{i}+\beta_{j}}{2}\right)-\log \cosh \left(\frac{\phi_{i}+\phi_{j}}{2}\right) \\
= & \frac{\beta_{i}+\beta_{j}}{2} \tanh \left(\frac{\phi_{i}+\phi_{j}}{2}\right)+O\left(\beta_{i}+\beta_{j}\right)^{2} \\
= & \frac{\beta_{i}+\beta_{j}}{2} \tanh (\bar{\phi})+O\left(\left(\beta_{i}+\beta_{j}\right)\left|\phi_{i}+\phi_{j}-2 \bar{\phi}\right|\right)+O\left(\beta_{i}+\beta_{j}\right)^{2} .
\end{aligned}
$$

Summing over $i<j$ and using (2.19) and (2.20) we get

$$
\begin{align*}
-\log f_{n, \theta, \boldsymbol{\beta}}(\phi) & =-\log f_{n, \theta, \mathbf{0}}(\phi)-\frac{(n-1) s A}{2} \tanh (\bar{\phi}) \\
& +O\left(n A \sum_{i=1}^{s}\left|\phi_{i}-\bar{\phi}\right|+s A \sum_{i=1}^{n}\left|\phi_{i}-\bar{\phi}\right|+n s A^{2}\right), \tag{2.24}
\end{align*}
$$

where

$$
\begin{align*}
-\log f_{n, \theta, \mathbf{0}}(\phi) & :=\sum_{i<j}\left[\frac{1}{8}\left(\phi_{i}-\phi_{j}\right)^{2}+q\left(\frac{\phi_{i}+\phi_{j}}{2}\right)\right] \\
& =\frac{n}{8} \sum_{i=1}^{n}\left(\phi_{i}-\bar{\phi}\right)^{2}+\sum_{i<j} q\left(\frac{\phi_{i}+\phi_{j}}{2}\right) \tag{2.25}
\end{align*}
$$

with $q($.$) as in (2.21). As the notation above suggests, f_{n, \theta, 0}$ defined above is the (unnormalized) density of $\phi$ under $\mathcal{H}_{0}$. Using (2.24), along with Lemma 2.3.2 we have

$$
-\log f_{n, \theta, \boldsymbol{\beta}}(\phi)=-\log f_{n, \theta, \mathbf{0}}(\phi)-\frac{n s A}{2} \tanh (\bar{\phi})-R_{n}
$$

where

$$
\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left|R_{n}\right| \lesssim \sqrt{n} s A+n s A^{2} \lesssim \sqrt{n} s A
$$

using $A \leq n^{-1 / 2}$. Thus, for any $K$ fixed and $K_{n}^{\prime}:=n^{3 / 4} s A$ we have

$$
\begin{aligned}
& \quad \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\tanh (\bar{\phi})<K n^{-1 / 4}\right) \\
& \leq \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\left|R_{n}\right|>K_{n}^{\prime}\right)+\mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\tanh (\bar{\phi})<K n^{-1 / 4},\left|R_{n}\right| \leq K_{n}^{\prime}\right) \\
& \leq \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\left|R_{n}\right|>K_{n}^{\prime}\right)+e^{K_{n}^{\prime}} \frac{\mathbb{E}_{\mathcal{H}_{0}} \exp \left[\frac{n s A}{2} \tanh (\bar{\phi})\right] 1\left\{\tanh (\bar{\phi})<K n^{-1 / 4}\right\}}{\mathbb{E}_{\mathcal{H}_{0}} \exp \left[\frac{n s A}{2} \tanh (\bar{\phi})\right] 1\left\{\left|R_{n}\right| \leq K_{n}^{\prime}\right\}} \\
& \leq \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\left|R_{n}\right|>K_{n}^{\prime}\right)+\frac{e^{K_{n}^{\prime}+\frac{K n^{3} / 4 s A}{2}-\frac{n s \tanh \left(2 K n^{-1 / 4}\right)}{2}}}{\mathbb{P}_{\mathcal{H}_{0}}\left(\bar{\phi}>2 K n^{-1 / 4},\left|R_{n}\right| \leq K_{n}^{\prime}\right)}
\end{aligned}
$$

On letting $n \rightarrow \infty$ and noting that $K_{n}^{\prime}=n^{3 / 4} s A \gg \sqrt{n} s A 2$ we have

$$
\lim _{n \rightarrow \infty} \sup _{\boldsymbol{\beta} \in \Xi(s, A)} \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\left|R_{n}\right| \leq K_{n}^{\prime}\right)=0, \text { and } \lim _{n \rightarrow \infty} \mathbb{P}_{\mathcal{H}_{0}}\left(\bar{\phi}>2 K n^{-1 / 4},\left|R_{n}\right| \leq K_{n}^{\prime}\right)=\mathbb{P}(\zeta>2 K)>0
$$

where $\zeta$ has density proportional to $e^{-\zeta^{4} / 12-\zeta^{2} / 24}$ (c.f. [41, Lemma 4.2]). Combining the last two displays we have

$$
\lim _{n \rightarrow \infty} \sup _{\boldsymbol{\beta} \in \Xi(s, A)} \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\tanh (\bar{\phi})<K n^{-1 / 4}\right)=0
$$

Since this holds for every fixed $K$, there exists $K_{n} \rightarrow \infty$ such that

$$
\limsup _{n \rightarrow \infty} \sup _{\boldsymbol{\beta} \in \Xi(s, A)} \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\tanh (\bar{\phi})<K_{n} n^{-1 / 4}\right)=0 .
$$

This verifies (2.22), and hence completes the proof of the theorem.
2.3.5 Proof of part (a) and (c) of Theorems 2.2.3

With $\Xi(s, A)$ as defined in (2.6), consider the following subset of $\Xi(s, A)$.

$$
\begin{equation*}
\tilde{\Xi}(s, A):=\left\{\boldsymbol{\beta}=\beta_{0} \mathbf{1}+\boldsymbol{\mu}:|\operatorname{supp}(\boldsymbol{\mu})|=s, \text { and } \mu_{i}=A, i \in \operatorname{supp}(\boldsymbol{\mu})\right\} . \tag{2.26}
\end{equation*}
$$

Let $\pi(d \boldsymbol{\beta})$ be a prior on $\Xi(s, A)$, which put probability mass $1 /\binom{n}{s}$ on each of configurations in $\tilde{\Xi}(s, A)$. And let $\mathbb{Q}_{\pi}():.=\int \mathbb{P}_{n, \theta, \boldsymbol{\beta}}(.) \pi(d \boldsymbol{\beta})$ denote the marginal distribution of $Y$ under this prior. To show that all tests for the problem (2.5) are asymptotically powerless, using the second moment method it suffices to show that

$$
\begin{equation*}
\lim _{n \rightarrow \infty} \mathbb{E}_{\mathcal{H}_{0}} L_{\pi}(Y)^{2}=1, \text { where } L_{\pi}(Y):=\frac{\mathbb{Q}_{\pi}(Y)}{\mathbb{P}_{\mathcal{H}_{0}}(Y)} \tag{2.27}
\end{equation*}
$$

is the likelihood ratio. The following lemma gives an upper bound to the second moment of $L_{\pi}($.$) .$

Lemma 2.3.3. For any $\left(\theta, \beta_{0}\right)$, with $L_{\pi}($.$) as defined in (2.27) we have$

$$
\begin{equation*}
\mathbb{E}_{\mathcal{H}_{0}} L_{\pi}^{2}(Y) \leq \exp \left\{A^{2} s^{2} \operatorname{Cov}_{\boldsymbol{\beta}=\left(\beta_{0} / 2\right) \mathbf{1}}\left(k_{1}, k_{2}\right)+\frac{2 s^{2}}{n}\left(e^{A^{2} \operatorname{Var}_{\boldsymbol{\beta}=\left(\beta_{0} / 2\right) \mathbf{1}}\left(k_{1}\right)}-1\right)\right\}, \tag{2.28}
\end{equation*}
$$

whenever $n>2 s$.

Proof. Define $\Lambda_{s}:=\left\{S|S \subset\{1,2, \ldots, n\},|S|=s\}\right.$. For any $S \in \Lambda_{s}$, define a vector $\boldsymbol{\beta}_{S}$ by setting

$$
\begin{aligned}
\beta_{S, i} & =\beta_{0}+A \text { if } i \in S, \\
& =\beta_{0} \text { if } i \notin S .
\end{aligned}
$$

By symmetry, the normalizing constant $Z_{n}\left(\boldsymbol{\beta}_{S}, \theta\right)$ is the same for all $S \in \Lambda_{s}$, which we denote by
$Z_{n}\left(\boldsymbol{\beta}_{[s]}, \theta\right)$ for the rest of this proof. Then, a direct calculation gives

$$
\begin{align*}
\mathbb{E}_{\mathcal{H}_{0}} L_{\pi}^{2}(Y) & =\frac{Z_{n}^{2}\left(\beta_{0}, \theta\right)}{Z_{n}^{2}\left(\boldsymbol{\beta}_{[s]}, \theta\right)} \frac{1}{\binom{n}{s}^{2}} \mathbb{E}_{\mathcal{H}_{0}} \sum_{S_{1}, S_{2} \in \Lambda_{s}} e^{\sum_{j \in S_{1}} \frac{A}{2} k_{j}+\sum_{j \in S_{2}} \frac{A}{2} k_{j}} \\
& =\frac{Z_{n}\left(\beta_{0}, \theta\right)}{Z_{n}^{2}\left(\boldsymbol{\beta}_{[s]}, \theta\right)} \frac{1}{\binom{n}{s}} \sum_{S_{1}, S_{2} \in \Lambda} \frac{Z_{n}\left(\boldsymbol{\beta}_{S_{1}}+\boldsymbol{\beta}_{S_{2}}, \theta\right)}{Z_{n}\left(\boldsymbol{\beta}_{S_{1}}+\boldsymbol{\beta}_{S_{2}}, \theta\right)} \\
& \sum_{Y} e^{\frac{\theta}{2 n} \sum_{i=1}^{n} k_{i}^{2}+\sum_{j=1}^{n} \frac{\beta_{S_{1}, j}+\beta \beta_{S_{2}, j}}{2} k_{j}}  \tag{2.29}\\
& =\frac{1}{\binom{n}{s}^{2}} \sum_{S_{1}, S_{2} \in \Lambda} \frac{Z_{n}\left(\beta_{0}, \theta\right) Z_{n}\left(\boldsymbol{\beta}_{S_{1}}+\boldsymbol{\beta}_{S_{2}}, \theta\right)}{Z_{n}\left(\boldsymbol{\beta}_{S_{1}}, \theta\right) Z_{n}\left(\boldsymbol{\beta}_{S_{2}}, \theta\right)}=\frac{1}{\binom{n}{s}^{2}} \sum_{S_{1}, S_{2} \in \Lambda} R_{S_{1}, S_{2}},
\end{align*}
$$

where

$$
\begin{aligned}
R_{S_{1}, S_{2}} & :=\log \left(\frac{Z_{n}\left(\beta_{0}, \theta\right) Z_{n}\left(\boldsymbol{\beta}_{S_{1}}+\boldsymbol{\beta}_{S_{2}}, \theta\right)}{Z_{n}\left(\boldsymbol{\beta}_{S_{1}}, \theta\right) Z_{n}\left(\boldsymbol{\beta}_{S_{2}}, \theta\right)}\right) \\
& =\log Z_{n}\left(\boldsymbol{\beta}_{S_{1}}+\boldsymbol{\beta}_{S_{2}}, \theta\right)-\log Z_{n}\left(\boldsymbol{\beta}_{S_{2}}, \theta\right)-\log Z_{n}\left(\boldsymbol{\beta}_{S_{1}}, \theta\right)+\log Z_{n}\left(\beta_{0}, \theta\right)
\end{aligned}
$$

Setting $W=S_{1} \cap S_{2}$, note that $R_{S_{1}, S_{2}}$ only depends on $|W|$ by symmetry. Thus, without loss of generality we assume that $S_{1}=\{1,2,3, \ldots, s\}$ and $S_{2}=\{1,2, \ldots, w, s+1, s+2, \ldots, 2 s-w\}$. Consequently we have
$R_{S_{1}, S_{2}}=\sum_{j \in S_{1}}\left[\log Z_{n}\left(\boldsymbol{\beta}_{[j]}+\boldsymbol{\beta}_{S_{2}}, \theta\right)-\log Z_{n}\left(\boldsymbol{\beta}_{[j-1]}+\boldsymbol{\beta}_{S_{2}}, \theta\right)-\log Z_{n}\left(\boldsymbol{\beta}_{[j]}, \theta\right)+\log Z_{n}\left(\boldsymbol{\beta}_{[j-1]}, \theta\right)\right]$,
where $\boldsymbol{\beta}_{[j]}$ denotes the vector $\boldsymbol{\beta}$ which equals $A$ on first $j$ entries, and $\beta_{0}$ for rest of its entries, The summand in the RHS above equals

$$
\begin{aligned}
& \log Z_{n}\left(\boldsymbol{\beta}_{[j]}+\boldsymbol{\beta}_{S_{2}}, \theta\right)-\log Z_{n}\left(\boldsymbol{\beta}_{[j-1]}+\boldsymbol{\beta}_{S_{2}}, \theta\right)-\log Z_{n}\left(\boldsymbol{\beta}_{[j]}, \theta\right)+\log Z_{n}\left(\boldsymbol{\beta}_{[j-1]}, \theta\right) \\
& =\int_{0}^{A} \frac{\partial \log Z_{n}\left(\boldsymbol{\beta}_{[j-1]}+\boldsymbol{\beta}_{S_{2}}+\gamma \mathbf{e}_{\mathbf{j}}, \theta\right)}{\partial \beta_{j}} d \gamma-\int_{0}^{A} \frac{\partial \log Z_{n}\left(\boldsymbol{\beta}_{[j-1]}+\gamma \mathbf{e}_{\mathbf{j}}, \theta\right)}{\partial \beta_{j}} d \gamma \\
& =\left.\int_{0}^{A} A \sum_{r \in S_{2}} \frac{\partial \log Z_{n}\left(\boldsymbol{\beta}_{[j-1]}+\boldsymbol{\xi}+\gamma \mathbf{e}_{\mathbf{j}}\right)}{\partial \beta_{j} \partial \beta_{r}}\right|_{\xi \leq \boldsymbol{\beta}_{S_{2}}} d \gamma \\
& =\int_{0}^{A} A \sum_{r \in S_{2}} \operatorname{Cov}_{\boldsymbol{\beta}=\boldsymbol{\beta}_{[j-1]}+\boldsymbol{\xi}+\gamma \mathbf{e}_{\mathbf{j}}}\left(k_{j}, k_{r}\right) d \gamma
\end{aligned}
$$

If $A \rightarrow 0$, then $\boldsymbol{\beta} \geq \mathbf{0}$ if $\beta_{0} \geq 0$, and $\boldsymbol{\beta} \leq \mathbf{0}$ for all $n$ large if $\beta_{0}<0$. Note that the GHS inequality [37] holds if either $\boldsymbol{\beta} \geq \mathbf{0}$ or $\boldsymbol{\beta} \leq \mathbf{0}$ (the second conclusion follows on noting that $\operatorname{Cov}_{\boldsymbol{\beta}}\left(k_{r}, k_{s}\right)=\operatorname{Cov}_{\boldsymbol{\beta}}\left(-k_{r},-k_{s}\right)$, thereby giving

$$
\operatorname{Cov}_{\boldsymbol{\beta}=\boldsymbol{\beta}_{[j-1]}+\xi_{+}+\mathbf{e}_{\mathbf{j}}}\left(k_{j}, k_{r}\right) \leq \operatorname{Cov}_{\boldsymbol{\beta}=\left(\beta_{0} / 2\right) \mathbf{1}}\left(k_{j}, k_{r}\right) .
$$

Combining the above two displays, this gives

$$
\begin{aligned}
R_{S_{1}, S_{2}} & \leq \sum_{j \in S_{1}} \int_{0}^{A} A \sum_{r \in S_{2}} \operatorname{Cov}_{\boldsymbol{\beta}=\left(\beta_{0} / 2\right) \mathbf{1}}\left(k_{j}, k_{r}\right) d \gamma \\
& =A^{2} w \operatorname{Var}_{\boldsymbol{\beta}=\left(\beta_{0} / 2\right) \mathbf{1}}\left(k_{1}\right)+A^{2}\left(s^{2}-w\right) \operatorname{Cov}_{\boldsymbol{\beta}=\left(\beta_{0} / 2\right) \mathbf{1}}\left(k_{1}, k_{2}\right)
\end{aligned}
$$

Along with (2.29), this further gives

$$
\mathbb{E}_{\mathcal{H}_{0}} L_{\pi}^{2}(Y) \leq \exp \left\{A^{2} s^{2} \operatorname{Cov}_{\boldsymbol{\beta}=\left(\beta_{0} / 2\right) \mathbf{1}}\left(k_{1}, k_{2}\right)\right\} \mathbb{E}_{W} \exp \left\{A^{2} \operatorname{Var}_{\boldsymbol{\beta}=\left(\beta_{0} / 2\right) \mathbf{1}}\left(k_{1}\right) W\right\}
$$

where $W$ follows Hypergeometric distribution with parameters $(n, s, s)$. Since $2 s<n$, $W$ is stochastically dominated by a binomial distribution with parameters $\left(s, \frac{s}{n-s}\right)$ ([39, Lemma 6.1]), which gives

$$
\mathbb{E}_{W} \exp \left\{A^{2} \operatorname{Var}_{\boldsymbol{\beta}=\left(\beta_{0} / 2\right) \mathbf{1}}\left(k_{1}\right) W\right\} \leq \exp \left\{\frac{2 s^{2}}{n}\left(e^{A^{2} \operatorname{Var}_{\boldsymbol{\beta}=\left(\beta_{0} / 2\right) 1}\left(k_{1}\right)}-1\right)\right\}
$$

Combining the last two displays, we have verified (2.28).

With $L_{\pi}$ as in defined in (2.27), it suffices to show that

$$
\lim _{n \rightarrow \infty} \mathbb{E}_{\mathcal{H}_{0}} L_{\pi}^{2}(Y)=1
$$

By [41, Lemma 4.4] we have

$$
\begin{equation*}
\operatorname{Var}_{\boldsymbol{\beta}=\left(\beta_{0} / 2\right) \mathbf{1}}\left(\sum_{e \in \mathcal{E}} Y_{e}\right) \lesssim n^{2}, \tag{2.30}
\end{equation*}
$$

which gives the existence of a constant $c$ depending on $\theta$ such that

$$
\begin{equation*}
\operatorname{Var}_{\boldsymbol{\beta}=\left(\beta_{0} / 2\right) \mathbf{1}}\left(k_{1}\right) \leq c n, \quad \operatorname{Cov}_{\boldsymbol{\beta}=\left(\beta_{0} / 2\right) \mathbf{1}}\left(k_{1}, k_{2}\right) \leq c \tag{2.31}
\end{equation*}
$$

Using this along with Lemma 2.3.3 gives

$$
\begin{equation*}
\mathbb{E}_{\mathcal{H}_{0}} L_{\pi}^{2}(Y) \leq \exp \left\{c A^{2} s^{2}+\frac{2 s^{2}}{n}\left(e^{c A^{2} n}-1\right)\right\} \tag{2.32}
\end{equation*}
$$

For part (a), in this regime we have $s=n^{b}$ and $A=n^{t}$ with $b \geq \frac{1}{2}$ and $b+t<0$. This gives

$$
\max \left(A^{2} n, A^{2} s^{2}\right)=\max \left(n^{2 t+1}, n^{2 t+2 b}\right)=n^{2 t+2 b} \rightarrow 0,
$$

using which the exponent in the RHS of (2.32) converges to 0 . This completes the proof of part (a).

For part (c), in this regime we have $s=n^{b}$ and $A=n^{t}$ with $b<\frac{1}{2}$ and $t \leq-\frac{1}{2}$. This gives $A^{2} s^{2}=n^{2 b+2 t} \rightarrow 0$. Also

$$
\frac{s^{2}}{n} e^{c A^{2} n-1} \leq e^{c-1} \frac{s^{2}}{n}=e^{c-1} n^{2 b-1} \rightarrow 0 .
$$

Consequently, the RHS of (2.32) again converges to 0 . This completes the proof of part (c).

### 2.3.6 Proof of Theorem 2.2.5 Part (a)

As before, with $L_{\pi}$ defined in (2.27), it is sufficient to show that

$$
\lim _{n \rightarrow \infty} \mathbb{E}_{\mathcal{H}_{0}} L_{\pi}^{2}(Y)=1
$$

To this effect, using Theorem 2.4 \& Lemma 4.8 in [41] we get

$$
\operatorname{Var}_{\mathcal{H}_{0}}\left(\sum_{e \in \mathcal{E}} Y_{e}\right) \lesssim n^{3} .
$$

Along with the non-negativity of covariance, this gives

$$
\operatorname{Cov}_{\mathcal{H}_{0}}\left(k_{1}, k_{2}\right)=O(n) .
$$

For getting the optimal bound on $\operatorname{Var}_{\mathcal{H}_{0}}\left(k_{1}\right)$, use (2.18) to get

$$
\operatorname{Var}_{\mathcal{H}_{0}}\left(k_{1}\right) \lesssim n^{2} \operatorname{Var}_{\mathcal{H}_{0}}\left(\phi_{1}\right)+n \lesssim n^{2}\left[\operatorname{Var}_{\mathcal{H}_{0}}(\bar{\phi})+\operatorname{Var}_{\mathcal{H}_{0}}\left(\phi_{1}-\bar{\phi}\right)\right]+n \lesssim n,
$$

where the last inequality uses [41, Lemma 4.1]. Combing the above two displays along with Lemma 2.3.3 gives the existence of a constant $c$ free of $n$, such that

$$
\begin{equation*}
\mathbb{E}_{\mathcal{H}_{0}} L_{\pi}^{2}(Y) \leq \exp \left\{c A^{2} s^{2} n+\frac{2 s^{2}}{n}\left(e^{c A^{2} n}-1\right)\right\} \tag{2.33}
\end{equation*}
$$

Now, recall that in this regime we have $s=n^{b}$ and $A=n^{t}$ with $b+t+\frac{1}{2}<0$. This gives $A^{2} s^{2} n=n^{2 b+2 t+1} \rightarrow 0$. Also, noting that $2 t+1<0$ we have

$$
\frac{s^{2}}{n}\left(e^{c A^{2} n}-1\right) \leq n^{2 b-1}\left(e^{c n^{2 t+1}}-1\right) \lesssim n^{2 b+2 t+1} \rightarrow 0 .
$$

Along with (2.33), this gives $\lim _{n \rightarrow \infty} \mathbb{E}_{\mathcal{H}_{0}} L_{\pi}^{2}(Y)=1$. This completes the proof of part (b).

### 2.3.7 Proof of part (a) and (c) of Theorem 2.2.4

We first state the following lemma about the function $q($.$) introduced in (2.21), the proof of$ which follows from straightforward calculus (see for e.g. [21]).

Lemma 2.3.4. If $\theta>1 / 2$, the equation $q^{\prime}(x)=2 \theta[x-\theta \tanh (2 \theta x)]$ has a unique positive root $t$, say, on $(0, \infty)$. Further, $t$ is the unique global minimizer of $q($.$) on [0, \infty)$.

We will use the notation $t$ introduced in the above lemma throughout the rest of the chapter.
Set

$$
\begin{equation*}
U:=\cap_{i=1}^{n} V_{i}, \quad V_{i}:=\left\{Y: k_{i} \geq(n-1) t / 2\right\} . \tag{2.34}
\end{equation*}
$$

Restricting the probability measure (2.13) to the set $U$, define the probability measure $\mathbb{P}_{n, \boldsymbol{\beta}, U}(\cdot)$ by setting

$$
\begin{equation*}
\mathbb{P}_{n, \boldsymbol{\beta}, U}(Y)=\frac{1}{Z_{n}^{+}(\boldsymbol{\beta}, \theta)} \exp \left\{\frac{\theta}{2 n} \sum_{i=1}^{n} k_{i}^{2}+\frac{1}{2} \sum_{i=1}^{n} \beta_{i} k_{i}\right\} 1\{Y \in U\} . \tag{2.35}
\end{equation*}
$$

where

$$
Z_{n}^{+}(\boldsymbol{\beta}, \theta)=\sum_{Y \in U} \exp \left\{\frac{\theta}{2 n} \sum_{i=1}^{n} k_{i}^{2}+\frac{1}{2} \sum_{i=1}^{n} \beta_{i} k_{i}\right\}
$$

is the restricted normalizing constant. As before, consider the sub parameter space $\tilde{\Xi}(s, A)$ defined in (2.26), let $\pi(d \boldsymbol{\beta})$ be a prior on $\tilde{\Xi}(s, A)$, which put probability mass $1 /\binom{n}{s}$ on each of configurations in $\tilde{\Xi}(s, A)$. And let $\mathbb{Q}_{\pi, U}():.=\int \mathbb{P}_{n, \boldsymbol{\beta}, U}(.) \pi(d \boldsymbol{\beta})$ denote the mixed alternative distribution of $Y$. Since [41, Lem 4.3] gives $\mathbb{P}_{\mathcal{H}_{0}}(U) \rightarrow 1 / 2$, to verify the absence of asymptotically powerful tests setting

$$
\begin{equation*}
L_{\pi, U}(Y):=\frac{\mathbb{Q}_{\pi, U}(Y)}{\mathbb{P}_{\mathcal{H}_{0}, U}(Y)}, \tag{2.36}
\end{equation*}
$$

it suffices to show:

$$
\begin{equation*}
\mathbb{E}_{\mathcal{H}_{0}, U} L_{\pi, U}^{2}(Y) \rightarrow 1 \tag{2.37}
\end{equation*}
$$

Proceeding similar to Lemma 2.3.3, we get

$$
\begin{equation*}
\mathbb{E}_{\mathcal{H}_{0}, U} L_{\pi, U}^{2}(Y)=\frac{1}{\binom{n}{s}^{2}} \sum_{S_{1}, S_{2} \in \Lambda} \frac{Z_{n}^{+}(0, \theta) Z_{n}^{+}\left(\boldsymbol{\beta}_{S_{1}}+\boldsymbol{\beta}_{S_{2}}, \theta\right)}{Z_{n}^{+}\left(\boldsymbol{\beta}_{S_{1}}, \theta\right) Z_{n}^{+}\left(\boldsymbol{\beta}_{S_{2}}, \theta\right)} \tag{2.38}
\end{equation*}
$$

Setting $R_{S_{1}, S_{2}}^{+}$as

$$
R_{S_{1}, S_{2}}^{+}:=\left(\log Z_{n}^{+}\left(\boldsymbol{\beta}_{S_{1}}+\boldsymbol{\beta}_{S_{2}}, \theta \log Z_{n}^{+}\left(\boldsymbol{\beta}_{S_{2}}, \theta\right)\right)-\left(\log Z_{n}^{+}\left(\boldsymbol{\beta}_{S_{1}}, \theta\right)-\log Z_{n}^{+}(0, \theta)\right)\right.
$$

A Taylor's series expansion gives

$$
\begin{equation*}
R_{S_{1}, S_{2}}=A^{2} \sum_{i \in S_{1}} \sum_{j \in S_{2}} \operatorname{Cov}_{\delta=\alpha \mathbf{1}_{S_{1}}+\gamma \mathbf{1}_{S_{2}}}\left(k_{i}, k_{j} \mid U\right) \tag{2.39}
\end{equation*}
$$

where $\alpha, \gamma \in(0, A)$ and $\mathbf{1}_{S}$ denote vector having unit signals at $S$, and $\delta:=\alpha \mathbf{1}_{S_{1}}+\gamma \mathbf{1}_{S_{2}} \in$ $\left[0,2 n^{-1 / 2}\right]^{n}$. We now claim that

Lemma 2.3.5.

$$
\max _{1 \leq i<j \leq n} \sup _{\boldsymbol{\beta} \in\left[0,2 n^{-1 / 2}\right]^{n}} \operatorname{Cov}_{\boldsymbol{\beta}}\left(k_{i}, k_{j} \mid U\right) \lesssim 1 .
$$

We defer the proof of Lemma 2.3.5 to the end of the section. Finally, use Lemma 2.3.6 to conclude that

$$
\begin{equation*}
\max _{1 \leq i \leq n} \operatorname{Var}_{\delta}\left(k_{i} \mid U\right) \lesssim n \tag{2.40}
\end{equation*}
$$

Given Lemma 2.3.5 along with (2.40) and (2.39), we have the existence of a constant $C$ free of $n$ such that

$$
R_{S_{1}, S_{2}} \leq C W A^{2} n+C s^{2} A^{2}
$$

which along with (2.38) gives

$$
\begin{equation*}
\mathbb{E}_{\mathcal{H}_{0}, U} L_{\pi, U}^{2}(Y) \leq \exp \left\{C A^{2} s^{2}\right\} \mathbb{E}_{W} \exp \left\{C A^{2} n W\right\} \tag{2.41}
\end{equation*}
$$

where $W$ follows Hypergeometric distribution with parameters ( $n, s, s$ ). As before, using the fact that $n>2 s, W$ is stochastically dominated by a binomial distribution with parameters $\left(s, \frac{s}{n-s}\right)$.

This gives

$$
\begin{equation*}
\mathbb{E}_{\mathcal{H}_{0}, U} L_{\pi, U}^{2}(Y) \leq \exp \left\{C A^{2} s^{2}+\frac{2 s^{2}}{n}\left(e^{C A^{2} n}-1\right)\right\} \tag{2.42}
\end{equation*}
$$

For part (a), in this regime we have $s=n^{b}$ and $A=n^{t}$ with $b \geq \frac{1}{2}$ and $b+t<0$. This gives $A^{2} s^{2}=n^{2 t+2 b} \rightarrow 0$. Also we have $A^{2} n=n^{2 t+1} \rightarrow 0$, and so

$$
\frac{s^{2}}{n}\left(e^{C A^{2} n}-1\right) \lesssim s^{2} A^{2}=n^{2 b+2 t} \rightarrow 0
$$

Combining the above two displays with (2.42), we have $\mathbb{E}_{\mathcal{H}_{0}, U} L_{\pi, U}^{2}(Y) \rightarrow 1$, as desired. This completes the proof of part (a).

For part (c), in this regime we have $s=n^{b}$ and $A=n^{t}$ with $b<\frac{1}{2}$ and $t+\frac{1}{2}<0$. This gives

$$
A^{2} s^{2}=n^{2 t+2 b} \leq n^{2 t+1} \rightarrow 0 .
$$

Also we have $A^{2} n=n^{2 t+1} \rightarrow 0$, and so

$$
\frac{s^{2}}{n}\left(e^{C A^{2} n}-1\right) \lesssim s^{2} A^{2}=n^{2 b+2 t} \rightarrow 0 .
$$

Combining the above two displays with (2.42), we have $\mathbb{E}_{\mathcal{H}_{0}, U} L_{\pi, U}^{2}(Y) \rightarrow 1$, as desired. This completes the proof of part (c).

### 2.3.8 Proof of Lemma 2.3.5

We first state two lemmas, which will be used in the proof of Lemma 2.3.5. The first lemma is the analogue of Lemma 2.3.2 for $\theta>1 / 2$.

Lemma 2.3.6. Suppose $\theta>1 / 2$, and $\boldsymbol{\beta} \in\left[0,2 n^{-1 / 2}\right]$. Then for every positive positive integer $\ell$
we have

$$
\begin{equation*}
\mathbb{E}_{n, \theta, \beta}\left(\left|\phi_{i}-t\right|^{\ell} \mid U\right) \leq C n^{-\ell / 2}, \tag{2.43}
\end{equation*}
$$

where $U$ is as defined in (2.34), and $C$ is a positive constant depending only on $\ell$ and $\theta$.

For stating the second lemma, we require the following definition. Analogous to (2.34), define

$$
\begin{equation*}
\widetilde{U}:=\cap_{i=1}^{n} \widetilde{V}_{i} \quad \widetilde{V}_{i}:=\left\{\phi_{i} \in[0,2]\right\} . \tag{2.44}
\end{equation*}
$$

The next lemma shows that the sets $U$ and $\widetilde{U}$ occur simultaneously with high probability, and so expectations involving $U$ can be transferred to expectations involving $\widetilde{U}$ at a very low cost. This lemma will be used frequently in the rest of this section, sometimes without an explicit mention.

Lemma 2.3.7. Suppose $\theta>1 / 2$, and $\boldsymbol{\beta} \in\left[0,2 n^{-1 / 2}\right]$. Then we have the following conclusions:
(a) $\log \mathbb{P}_{n, \theta, \boldsymbol{\beta}}(U \Delta \widetilde{U}) \lesssim-n$.
(b) For any random variable $W$ such that $\mathbb{E} W^{2} \leq 1$, we have

$$
|\mathbb{E} W 1\{U\}-\mathbb{E} W 1\{\widetilde{U}\}| \leq \sqrt{\mathbb{P}_{n, \theta, \boldsymbol{\beta}}(U \Delta \widetilde{U})} .
$$

The proofs of Lemmas 2.3.6 and 2.3.7 are deferred to section 3.7. We now prove a correlation bound for higher order terms, which will be used for proved Lemma 2.3.5.

Lemma 2.3.8. Suppose $\theta>1 / 2$, and $\boldsymbol{\beta} \in\left[0,2 n^{-1 / 2}\right]$. Then for any pair of indices $\left\{i_{1}, i_{2}, i_{3}\right\}$ (not necessarily distinct), we have

$$
\operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(\left(\phi_{i_{1}}-t\right)\left(\phi_{i_{2}}-t\right), \phi_{i_{3}}-t \mid \widetilde{U}\right) \lesssim n^{-2} .
$$

Proof. Setting $M\left(i_{1}, i_{2}, i_{3}\right):=\operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(\left(\phi_{i_{1}}-t\right)\left(\phi_{i_{2}}-t\right), \phi_{i_{3}}-t \mid \widetilde{U}\right)$, we claim that

$$
\begin{equation*}
\max _{1 \leq i_{1}, i_{2} \leq n}\left|M\left(i_{1}, i_{2}, i_{3}\right)-\frac{\theta^{3} \operatorname{sech}^{6}(2 \theta t)}{(n-1)^{3}} \sum_{j_{1} \neq i_{1}, j_{2} \neq i_{2}, j_{3} \neq i_{3}} \sum_{u \in\left(i_{1}, j_{1}\right), v \in\left(i_{2}, j_{2}\right), w \in\left(i_{3}, j_{3}\right)} M(u, v, w)\right|=O\left(n^{-2}\right) . \tag{2.45}
\end{equation*}
$$

We first complete the proof of the lemma, deferring the proof of (2.45). The above display implies the existence of a constant $C$ free of $n$, such that

$$
\begin{equation*}
\max _{1 \leq i_{1}, i_{2}, i_{3} \leq n}\left|M\left(i_{1}, i_{2}, i_{3}\right)-\sum_{1 \leq j_{1}, j_{2}, j_{3} \leq n} B_{n}\left(\left(i_{1}, i_{2}, i_{3}\right),\left(j_{1}, j_{2}, j_{3}\right)\right)\right| \leq \frac{C}{n^{2}} \tag{2.46}
\end{equation*}
$$

where $B_{n}$ is a symmetric $n^{3} \times n^{3}$ matrix with non-negative entries, satisfying

$$
\sum_{1 \leq j_{1}, j_{2}, j_{3} \leq n} B_{n}\left(\left(i_{1}, i_{2}, i_{3}\right),\left(j_{1}, j_{2}, j_{3}\right)\right)=8 \theta^{3} \operatorname{sech}^{6}(2 \theta t)<1 .
$$

Thus the matrix $\left(\mathbf{I}-B_{n}\right)^{-1}$ has $\ell_{\infty}$ operator norm equal to $\left(1-8 \theta^{3} \operatorname{sech}^{6}(2 \theta t)\right)^{-1}<\infty$, and so (2.46) gives

$$
\max _{1 \leq i_{1}, i_{2}, i_{3} \leq n}\left|M\left(i_{1}, i_{2}, i_{3}\right)\right| \leq C\left(1-8 \theta^{3} \operatorname{sech}^{6}(2 \theta t)\right)^{-1} n^{-2}
$$

from which the desired conclusion follows.

It thus remains to verify (2.45). There are various possibilities depending on which of the indices $\{i, j, \ell\}$ are distinct. Below we argue the case $i_{1}=i_{2}=i$ and $i_{3}=j$, with $\{i, j\}$ distinct, noting that the bound follows by similar calculations for other choices. To this end, setting $k_{i, t}:=$
$k_{i}-(n-1) t$ note that $\left(\phi_{i}-t \mid Y\right) \sim N\left(\frac{k_{i, t}}{n-1}, \frac{1}{(n-1) \theta}\right)$. Consequently, we have

$$
\begin{align*}
& \mathbb{P}_{n, \theta, \boldsymbol{\beta}}(\widetilde{U}) \operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(\left(\phi_{i}-t\right)^{2}, \phi_{j}-t \mid \widetilde{U}\right) \\
= & \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[\left(\phi_{i}-t\right)^{2}\left(\phi_{j}-t\right) 1\{\widetilde{U}\}\right]-\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[\left(\phi_{i}-t\right)^{2} 1\{\widetilde{U}\}\right] \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[\left(\phi_{j}-t\right) 1\{\widetilde{U}\}\right] \\
= & \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[\left(\phi_{i}-t\right)^{2}\left(\phi_{j}-t\right) 1\{U\}\right]-\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[\left(\phi_{i}-t\right)^{2} 1\{U\}\right] \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[\left(\phi_{j}-t\right) 1\{U\}\right]+O\left(e^{-c n}\right) \\
= & \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[\left(\frac{k_{i, t}^{2}}{(n-1)^{2}}+\frac{1}{(n-1) \theta}\right) \frac{k_{i, t}}{n-1} 1\{U\}\right] \\
- & \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[\left(\frac{k_{i, t}^{2}}{(n-1)^{2}}+\frac{1}{(n-1) \theta}\right) 1\{U\}\right] \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[\frac{k_{i, t}}{n-1} 1\{U\}\right]+O\left(e^{-c n}\right) \\
= & \mathbb{P}_{n, \theta, \boldsymbol{\beta}}(U) \operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(\frac{k_{i, t}^{2}}{(n-1)^{2}}+\frac{1}{(n-1) \theta}, \left.\frac{k_{j, t}}{n-1} \right\rvert\, U\right)+O\left(e^{-c n}\right) \\
= & \frac{1}{(n-1)^{3}} \mathbb{P}_{n, \theta, \boldsymbol{\beta}}(U) \operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(k_{i, t}^{2}, k_{j, t} \mid U\right)+O\left(e^{-c n}\right) \\
= & \frac{1}{(n-1)^{3}} \mathbb{P}_{n, \theta, \boldsymbol{\beta}}(U) \sum_{a_{1}, a_{2} \neq i, b \neq j} \operatorname{Cov}\left(Y_{i a_{1}, t} Y_{i a_{2}, t}, Y_{j b, t} \mid U\right)+O\left(e^{-c n}\right), \tag{2.47}
\end{align*}
$$

where $Y_{i j, t}:=Y_{i j}-t$, and the change from $U$ to $\widetilde{U}$ uses Lemma 2.3.7 and incurs the cost $O\left(e^{-c n}\right)$. Proceeding to estimate the RHS of (2.47), set $r_{i j, t}:=\mathbb{E}\left(Y_{i j} \mid \phi\right)$, and for $a_{1} \neq a_{2}$ note that

$$
\begin{align*}
& \mathbb{P}_{n, \theta, \boldsymbol{\beta}}(U) \operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(Y_{i a_{1}, t} Y_{i a_{2}, t}, Y_{j b, t} \mid U\right) \\
= & \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left(Y_{i a_{1}, t} Y_{i a_{2}, t} Y_{j b, t} 1\{U\}\right)-\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left(Y_{i a_{1}, t} Y_{i a_{2}, t} 1\{U\}\right) \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left(Y_{j b} 1\{U\}\right) \\
= & \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[Y_{i a_{1}, t} Y_{i a_{2}, t} Y_{j b, t} 1\{\widetilde{U}\}\right]-\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[Y_{i a_{1}, t} Y_{i a_{2}, t} 1\{\widetilde{U}\}\right] \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[Y_{j b, t} 1\{\widetilde{U}\}\right]+O\left(e^{-c n}\right) \\
= & \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[r_{i a_{1}, t} r_{i a_{2}, t} r_{j b, t} 1\{\widetilde{U}\}\right]-\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[r_{i a_{1}, t} r_{i a_{2}, t}\{\widetilde{U}\}\right] \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[r_{j b} 1\{\widetilde{U}\}\right]+O\left(e^{-c n}\right) \\
= & \mathbb{P}_{n, \theta, \boldsymbol{\beta}}(\widetilde{U}) \operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(r_{i a_{1}, t} r_{i a_{2}, t}, r_{j b, t} \mid \widetilde{U}\right)+O\left(e^{-c n}\right) . \tag{2.48}
\end{align*}
$$

In the above display, we have again moved from $U$ to $\widetilde{U}$ at a cost $O\left(e^{-c n}\right)$, using Lemma 2.3.7. A
one term Taylor's series expansion gives

$$
\begin{align*}
r_{i j, t} & =\tanh \left[\theta\left(\phi_{i}+\phi_{i}\right)+\left(\beta_{i}+\beta_{j}\right)\right]-\tanh (2 \theta t) \\
& =\left[\theta\left(\phi_{i}-t+\phi_{j}-t\right)+\frac{1}{2}\left(\beta_{i}+\beta_{j}\right)\right] \operatorname{sech}^{2}(2 \theta t)+\xi_{i j} \tag{2.49}
\end{align*}
$$

where

$$
\left|\xi_{i j}\right| \lesssim\left(\phi_{i}-t\right)^{2}+\left(\phi_{j}-t\right)^{2}+\beta_{i}^{2}+\beta_{j}^{2} \lesssim\left(\phi_{i}-t\right)^{2}+\left(\phi_{j}-t\right)^{2}+n^{-1} .
$$

On taking expectations, $\operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(r_{i a_{1}, t} r_{i a_{2}, t}, r_{j b, t} \mid \widetilde{U}\right)$ equals

$$
\begin{equation*}
\theta^{3} \operatorname{sech}^{6}(2 \theta t) \sum_{u \in\left\{i, a_{1}\right\}, v \in\left\{i, a_{2}\right\}, w \in\{j, b\}} \operatorname{Cov}_{n, \theta, \beta}\left(\left(\phi_{u}-t\right)\left(\phi_{v}-t\right),\left(\phi_{w}-t\right) \mid \widetilde{U}\right)+O\left(n^{-2}\right), \tag{2.50}
\end{equation*}
$$

where we have used Lemma 2.3.6. On the other hand, if $a_{1}=a_{2}=a$, then using the fact that $Y_{i a, t}^{2}=1+t^{2}-2 t Y_{i a}=1-t^{2}-2 t Y_{i a, t}$ we have

$$
\begin{align*}
& -\frac{1}{2 t} \mathbb{P}_{n, \theta, \boldsymbol{\beta}}(U) \operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(Y_{i a, t}^{2}, Y_{j b, t} \mid U\right) \\
= & \mathbb{P}_{n, \theta, \boldsymbol{\beta}}(U) \operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(Y_{i a, t}, Y_{j b, t} \mid U\right) \\
= & \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left(Y_{i a, t} Y_{j b, t} 1\{U\}\right)-\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left(Y_{i a, t} 1\{U\}\right) \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left(Y_{j b} 1\{U\}\right) \\
= & \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[Y_{i a, t} Y_{j b, t} 1\{\widetilde{U}\}\right]-\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[Y_{i a, t} 1\{\widetilde{U}\}\right] \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[Y_{j b, t} 1\{\widetilde{U}\}\right]+O\left(e^{-c n}\right) \\
= & \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[r_{i a} r_{j b} 1\{\widetilde{U}\}\right]-\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[r_{i a} 1\{\widetilde{U}\}\right] \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[r_{j b} 1\{\widetilde{U}\}\right]+O\left(e^{-c n}\right) \\
= & \mathbb{P}_{n, \theta, \boldsymbol{\beta}}(\widetilde{U}) \operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(r_{i a}, r_{j b} \mid \widetilde{U}\right)+O\left(e^{-c n}\right)=O\left(n^{-1}\right), \tag{2.51}
\end{align*}
$$

where the last equality again uses Lemma 2.3.6, along with (2.49). Combining (2.47), (2.48), (2.50) and (2.51) we have

$$
\begin{aligned}
& \operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(\left(\phi_{i}-t\right)^{2}, \phi_{j}-t \mid \widetilde{U}\right) \\
= & \frac{\theta^{3} \operatorname{sech}^{6}(2 \theta t)}{(n-1)^{3}} \sum_{a_{1}, a_{2} \neq i, b \neq j} \sum_{u \in\left\{i, a_{1}\right\}, v \in\left\{i, a_{2}\right\}, w \in\{j, b\}} \operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(\left(\phi_{u}-t\right)\left(\phi_{v}-t\right),\left(\phi_{w}-t\right) \mid \widetilde{U}\right)+O\left(n^{-2}\right),
\end{aligned}
$$

which verifies (2.45) for the choice $\left\{i_{1}=i_{2}=i, i_{3}=j\right\}$. This completes the proof of the claim.

Proof of Lemma 2.3.5. We proceed via a similar argument as in the proof of Lemma 2.3.8. Setting $M\left(i_{1}, i_{2}\right):=\operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(k_{i_{1}}, k_{i_{2}} \mid U\right)$ for $1 \leq i_{1}, i_{2} \leq n$, we begin by claiming

$$
\begin{equation*}
\max _{1 \leq i_{1}, i_{2} \leq n}\left|M\left(i_{1}, i_{2}\right)-\frac{1}{(n-1)^{2}} \sum_{j_{1} \neq i_{1}, j_{2} \neq i_{2}} \sum_{u \in\left\{i_{1}, j_{1}\right\}, v \in\left\{i_{2}, j_{2}\right\}}\left[C_{0}+C_{1}\left(\beta_{u}+\beta_{v}\right)\right] M(u, v)\right|=O(1), \tag{2.52}
\end{equation*}
$$

where

$$
\begin{equation*}
C_{0}:=\theta^{2} \operatorname{sech}^{4}(2 \theta t), \quad C_{1}:=\frac{\theta^{2}}{2} \operatorname{sech}^{2}(2 \theta t) \tag{2.53}
\end{equation*}
$$

Given (2.52), and noting that $\operatorname{Var}_{n, \theta, \boldsymbol{\beta}}\left(k_{i} \mid U\right)=O(n)$ by Lemma 2.3.6, we conclude

$$
\begin{equation*}
\max _{i_{1} \neq i_{2}}\left|M\left(i_{1}, i_{2}\right)-\sum_{j_{1} \neq j_{2}} B_{n}\left(\left(i_{1}, i_{2}\right),\left(j_{1}, j_{2}\right)\right) M\left(j_{1}, j_{2}\right)\right|=O(1) \tag{2.54}
\end{equation*}
$$

where $B_{n}$ is a symmetric $n(n-1)$ matrix with non-negative entries, satisfying

$$
\sum_{j_{1} \neq j_{2}} B_{n}\left(\left(i_{1}, i_{2}\right),\left(j_{1}, j_{2}\right)\right) \leq 8\left(C_{0}+2 A\right) \xrightarrow{A \rightarrow 0} 8 C_{0}=8 \theta^{3} \operatorname{sech}^{6}(2 \theta t)<1 .
$$

Thus the $\ell_{\infty}$ operator norm of $\left(\mathbf{I}-B_{n}\right)^{-1}$ converges to $\left(1-8 \theta^{3} \operatorname{sech}^{6}(2 \theta t)\right)^{-1}<\infty$, which along with (2.54) gives

$$
\max _{i_{1} \neq i_{2}} M\left(i_{1}, i_{2}\right)=O(1)
$$

as desired.

It thus remains to verify (2.52). To this end, for any $i \neq j$, we have

$$
\begin{align*}
& \mathbb{P}_{n, \theta, \boldsymbol{\beta}}(U) \operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(k_{i}, k_{j} \mid U\right) \\
= & \mathbb{P}_{n, \theta, \boldsymbol{\beta}}(U) \sum_{a \neq i, b \neq j} \operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(Y_{i a}, Y_{j b} \mid U\right) \\
= & \sum_{a \neq i, b \neq j}\left\{\mathbb{E}_{\beta}\left[Y_{i a} Y_{j b} 1\{U\}\right]-\mathbb{E}_{\beta}\left[Y_{i a} 1\{U\}\right] \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[Y_{j b} 1\{U\}\right]\right\} \\
= & \sum_{a \neq i, b \neq j}\left\{\mathbb{E}_{\beta}\left[Y_{i a} Y_{j b} 1\{\widetilde{U}\}\right]-\mathbb{E}_{\beta}\left[Y_{i a} 1\{\widetilde{U}\}\right] \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[Y_{j b} 1\{\widetilde{U}\}\right]\right\}+O\left(e^{-c n}\right) \\
= & \sum_{a \neq i, b \neq j}\left\{\mathbb{E}_{\beta}\left[r_{i a} r_{j b} 1\{\widetilde{U}\}\right]-\mathbb{E}_{\beta}\left[r_{i a} 1\{\widetilde{U}\}\right]-\mathbb{E}_{\beta}\left[r_{j b} 1\{\widetilde{U}\}\right]\right\}+O\left(e^{-c n}\right) \\
= & \mathbb{P}_{n, \theta, \boldsymbol{\beta}}(\widetilde{U}) \operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(r_{i a}, r_{j b} \mid \widetilde{U}\right)+O\left(e^{-c n}\right) . \tag{2.55}
\end{align*}
$$

In the above display, $r_{i j}:=\tanh \left[\theta\left(\phi_{i}+\phi_{a}\right)+\frac{1}{2}\left(\beta_{i}+\beta_{a}\right)\right]$. A Taylor's series expansion gives

$$
\begin{aligned}
r_{i j} & =\tanh \left[\theta\left(\phi_{i}+\phi_{j}\right)+\frac{1}{2}\left(\beta_{i}+\beta_{j}\right)\right] \\
& =\tanh (2 \theta t)+\left[\theta\left(\phi_{i}-t+\phi_{j}-t\right)+\frac{1}{2}\left(\beta_{i}+\beta_{j}\right)\right] \operatorname{sech}^{2}(2 \theta t) \\
& +\frac{1}{2}\left[\theta\left(\phi_{i}-t+\phi_{j}-t\right)+\frac{1}{2}\left(\beta_{i}+\beta_{j}\right)\right]^{2} \tanh ^{\prime \prime}(2 \theta t)+\xi_{i j},
\end{aligned}
$$

where

$$
\left|\xi_{i j}\right| \lesssim\left|\phi_{i}-t\right|^{3}+\left|\phi_{j}-t\right|^{3}+\left|\beta_{i}\right|^{3}+\left|\beta_{j}\right|^{3} \lesssim\left|\phi_{i}-t\right|^{3}+\left|\phi_{j}-t\right|^{3}+n^{-3 / 2} .
$$

Using the above display we have

$$
\begin{equation*}
\operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(r_{i a}, r_{j b} \mid \widetilde{U}\right)=\sum_{u \in\{i, a\}, v \in\{j, b\}}\left[C_{0}+C_{1}\left(\beta_{u}+\beta_{v}\right)\right] \operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(\phi_{u}, \phi_{v} \mid \widetilde{U}\right)+O\left(n^{-2}\right), \tag{2.56}
\end{equation*}
$$

where the bound on the error term uses Lemma 2.3.6 and Lemma 2.3.8. In the above display, the constants $C_{0}, C_{1}$ are as in (2.53).

Finally, we have

$$
\begin{align*}
& \mathbb{P}_{n, \theta, \boldsymbol{\beta}}(\widetilde{U}) \operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(\phi_{u}, \phi_{v} \mid \widetilde{U}\right) \\
= & \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[\phi_{u} \phi_{v} 1\{\widetilde{U}\}\right]-\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[\phi_{u} 1\{\widetilde{U}\}\right] \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[\phi_{v} 1\{\widetilde{U}\}\right] \\
= & \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[\phi_{u} \phi_{v} 1\{U\}\right]-\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[\phi_{u} 1\{U\}\right] \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[\phi_{v} 1\{U\}\right]+O\left(e^{-c n}\right) \\
= & \frac{1}{(n-1)^{2}} \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[k_{u} k_{v} 1\{U\}\right]-\frac{1}{(n-1)^{2}} \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[k_{u} 1\{U\}\right] \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left[k_{v} 1\{U\}\right]+O\left(e^{-c n}\right) \\
= & \frac{\mathbb{P}_{n, \theta, \boldsymbol{\beta}}(U)}{(n-1)^{2}} \operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(k_{u}, k_{v} \mid U\right)+O\left(e^{-c n}\right) . \tag{2.57}
\end{align*}
$$

Combining (2.55), (2.56) and (2.57) we have

$$
\operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(k_{i}, k_{j}\right)=\frac{1}{(n-1)^{2}} \sum_{a \neq i, b \neq j} \sum_{u \in\{i, a\}, v \in\{j, b\}}\left[C_{0}+C_{1}\left(\beta_{u}+\beta_{v}\right)\right] \operatorname{Cov}_{n, \theta, \boldsymbol{\beta}}\left(k_{u}, k_{v} \mid U\right)+O(1)
$$

from which (2.52) follows. This completes the proof of the lemma.

### 2.4 Proofs of Auxiliary Variable Lemmas

### 2.4.1 Proof of Proposition 2.3.1

The conditional distribution $(\phi \mid Y)$ has a density on $\mathbb{R}^{n}$ proportional to

$$
\begin{aligned}
\exp \left[-\frac{(n-1) \theta}{2} \sum_{i=1}^{n}\left(\phi_{i}-\frac{k_{i}}{n-1}\right)^{2}\right] & =\exp \left[-\frac{(n-1) \theta}{2} \sum_{i=1}^{n} \phi_{i}^{2}+\theta \sum_{i=1}^{n} \phi_{i} k_{i}-\frac{\theta}{2(n-1)} \sum_{i=1}^{n} k_{i}^{2}\right] \\
& =\exp \left[-\frac{(n-1) \theta}{2} \sum_{i=1}^{n} \phi_{i}^{2}-\frac{\theta}{2(n-1)} \sum_{i=1}^{n} k_{i}^{2}+\theta \sum_{i<j} Y_{i j}\left(\phi_{i}+\phi_{j}\right)\right] .
\end{aligned}
$$

Since $Y$ has a p.m.f. proportional to $\exp \left(\frac{\theta}{2} \sum_{i=1}^{n} k_{i}^{2}\right)$, the joint distribution of $(Y, \phi)$ has a density on $\{-1,1\}{ }^{\binom{n}{2}} \times \mathbb{R}^{n}$ proportional to

$$
\begin{equation*}
\exp \left[-\frac{(n-1) \theta}{2} \sum_{i=1}^{n} \phi_{i}^{2}+\theta \sum_{i<j} Y_{i j}\left(\phi_{i}+\phi_{j}\right)\right] \tag{2.58}
\end{equation*}
$$

(a) From (2.58), it follows that conditional on $\phi$ the random variables $\left\{Y_{i j}, 1 \leq i<j \leq n\right\}$ are mutually independent, with $Y_{i j}$ having the distribution as in part (a).
(b) Summing over the expression in (2.58), the marginal density of $\phi$ is proportional to

$$
\begin{aligned}
& \sum_{Y \in\{-1,1\}^{\binom{(2)}{2}}} \exp \left[-\frac{(n-1) \theta}{2} \sum_{i=1}^{n} \phi_{i}^{2}+\theta \sum_{i<j} Y_{i j}\left(\phi_{i}+\phi_{j}\right)\right] \\
= & 2^{\binom{n}{2}} \exp \left[-\frac{(n-1) \theta}{2} \sum_{i=1}^{n} \phi_{i}^{2}+\log \cosh \left(\theta\left(\phi_{i}+\phi_{j}\right)\right)\right] .
\end{aligned}
$$

Since the RHS above is proportional to $f_{n, \theta, \boldsymbol{\beta}}(\phi)$, the conclusion of part (b) follows.

### 2.4.2 Proof of Lemma 2.3.2

For proving Lemma 2.3.2, we need the following two lemmas.

Lemma 2.4.1. Suppose $\theta=1 / 2$, and $\boldsymbol{\beta} \in\left[0, n^{-1 / 2}\right]$. Then there exists a positive constant $M$ free of $n$, such that

$$
\begin{equation*}
\log \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\sum_{i=1}^{n}\left(\phi_{i}-\bar{\phi}\right)^{2}>M\right) \lesssim-n . \tag{2.59}
\end{equation*}
$$

Lemma 2.4.2. Suppose $\theta=1 / 2$, and $\boldsymbol{\beta} \in\left[0, n^{-1 / 2}\right]$. Then there exists a positive constant $M$ free of n, such that

$$
\begin{equation*}
\log \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(|\bar{\phi}| \geq M n^{-1 / 4}\right) \lesssim-n \tag{2.60}
\end{equation*}
$$

## Proof of Lemma 2.4.1

To begin, use (2.24) we get the existence of a finite positive constant $C$ free of $n$ such that

$$
\begin{align*}
& \left|\log f_{n, \theta, \boldsymbol{\beta}}-\log f_{n, \theta, \mathbf{0}}(\phi)-\frac{(n-1) s A}{2} \tanh (\bar{\phi})\right| \\
\leq & 2 C n A \sum_{i=1}^{s}\left|\phi_{i}-\bar{\phi}\right|+2 C s A \sum_{i=1}^{n}\left|\phi_{i}-\bar{\phi}\right|+C n s A^{2} \\
\leq & C n\left[\delta^{2} \sum_{i=1}^{s}\left(\phi_{i}-\bar{\phi}\right)^{2}+\frac{s A^{2}}{\delta^{2}}\right]+C s\left[\delta^{2} \sum_{i=1}^{n}\left(\phi_{i}-\bar{\phi}\right)^{2}+\frac{n A^{2}}{\delta^{2}}\right]+C n s A^{2} \\
\leq & C n \delta^{2} \sum_{i=1}^{n}\left(\phi_{i}-\bar{\phi}\right)^{2}+\frac{3 C n s A^{2}}{\delta^{2}} . \tag{2.61}
\end{align*}
$$

for any $\delta \in(0,1)$, where we use the bound $2 a b \leq a^{2}+b^{2}$ in the third inequality. Also, with $q(x)=\frac{x^{2}}{2}-\log \cosh (x)$ as in (2.21), we have $q^{\prime \prime}(x)=1-\operatorname{sech}^{2}(x) \in[0,1]$, where we use the fact that $\theta=\frac{1}{2}$. A Taylor's series expansion then gives

$$
\begin{equation*}
\left(\frac{\phi_{i}+\phi_{j}}{2}-\bar{\phi}\right) q^{\prime}(\bar{\phi}) \leq q\left(\frac{\phi_{i}+\phi_{j}}{2}\right)-q(\bar{\phi}) \leq\left(\frac{\phi_{i}+\phi_{j}}{2}-\bar{\phi}\right) q^{\prime}(\bar{\phi})+\frac{1}{2}\left(\frac{\phi_{i}+\phi_{j}}{2}-\bar{\phi}\right)^{2}, \tag{2.62}
\end{equation*}
$$

which on summing over $i<j$ and invoking with (2.25)

$$
\begin{equation*}
\frac{n(n-1)}{2} q(\bar{\phi})+\frac{n}{8} \sum_{i=1}^{n}\left(\phi_{i}-\bar{\phi}\right)^{2} \leq-\log f_{n, \theta, \mathbf{0}}(\phi) \leq \frac{n(n-1)}{2} q(\bar{\phi})+\frac{n}{4} \sum_{i=1}^{n}\left(\phi_{i}-\bar{\phi}\right)^{2} \tag{2.63}
\end{equation*}
$$

This gives

$$
\begin{align*}
& \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\sum_{i=1}^{n}\left(\phi_{i}-\bar{\phi}\right)^{2}>M\right) \\
= & \frac{\int_{\mathbb{R}^{n}} e^{-f_{n, \theta, \boldsymbol{\beta}}(\phi)} 1\left\{\sum_{i=1}^{n}\left(\phi_{i}-\bar{\phi}\right)^{2}>M\right\} d \phi}{\int_{\mathbb{R}^{n}} e^{-f_{n, \theta, \boldsymbol{\beta}}(\phi)} d \phi} \\
\leq & e^{\frac{3 C n A^{2}}{\delta^{2}}} \frac{\int_{\mathbb{R}^{n}} \exp \left(-\frac{n(n-1)}{2} q(\bar{\phi})-\frac{\lambda_{1}}{2} \sum_{i=1}^{n}\left(\phi_{i}-\bar{\phi}\right)^{2}\right) 1\left\{\sum_{i=1}^{n}\left(\phi_{i}-\bar{\phi}\right)^{2}>M\right\} d \phi}{\int_{\mathbb{R}^{n}} \exp \left(-\frac{n(n-1)}{2} q(\bar{\phi})-\frac{\lambda_{2}}{2} \sum_{i=1}^{n}\left(\phi_{i}-\bar{\phi}\right)^{2}\right) d \phi} \tag{2.64}
\end{align*}
$$

where $\lambda_{1}:=\frac{\theta}{2}-2 C \delta^{2}$ and $\lambda_{2}:=\theta+2 C \delta^{2}$ are positive reals, for the choice $\delta^{2}:=\frac{\theta}{8 C}$. Let $O_{n}$ be an orthogonal matrix with first row equal to $n^{-1 / 2} \mathbf{1}$. Then, setting $\psi=O_{n} \phi$ we have $\psi_{1}=\sqrt{n} \bar{\phi}$, and $\sum_{i=2}^{n} \psi_{i}^{2}=\sum_{i=1}^{n}\left(\phi_{i}-\bar{\phi}\right)^{2}$. Using this transformation, the ratio of integrals in the RHS of (2.64) equals

$$
\frac{\int_{\mathbb{R}^{n}} \exp \left(-\frac{n(n-1)}{2} q\left(n^{-1 / 2} \psi_{1}\right)-\frac{\lambda_{1}}{2} \sum_{i=2}^{n} \psi_{i}^{2}\right) 1\left\{\sum_{i=2}^{n} \psi_{i}^{2}>M\right\} d \psi}{\int_{\mathbb{R}^{n}} \exp \left(-\frac{n(n-1)}{2} q\left(n^{-1 / 2} \psi_{1}\right)-\frac{\lambda_{2}}{2} \sum_{i=2}^{n} \psi_{i}^{2}\right) d \psi}=\left(\frac{\lambda_{2}}{\lambda_{1}}\right)^{\frac{n-1}{2}} \mathbb{P}\left(\chi_{n-1}^{2}>M \lambda_{1}\right)
$$

The desired conclusion is immediate from standard tail bounds of the $\chi_{n-1}^{2}$ distribution.

## Proof of Lemma 2.4.2

To begin, note that

$$
\mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(|\bar{\phi}| \geq M n^{-1 / 4}\right)=\mathbb{P}_{n, \theta, \boldsymbol{\beta}}(|\bar{\phi}| \geq 2)+\mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(M n^{-1 / 4} \leq|\bar{\phi}| \leq 2\right),
$$

where $\log \mathbb{P}_{n, \theta, \boldsymbol{\beta}}(|\bar{\phi}| \geq 2) \lesssim-n$ using (2.18). Proceeding to bound the first term in the RHS of the above display, using an argument similar to the derivation of (2.64) we get

$$
\begin{aligned}
& \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(M n^{-1 / 4} \leq|\bar{\phi}| \leq 2\right) \\
\leq & e^{\frac{3 C n s A^{2}}{\delta^{2}}} \frac{\int_{\mathbb{R}^{n}} \exp \left(-\frac{n(n-1)}{2} q(\bar{\phi})-\frac{\lambda_{1}}{2} \sum_{i=1}^{n}\left(\phi_{i}-\bar{\phi}\right)^{2}\right) 1\left\{M n^{-1 / 4} \leq|\bar{\phi}| \leq 2\right\} d \phi}{\int_{\mathbb{R}^{n}} \exp \left(-\frac{n(n-1)}{2} q(\bar{\phi})-\frac{\lambda_{2}}{2} \sum_{i=1}^{n}\left(\phi_{i}-\bar{\phi}\right)^{2}\right) d \phi} \\
= & e^{\frac{3 C n s A^{2}}{\delta^{2}}} \frac{\int_{\mathbb{R}^{n}} \exp \left(-\frac{n(n-1)}{2} q\left(n^{-1 / 2} \psi_{1}\right)-\frac{\lambda_{1}}{2} \sum_{i=2}^{n} \psi_{i}^{2}\right) 1\left\{M n^{-1 / 4} \leq\left|n^{-1 / 2} \psi_{1}\right| \leq 2\right\} d \psi}{\int_{\mathbb{R}^{n}} \exp \left(-\frac{n(n-1)}{2} q\left(n^{-1 / 2} \psi_{1}\right)-\frac{\lambda_{2}}{2} \sum_{i=2}^{n} \psi_{i}^{2}\right) d \psi} \\
= & e^{\frac{3 C n s A^{2}}{\delta^{2}}}\left(\frac{\lambda_{2}}{\lambda_{1}}\right)^{\frac{n-1}{2}} \frac{\int_{\mathbb{R}} \exp \left(-\frac{n(n-1)}{2} q\left(n^{-1 / 2} \psi_{1}\right)\right) 1\left\{M n^{-1 / 4} \leq\left|n^{-1 / 2} \psi_{1}\right| \leq 2\right\} d \psi_{1}}{\int_{\mathbb{R}} \exp \left(-\frac{n(n-1)}{2} q\left(n^{-1 / 2} \psi_{1}\right)\right) d \psi_{1}}
\end{aligned}
$$

where we use the orthogonal transformation $\phi \mapsto \psi$ introduced in Lemma 2.4.1 in the last step. Now the function $q($.$) satisfies q^{\prime}(0)=q^{\prime \prime}(0)=q^{\prime \prime \prime}(0)=0$, and $q^{\prime \prime \prime \prime}(0)>0$. Since $q($.$) is$ continuous and does not vanish anywhere else on $\mathbb{R}$, there exists finite positive constants $c_{1}, c_{2}$ such that $c_{1} x^{4} \leq q(x) \leq c_{2} x^{4}$ for all $x \in[-2,2]$. Using this, the ratio of integrals in the above display can be bounded by

$$
\frac{\int_{\mathbb{R}} \exp \left(-c_{1}^{\prime} \psi_{1}^{4}\right) 1\left\{\left|\psi_{1}\right| \geq M n^{1 / 4}\right\} d \psi_{1}}{\int_{\mathbb{R}} \exp \left(-c_{2}^{\prime} \psi_{1}^{4}\right) d \psi_{1}}
$$

The desired conclusion follows from the above display using Laplace method for a suitable choice of $M$.

## Proof of Lemma 2.3.2

Without loss of generality, it suffices to work with $\phi_{1}$. For $2 \leq i \leq n$, using (2.20) we have

$$
\begin{align*}
p_{1 i}\left(\phi_{1}, \phi_{i}\right) & =\frac{1}{8}\left(\phi_{1}-\phi_{i}\right)^{2}+q\left(\frac{\phi_{1}+\phi_{i}}{2}\right)-\log \cosh \left(\frac{\phi_{1}+\phi_{i}}{2}+\frac{\beta_{1}+\beta_{i}}{2}\right)+\log \cosh \left(\frac{\phi_{1}+\phi_{j}}{2}\right) \\
& =\frac{1}{8}\left(\phi_{1}-\phi_{i}\right)^{2}+q\left(\frac{\phi_{1}+\phi_{i}}{2}\right)-\frac{\beta_{1}+\beta_{i}}{2} \tanh \left(\theta\left(\phi_{1}+\phi_{i}\right)\right)+O\left(\beta_{1}+\beta_{i}\right)^{2} . \tag{2.65}
\end{align*}
$$

Note that $q^{\prime \prime}(x) \in[0,1]$, which along with a Taylor's series expansion around $\bar{\phi}_{1}:=\frac{\sum_{j=2}^{n} \phi_{j}}{n-1}$ gives

$$
0 \leq q\left(\frac{\phi_{1}+\phi_{i}}{2}\right)-q\left(\bar{\phi}_{1}\right)-\left(\frac{\phi_{1}+\phi_{i}}{2}-\bar{\phi}_{1}\right) q^{\prime}\left(\bar{\phi}_{1}\right) \leq \frac{1}{2}\left(\frac{\phi_{1}+\phi_{i}}{2}-\bar{\phi}_{1}\right)^{2} .
$$

On adding over $i \in[2, n]$ and using the previous display, this gives

$$
\begin{align*}
& \frac{n-1}{8}\left(\phi_{1}-\bar{\phi}_{1}\right)^{2}+\frac{1}{8} \sum_{i=2}^{n}\left(\phi_{i}-\bar{\phi}_{1}\right)^{2} \\
\leq & \frac{\theta}{4} \sum_{i=2}^{n}\left(\phi_{1}-\phi_{i}\right)^{2}+\sum_{i=2}^{n} q\left(\frac{\phi_{1}+\phi_{i}}{2}\right)-(n-1)\left[q\left(\bar{\phi}_{1}\right)+\frac{1}{2} q^{\prime}\left(\bar{\phi}_{1}\right)\left(\phi_{1}-\bar{\phi}_{1}\right)\right] \\
\leq & \frac{n-1}{4}\left(\phi_{1}-\bar{\phi}_{1}\right)^{2}+\frac{1}{4} \sum_{i=2}^{n}\left(\phi_{i}-\bar{\phi}_{1}\right)^{2} . \tag{2.66}
\end{align*}
$$

Another Taylor's series approximation gives

$$
\tanh \left(\frac{\phi_{1}+\phi_{i}}{2}\right)=\tanh \left(\bar{\phi}_{1}\right)+\left(\frac{\phi_{1}+\phi_{i}}{2}-\bar{\phi}_{1}\right) \operatorname{sech}^{2}\left(\bar{\phi}_{1}\right)+O\left(\phi_{1}+\phi_{i}-2 \bar{\phi}_{1}\right)^{2},
$$

which on summing over $i \in[2, n]$ gives

$$
\begin{align*}
& \sum_{i=2}^{n} \frac{\beta_{1}+\beta_{i}}{2} \tanh \left(\theta\left(\phi_{1}+\phi_{i}\right)\right) \\
= & \tanh \left(\bar{\phi}_{1}\right) \sum_{i=2}^{n} \frac{\beta_{1}+\beta_{i}}{2}+\frac{1}{4} \sum_{i=2}^{n} \beta_{i}\left(\phi_{i}-\bar{\phi}_{1}\right) \operatorname{sech}^{2}\left(\bar{\phi}_{1}\right) \\
+ & O\left((n-1) A\left|\phi_{1}-\bar{\phi}_{1}\right|+(n-1) A\left(\phi_{1}-\bar{\phi}_{1}\right)^{2}+A \sum_{i=2}^{n}\left(\phi_{i}-\bar{\phi}_{1}\right)^{2}\right) . \tag{2.67}
\end{align*}
$$

Combining (2.66) and (2.67) along with (2.65) we get the existence of a positive constant $C$ free of $n$, such that

$$
\begin{align*}
& \frac{1}{9}\left[(n-1)\left(\phi_{1}-\bar{\phi}_{1}\right)^{2}+\sum_{i=2}^{n}\left(\phi_{i}-\bar{\phi}_{1}\right)^{2}\right]-C(n-1) A\left|\phi_{1}-\bar{\phi}_{1}\right|-C n A^{2} \\
\leq & \sum_{i=2}^{n} p_{1 i}\left(\phi_{1}, \phi_{i}\right)-\varphi\left(\phi_{i}, 2 \leq i \leq n\right)-\frac{n-1}{2} q^{\prime}\left(\bar{\phi}_{1}\right)\left(\phi_{1}-\bar{\phi}_{1}\right) \\
\leq & \frac{1}{3}\left[(n-1)\left(\phi_{1}-\bar{\phi}_{1}\right)^{2}+\sum_{i=2}^{n}\left(\phi_{i}-\bar{\phi}_{1}\right)^{2}\right]+C(n-1) A\left|\phi_{1}-\bar{\phi}_{1}\right|+C n A^{2} . \tag{2.68}
\end{align*}
$$

In (2.68), we have set

$$
\varphi\left(\phi_{i}, 2 \leq i \leq n\right):=(n-1) q\left(\bar{\phi}_{1}\right)-\tanh \left(\bar{\phi}_{1}\right) \sum_{i=2}^{n} \frac{\beta_{1}+\beta_{i}}{2}-\frac{1}{4} \sum_{i=2}^{n} \beta_{i}\left(\phi_{i}-\bar{\phi}_{1}\right) \operatorname{sech}^{2}\left(\bar{\phi}_{1}\right)
$$

which is a function which does not depend on $\phi_{1}$. Set

$$
D:=\left\{\sum_{i=2}^{n}\left(\phi_{i}-\bar{\phi}_{1}\right)^{2} \leq M,\left|\bar{\phi}_{1}\right| \leq M n^{-1 / 4}\right\},
$$

where $M$ is a constant free of $n$ such that $\log \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(U^{c}\right) \lesssim-n$. The existence of such a constant follows from Lemmas 2.4.1 and 2.4.2. Then we have

$$
\begin{aligned}
& \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left|\phi_{1}-\bar{\phi}_{1}\right|^{l}=\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left|\phi_{1}-\bar{\phi}_{1}\right|^{l} 1_{D}+\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left|\phi_{1}-\bar{\phi}_{1}\right|^{l} 1_{D^{c}} \\
\leq & \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left(\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left(\left|\phi_{1}-\bar{\phi}_{1}^{l}\right| \phi_{i}, 2 \leq i \leq n\right) 1_{D}\right)+\sqrt{\mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left|\phi_{1}-\bar{\phi}_{1}\right|^{2 l}} \sqrt{\mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(D^{c}\right)} .
\end{aligned}
$$

Since $\mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(D^{c}\right)$ decays exponentially, to complete the argument it suffices to show that

$$
\begin{equation*}
\sup _{\left(\phi_{2}, \ldots, \phi_{n}\right) \in D} \mathbb{E}_{n, \theta, \beta}\left(\left|\phi_{1}-\bar{\phi}_{1}\right|^{l} \mid \phi_{i}, 2 \leq i \leq n\right) \lesssim n^{-\ell / 2} \tag{2.69}
\end{equation*}
$$

Proceeding to show (2.69), using (2.68) we have

$$
\begin{align*}
& \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left(\left|\phi_{1}-\bar{\phi}_{1}\right|^{l} \mid \phi_{i}, 2 \leq i \leq n\right) \\
&= \frac{\int_{\mathbb{R}}\left|\phi_{1}-\bar{\phi}_{1}\right|^{l} \exp \left(-\sum_{i=2}^{n} p_{1 i}\left(\phi_{1}, \phi_{i}\right)\right) d \phi_{1}}{\int_{\mathbb{R}} \exp \left(-\sum_{i=2}^{n} p_{1 i}\left(\phi_{1}, \phi_{i}\right)\right) d \phi_{1}} \\
& \leq \exp \left(\left(\frac{1}{3}-\frac{1}{9}\right) \sum_{i=2}^{n}\left(\phi_{i}-\bar{\phi}_{1}\right)^{2}+2 C n A^{2}\right) \\
& \times \frac{\int_{\mathbb{R}}\left|\phi_{1}-\bar{\phi}_{1}\right|^{\ell} \exp \left(-\frac{n-1}{3}\left(\phi_{1}-\bar{\phi}_{1}\right)^{2}-\frac{n-1}{2}\left(\phi_{1}-\bar{\phi}_{1}\right) q^{\prime}\left(\bar{\phi}_{1}\right)+C \sqrt{n-1}\left|\phi_{1}-\bar{\phi}_{1}\right|\right) d \phi_{1}}{\int_{\mathbb{R}} \exp \left(-\frac{n-1}{9}\left(\phi_{1}-\bar{\phi}_{1}\right)^{2}-\frac{n-1}{2}\left(\phi_{1}-\bar{\phi}_{1}\right) q^{\prime}\left(\bar{\phi}_{1}\right)-C \sqrt{n-1}\left|\phi_{1}-\bar{\phi}_{1}\right|\right) d \phi_{1}} \\
& \leq \exp \left(\frac{2 M}{9}+2 C\right) \\
& \times \frac{\int_{\mathbb{R}}\left|\phi_{1}-\bar{\phi}_{1}\right|^{\ell} \exp \left(-\frac{n-1}{3}\left(\phi_{1}-\bar{\phi}_{1}\right)^{\left.+\frac{c M^{3}(n-1)}{2 n^{3 / 4}}\left|\phi_{1}-\bar{\phi}_{1}\right|+C \sqrt{n-1}\left|\phi_{1}-\bar{\phi}_{1}\right|\right) d \phi_{1}}\right.}{\int_{\mathbb{R}} \exp \left(-\frac{n-1}{9}\left(\phi_{1}-\bar{\phi}_{1}\right)^{2}-\frac{c M^{3}(n-1)}{2 n^{3 / 4}}\left|\phi_{1}-\bar{\phi}_{1}\right|-C \sqrt{n-1}\left|\phi_{1}-\bar{\phi}_{1}\right|\right) d \phi_{1}}, \tag{2.70}
\end{align*}
$$

where $c:=\sup _{x \in[-1,1]} \frac{\left|q^{\prime}(x)\right|}{|x|^{3}}$. By a change of variable, the RHS of (2.70) becomes

$$
\frac{e^{\frac{2 M}{9}+2 C}}{(n-1)^{\ell / 2}} \frac{\int_{\mathbb{R}}|x|^{\ell} e^{-\frac{x^{2}}{3}+\frac{c M^{3} \sqrt{n-1}}{2 n^{3 / 4}}|x|+C|x|} d x}{\int_{\mathbb{R}} e^{-\frac{x^{2}}{9}-\frac{c M^{3} \sqrt{n-1}}{2 n^{3 / 4}}-C|x|} d x} \lesssim n^{-\ell / 2}
$$

from which (2.69) follows. This completes the proof of the lemma.

### 2.4.3 Proof of Lemma 2.3.6

We begin by proving two lemmas, which will be useful for the proof.

Lemma 2.4.3. For every $a \in \mathbb{R}$, define the function $q_{a}: \mathbb{R} \mapsto \mathbb{R}$ by setting

$$
q_{a}(x):=\theta x^{2}-\log \cosh (2 \theta x+a) .
$$

Denote by $t(a)$ the largest root of the equation $q_{a}^{\prime}(x)=0$. Then the following conclusions hold:
(a) The map $t($.$) is well defined and C_{1}$ on $(-2 \delta, 2 \delta)$, for some $\delta>0$.
(b) If $a \geq 0$, then $t(a)$ is the unique global maximizer of $q_{a}($.$) in [0, \infty)$.
(c) There exists finite positive reals $\lambda_{1}^{\prime}, \lambda_{2}^{\prime}$ such that for all $x, y \in[0,2]$ and $a \in[0, \delta]$ we have

$$
\lambda_{1}^{\prime}\left[(x-t)^{2}+(y-t)^{2}\right]-\lambda_{1}^{\prime} a^{2} \leq q_{a}(x)-q_{a}(t(a)) \leq \lambda_{2}^{\prime}\left[(x-t)^{2}+(y-t)^{2}\right]+\lambda_{2}^{\prime} a^{2} .
$$

Lemma 2.4.4. Suppose $\theta>1 / 2$, and $\boldsymbol{\beta} \in\left[0,2 n^{-1 / 2}\right]^{n}$. Then there exists a positive constant $M$ depending on $n$ such that

$$
\begin{equation*}
\log \mathbb{P}_{n, \theta, \beta}\left(\sum_{i=1}^{n}\left(\phi_{i}-t\right)^{2}>M \mid \phi \in[0,2]^{n}\right) \lesssim-n \tag{2.71}
\end{equation*}
$$

## Proof of Lemma 2.4.3

(a) Since $q_{a}^{\prime}(x)=2 \theta[x-\tanh (2 \theta x+a)]$, it follows that $q_{a}($.$) has an odd number of roots in$ $\mathbb{R}$, and so the maximum root is well defined for all $a \in \mathbb{R}$, and satisfies $x=\tanh (2 \theta x+a)$. If $a=0$, then the desired conclusion follows from Lemma 2.3.4, with $t(0)=t$. Since $q_{0}^{\prime \prime \prime}(t) \neq 0$, we must have $q_{0}^{\prime \prime}(t)>0$, and so using Implicit function theorem, there exists
$\delta>0$ (depending on $\theta$ ) such that for all $a \in(-2 \delta, 2 \delta)$ the map $t($.$) is \mathcal{C}_{1}$.
(b) If $a=0$ then the conclusion follows from Lemma 2.3.4, and so we assume $a>0$. Since $q_{a}(x) \rightarrow \infty$ as $x \rightarrow \infty$, the function $q_{a}($.$) attains a global minima at a finite number in$ $[0, \infty)$. Also since $q_{a}^{\prime}(0+)=-2 \theta \tanh (2 \theta+a)<0,0$ is not a minima of $q_{a}($.$) . Since q_{a}($. has a unique root in $(0, \infty)$ for $a>0$, the desired conclusion follows.
(c) Define the function $Q(.,):.[0, \delta] \times[0,2] \mapsto \mathbb{R}$ by setting

$$
\begin{aligned}
Q(a, t): & :=\frac{q_{a}(x)-q_{a}(t(a))}{\left(x-t((a))^{2}\right.} \text { if } x \neq t(a), \\
& =\frac{1}{2} q_{a}^{\prime \prime}(t(a)) \text { if } x=t(a) .
\end{aligned}
$$

Using part (b) we have $Q(.,$.$) is strictly positive point-wise, as t(a)$ is the unique global minimizer of $q_{a}($.$) in [0, \infty)$. On the other hand, using part (a) we have $Q(.,$.$) is continuous.$ Since a continuous function on a compact set attains its maximum and minimum, we have

$$
\begin{equation*}
\lambda_{1}^{\prime \prime}:=\inf _{a \in[0, \delta], x \in[0,2]} Q(a, t) \leq \sup _{a \in[0, \delta], x \in[0,2]} Q(a, t)=: \lambda_{2}^{\prime \prime}, \tag{2.72}
\end{equation*}
$$

Using (2.72) we get

$$
\begin{aligned}
& q_{a}(x)-q_{a}(t(a)) \leq \lambda_{2}^{\prime \prime}\left[(x-t(a))^{2}+(y-t(a))^{2}\right] \leq \lambda_{2}^{\prime \prime}\left[(x-t)^{2}+(y-t)^{2}\right]+4 \lambda_{2}^{\prime \prime}(t(a)-t)^{2}, \\
& q_{a}(x)-q_{a}(t(a)) \geq \lambda_{1}^{\prime \prime}\left[(x-t(a))^{2}+(y-t(a))^{2}\right] \geq \lambda_{1}^{\prime \prime}\left[(x-t)^{2}+(y-t)^{2}\right]-4 \lambda_{1}^{\prime \prime}(t(a)-t)^{2} .
\end{aligned}
$$

The desired conclusion then follows on using part (a) to note the existence of $c>0$ such that $|t(a)-t| \leq c|a|$ for all $a \in[-\delta, \delta]$.

## Proof of Lemma 2.4.4

With $p_{i j}\left(\phi_{i}, \phi_{j}\right)$ as in (2.20) we can write

$$
\begin{equation*}
p_{i j}\left(\phi_{i}, \phi_{j}\right)=\frac{\theta}{4}\left(\phi_{i}-\phi_{j}\right)^{2}+q_{\frac{\beta_{i}+\beta_{j}}{2}}\left(\frac{\phi_{i}+\phi_{j}}{2}\right), \tag{2.73}
\end{equation*}
$$

where the function $q_{a}($.$) is defined in Lemma 2.4.3. For \phi_{i}, \phi_{j} \in[0,2]$ and $\beta_{i}, \beta_{j} \in\left[0,2 n^{-1 / 2}\right]$, using part (c) of Lemma 2.4.3 gives

$$
\lambda_{1}^{\prime}\left[\frac{\phi_{i}+\phi_{j}}{2}-t\right]^{2}-\frac{\lambda_{1}^{\prime}}{n} \leq q_{\frac{\beta_{i}+\beta_{j}}{2}}\left(\frac{\phi_{i}+\phi_{j}}{2}\right)-q_{\frac{\beta_{i}+\beta_{j}}{2}}\left(t\left(\frac{\beta_{i}+\beta_{j}}{2}\right)\right) \leq \lambda_{2}^{\prime}\left[\frac{\phi_{i}+\phi_{j}}{2}-t\right]^{2}+\frac{\lambda_{2}^{\prime}}{n} .
$$

Using this along with (2.73), this gives the existence of finite positive constants $\lambda_{1}$ and $\lambda_{2}$, such that

$$
\begin{align*}
\frac{\lambda_{1}}{2}\left[\left(\phi_{i}-t\right)^{2}+\left(\phi_{j}-t\right)^{2}\right]-n \lambda_{1} & \leq p_{i j}\left(\phi_{i}, \phi_{j}\right)-q_{\frac{\beta_{i}+\beta_{j}}{2}}\left(t\left(\frac{\beta_{i}+\beta_{j}}{2}\right)\right)  \tag{2.74}\\
& \leq \frac{\lambda_{2}}{2}\left[\left(\phi_{i}-t\right)^{2}+\left(\phi_{j}-t\right)^{2}\right]+n \lambda_{2}
\end{align*}
$$

Summing over $i<j$ we get

$$
\begin{aligned}
\frac{(n-1) \lambda_{1}}{2} \sum_{i=1}^{n}\left(\phi_{i}-t\right)^{2}-n \lambda_{1} & \leq \sum_{i<j} p_{i j}\left(\phi_{i}, \phi_{j}\right)-\sum_{i<j} q_{\frac{\beta_{i}+\beta_{j}}{2}}\left(t\left(\frac{\beta_{i}+\beta_{j}}{2}\right)\right) \\
& \leq \frac{(n-1) \lambda_{2}}{2} \sum_{i=1}^{n}\left(\phi_{i}-t\right)^{2}+n \lambda_{2}
\end{aligned}
$$

which gives

$$
\begin{aligned}
\mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\sum_{i=1}^{n}\left(\phi_{i}-t\right)^{2}>M \mid \phi \in[0,2]^{n}\right) & =\frac{\int_{\mathbb{R}^{n}} e^{-\sum_{i<j} p\left(\phi_{i}, \phi_{i}\right)} 1\left\{\sum_{i=1}^{n}\left(\phi_{i}-t\right)^{2}>M\right\} d \phi}{\int_{\mathbb{R}^{n}} e^{-\sum_{i<j} p\left(\phi_{i}, \phi_{i}\right)} d \phi_{1}} \\
& \leq e^{n\left(\lambda_{1}+\lambda_{2}\right)} \frac{\int_{\mathbb{R}^{n}} e^{-\frac{(n-1) \lambda_{1}}{2} \sum_{i=1}^{n}\left(\phi_{i}-t\right)^{2}} 1\left\{\sum_{i=1}^{n}\left(\phi_{i}-t\right)^{2}>M\right\} d \phi}{\int_{[0,2]^{n}} e^{-\frac{(n-1) \lambda_{2}}{2} \sum_{i=1}^{n}\left(\phi_{i}-t\right)^{2}} d \phi} \\
& \leq e^{n\left(\lambda_{1}+\lambda_{2}\right)}\left(\frac{\lambda_{2}}{\lambda_{1}}\right)^{n / 2} \frac{\mathbb{P}\left(\chi_{n}^{2}>(n-1) M \lambda_{1}\right)}{\mathbb{P}\left(N\left(0, \lambda_{2}^{-1}\right) \in \sqrt{n-1}[-t, 2-t]\right)^{n}},
\end{aligned}
$$

The desired conclusion then follows on using standard tail bounds of the $\chi_{n}^{2}$ distribution.

## Proof of Lemma 2.3.6

Summing over display (2.74) for $j \in[2, n]$ gives

$$
\begin{aligned}
\frac{\lambda_{1}}{2} \sum_{j=2}^{n}\left(\phi_{j}-t\right)^{2}-\frac{\lambda_{1}}{2}\left(\phi_{1}-t\right)^{2}-\lambda_{1} & \leq \sum_{j=2}^{n}\left[p_{1 j}\left(\phi_{1}, \phi_{j}\right)-q_{\frac{\beta_{1}+\beta_{j}}{2}}\left(t_{1}, t_{j}\right)\right] \\
& \leq \frac{\lambda_{2}}{2} \sum_{j=2}^{n}\left(\phi_{j}-t\right)^{2}+\frac{(n-1) \lambda_{2}}{2}\left(\phi_{1}-t\right)^{2}+\lambda_{2}
\end{aligned}
$$

Thus, with $D:=\left\{\sum_{j=2}^{n}\left(\phi_{j}-t\right)^{2} \leq M\right\}$, for $\left(\phi_{2}, \ldots, \phi_{n}\right) \in D \cap[0,2]^{n-1}$ we have

$$
\begin{aligned}
& \mathbb{E}_{n, \theta, \boldsymbol{\beta}}\left(\left|\phi_{1}-t\right|^{\ell} 1\left\{\phi_{1} \in[0,2]\right\} \mid \phi_{i}, i \neq 1\right) \\
= & \frac{\int_{[0,2]}\left|\phi_{1}-t\right|^{\ell} \prod_{i=2}^{n} e^{-p_{1 i}\left(\phi_{1}, \phi_{i}\right)} d \phi_{1}}{\int_{[0,2]} \prod_{i=2}^{n} e^{\left.-p_{1 i}\left(\phi_{1}, \phi_{i}\right)\right)} d \phi_{1}} \\
\leq & e^{\frac{\lambda_{2}-\lambda_{1}}{2} \sum_{i=2}^{n}\left(\phi_{i}-t\right)^{2}+\lambda_{1}+\lambda_{2} \int_{[0,2]} \exp \left(\frac{-(n-1) \lambda_{1}\left(\phi_{1}-t\right)^{2}}{2}\right)\left|\phi_{1}-t\right|^{\ell} d \phi_{1}} \\
\leq & e^{\left(\lambda_{2}-\lambda_{1}\right) M+\lambda_{1}+\lambda_{2}} \sqrt{\frac{\lambda_{2}}{(n-1)^{\ell} \lambda_{1}}} \frac{\exp \left(\frac{-(n-1) \lambda_{1}\left(\phi_{1}-t\right)^{2}}{2}\right) d \phi_{1}}{\mathbb{E}\left|N\left(0, \lambda_{1}^{-1}\right)\right|^{\ell}} \\
&
\end{aligned}
$$

Since the probability in the denominator above converges to 1 , using tail estimates of the normal distribution we get

$$
\sup _{\left(\phi_{2}, \ldots, \phi_{n}\right) \in\left\{D \cap[0,2]^{n-1}\right\}} \mathbb{E}\left[\left|\phi_{1}-t\right|^{\ell} 1\left\{\phi_{1} \in[0,2]\right\} \mid \phi_{i}, 2 \leq i \leq n\right] \lesssim n^{-\ell / 2} .
$$

The desired conclusion follows from the last display above, and using Lemma 2.4.4 to get that $\log \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(D^{c} \mid \phi \in[0,2]^{n}\right) \lesssim-n$.

### 2.4.4 Proof of Lemma 2.3.7

We first prove the following lemma which will be used in proving Lemma 2.3.7.

Lemma 2.4.5. Suppose $\theta>1 / 2$, and $\boldsymbol{\beta} \in\left[0,2 n^{-1 / 2}\right]^{n}$. Then for any $\delta>0$ there exists a constant c such that

$$
\log \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\max _{i \in[n]}\left|\phi_{i}-t\right|>\delta \mid \phi \in[0,2]^{n}\right) \leq-c n
$$

## Proof of Lemma 2.4.5

Set $D:=\left\{\sum_{i=2}^{n}\left(\phi_{i}-t\right)^{2} \leq M\right\}$, and use (2.74) to note that for any $\delta>0$ and $\left(\phi_{2}, \ldots, \phi_{n}\right) \in$ $D \cap[0,2]^{n-1}$ we have

$$
\begin{aligned}
& \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\left|\phi_{1}-t\right|>\delta, \phi_{1} \in[0,2] \mid \phi_{i}, 2 \leq i \leq n\right) \\
= & \frac{\int_{[0,2]} 1\left\{\left|\phi_{i}-t\right|>\delta\right\}}{\prod_{[0,2]}^{n} e_{i=2}^{n} e^{-p_{1 i}\left(\phi_{1}, \phi_{i}\right)} d \phi_{1}} \\
\leq & e^{\frac{\lambda_{2}-\lambda_{1}\left(\phi_{1}, \phi_{i}\right)}{2} \sum_{i=2}^{n}\left(\phi_{i}-t\right)^{2}+\lambda_{1}+\lambda_{2}} \frac{\int_{[0,2]} \exp \left(\frac{-(n-1) \lambda_{1}\left(\phi_{1}-t\right)^{2}}{2}\right) 1\left\{\left|\phi_{i}-t\right|>\delta\right\} d \phi_{1}}{\int_{[0,2]} \exp \left(\frac{-(n-1) \lambda_{1}\left(\phi_{1}-t\right)^{2}}{2}\right) d \phi_{1}} \\
\leq & e^{\left(\lambda_{2}-\lambda_{1}\right) M+\lambda_{1}+\lambda_{2}} \sqrt{\frac{\lambda_{2}}{\lambda_{1}}} \frac{\mathbb{P}\left(N\left(0, \lambda_{1}^{-1}\right) \mid>\delta \sqrt{n-1}\right)}{\mathbb{P}\left(N\left(0, \lambda_{2}^{-1}\right) \in[-t \sqrt{n-1}, 2 \sqrt{n-1}]\right.} .
\end{aligned}
$$

From the above display, we get

$$
\sup _{\left(\phi_{2}, \ldots, \phi_{n}\right) \in D \cap[0,2]^{n-1}} \log \mathbb{P}\left(\left|\phi_{1}-t\right|>\delta 1\left\{\phi_{1} \in[0,2] \mid \phi_{i}, 2 \leq i \leq n\right) \lesssim-n .\right.
$$

Recalling that $\log \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(D^{c} \mid \phi \in[0,2]^{n}\right) \lesssim-n$, we get

$$
\log \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\left|\phi_{1}-t\right|>\delta, \mid \phi \in[0,2]^{n}\right) \lesssim-n .
$$

A similar argument applies to all co-ordinates of $\phi$, and so a union bound gives

$$
\log \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\max _{i \in[n]}\left|\phi_{i}-t\right|>\delta \mid \phi \in[0,2]^{n}\right) \lesssim-n,
$$

as desired.

## Proof of Lemma 2.3.7

(a) Note that

$$
\begin{aligned}
\mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(U \cap \widetilde{U}^{c}\right) & =\mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\min _{i \in[n]} \phi_{i} \geq 0, \max _{i \in[n]} k_{i}<\frac{(n-1) t}{2}\right) \\
& \leq \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\min _{i \in[n]} \phi_{i} \geq \frac{3 t}{4}, \max _{i \in[n]} k_{i}<\frac{(n-1) t}{2}\right)+\mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(\left.\min _{i \in[n]} \phi_{i} \leq \frac{3 t}{4} \right\rvert\, \phi \in[0,2]^{n}\right) .
\end{aligned}
$$

The two terms in the RHS above decays exponentially using (2.18) and Lemma 2.4.5 respectively, and so

$$
\log \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(U \cap \widetilde{U}^{c}\right) \lesssim-n
$$

Also,

$$
\log \mathbb{P}_{n, \theta, \boldsymbol{\beta}}\left(U^{c} \cap \widetilde{U}\right) \lesssim-n
$$

using (2.18). The desired conclusion follows on combining the last two displays.
(b) This follows on using Cauchy-Schwarz inequality to note that

$$
\begin{aligned}
\mathbb{E}_{n, \theta, \boldsymbol{\beta}}(W 1\{U\}-W 1\{\widetilde{U}\}) & \leq \sqrt{\mathbb{E}_{n, \theta, \boldsymbol{\beta}} W^{2}} \sqrt{\mathbb{P}_{n, \theta, \boldsymbol{\beta}}(U)+\mathbb{P}_{n, \theta, \boldsymbol{\beta}}(\widetilde{U})-2 \mathbb{P}_{n, \theta, \boldsymbol{\beta}}(U \cap \widetilde{U})} \\
& \leq \sqrt{\mathbb{P}_{n, \theta, \boldsymbol{\beta}}(U \Delta \widetilde{U})}
\end{aligned}
$$

## Chapter 3: Ising models on dense regular graphs

### 3.1 Introduction

The Ising Model is possibly the most well-known discrete graphical model, originating in statistical physics (see [28, 34, 42]), and henceforth studied in depth across several disciplines, including statistics and machine learning (c.f. $[3,7,18,19,26,40]$ 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 ([20, 23, 35, 42]). 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 3.2.2 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 (3.1)), we study limits of experiments in the sense of Lucien Le Cam ([36], see also [48]). 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 critical $(\theta=1)$ and high $(\theta<1)$ temperature regimes. 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 asymp-
totic 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 ([17]), 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.

### 3.2 Main results

In this section we formally introduce the Ising model (section 3.2.1), and state our main results (section 3.2.6), 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 3.2.3. To obtain convergence in experiments, we require the sequence coupling matrices for the Ising model to converge in cut metric, a notion which we recall in section 3.2.2. We also introduce the estimators and test statistics that we study in sections 3.2.4 and 3.2.5 respectively. Section 3.2.7 illustrates our results with two concrete examples. Finally, section 3.2.8 discusses the main contributions of this paper, and possible avenues of future research.

### 3.2.1 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

$$
\begin{equation*}
\mathbb{P}_{\theta, Q_{n}}(\mathbf{X}=\mathbf{x}):=\exp \left(\frac{\theta}{2} \mathbf{x}^{T} Q_{n} \mathbf{x}-Z_{n}\left(\theta, Q_{n}\right)\right), \text { for } \mathbf{x} \in\{-1,1\}^{n} \tag{3.1}
\end{equation*}
$$

Here $Z\left(\theta, Q_{n}\right)$ is the $\log$ normalizing constant which makes (3.1) into a probability distribution.

One of the major challenges in analyzing Ising models is that $Z_{n}\left(\theta, Q_{n}\right)$ 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

$$
\begin{equation*}
\sum_{j=1}^{n} Q_{n}(i, j)=1, \text { for all } i \in[n] \tag{3.2}
\end{equation*}
$$

- There exists a finite positive constant $C_{w}$ free of $n$ such that

$$
\begin{equation*}
\max _{i, j \in[n]} Q_{n}(i, j)<\frac{C_{w}}{n} \tag{3.3}
\end{equation*}
$$

- The Frobenius norm of $Q_{n}$ converges, i.e.

$$
\begin{equation*}
\left\|Q_{n}\right\|_{F}:=\sqrt{\sum_{i, j=1}^{n} Q_{n}^{2}(i, j)} \rightarrow \gamma \tag{3.4}
\end{equation*}
$$

### 3.2.2 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 $[9,10,38]$.

By a graphon, we will mean a symmetric bounded measurable function $f:[0,1]^{2} \rightarrow\left[0, C_{w}\right]$. Let $\mathcal{W}$ denote the space of all graphons. Equip the space $\mathcal{W}$ by the cut distance, defined by

$$
d_{\square}(f, g):=\sup _{S, T \subset[0,1]}\left|\int_{S \times T}(f(x, y)-g(x, y)) d x d y\right| .
$$

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

$$
f \sim g \quad \text { if } \quad f(x, y) \stackrel{\text { a.s. }}{=} g_{\iota}(x, y):=g(\iota(x), \iota(y)), \text { for some } \iota \in \mathcal{M} .
$$

Let $\tilde{\mathcal{W}}$ denote the quotient space $\mathcal{W} / \sim$ under the above equivalence relation. Equip $\tilde{\mathcal{W}}$ with the cut metric, defined as follows:

$$
\delta_{\square}(\tilde{f}, \tilde{g}):=\inf _{\iota \in \mathcal{M}} d_{\square}\left(f, g_{\iota}\right)=\inf _{\iota \in \mathcal{M}} d_{\square}\left(f_{\iota}, g\right)=\inf _{\iota_{1}, \iota_{2} \in \mathcal{M}} d_{\square}\left(f_{\iota_{1}}, g_{\iota_{2}}\right) .
$$

Then it follows from [9] that $\delta_{\square}$ is well defined, and ( $\left.\tilde{\mathcal{W}}, \delta_{\square}\right)$ is a compact metric space. We say a sequence of graphons $\left\{f_{n}\right\}_{n \geq 1}$ converge to a graphon $f$ in cut metric, if

$$
\delta_{\square}\left(f_{n}, f\right) \rightarrow 0 .
$$

Given a symmetric $n \times n$ matrix $A_{n}$ with 0 on the diagonal, define a corresponding graphon $f^{A_{n}}$ by setting

$$
f^{A_{n}}(x, y):=A_{\lceil n x\rceil,\lceil n y\rceil} .
$$

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

$$
\delta_{\square}\left(f^{A_{n}}, f\right) \rightarrow 0 .
$$

Given $f \in \mathcal{W}$, define a Hilbert-Schmidt operator $T_{f}$ from $L_{2}[0,1]$ to $L_{2}[0,1]$ by setting

$$
T_{f}(g)(\cdot):=\int_{[0,1]} f(\cdot, y) g(y) d y
$$

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

$$
\left|\lambda_{1}\right| \geq\left|\lambda_{2}\right| \geq \cdots
$$

Throughout the paper we assume that the sequence of matrices $\left\{n Q_{n}\right\}_{n \geq 1}$ converge in cut metric to a graphon $f \in \mathcal{W}$ such that $\lambda_{2}<\lambda_{1}\left(\right.$ can be $\left.-\lambda_{1}\right)$, i.e.

$$
\begin{equation*}
\delta_{\square}\left(f^{n Q_{n}}, f\right) \rightarrow 0, \quad \lambda_{2}<\lambda_{1} . \tag{3.5}
\end{equation*}
$$

A similar spectral gap assumption on the eigenvalue was utilized in [19, p. 1.7], where the authors study universal limiting distribution of $\overline{\mathbf{X}}$ for Ising models on dense regular graphs. In particular, it was shown in [19, Ex 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 3.2.1. To understand the exact nature of the assumptions (3.2), (3.3), (3.4) and (3.5), 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 $Q_{n}=\frac{1}{d} G_{n}$, where $\bar{d}:=\frac{1}{n} \sum_{i=1}^{n} d_{i}$ is the average degree of the graph $G_{n}$, and $\left(d_{1}, \ldots, d_{n}\right)$ is the labeled degree sequence of the graph $G_{n}$. This particular choice ensures that the resulting model is non trivial (see [2, Cor 1.2]). For the above choice, the assumptions (3.2) and (3.3) 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 (3.4) follows from (3.5), as

$$
\sum_{i, j=1}^{n} Q_{n}^{2}(i, j)=\frac{1}{\bar{d}_{n}^{2}} \sum_{i, j=1}^{n} G_{n}(i, j)=\frac{n}{\bar{d}_{n}} \rightarrow \frac{1}{\int_{[0,1]} f(x, y) d x d y} .
$$

Thus our results apply to Ising models on dense regular graphs converging in cut metric (i.e. under (3.2), (3.3) and (3.5)).

### 3.2.3 Limits of Experiments

In this section we briefly introduce the notion of convergence of experiments. For more details we refer the reader to [48, Chapter 9].

Suppose that for each $1 \leq n \leq \infty$ we have a measure space $\left(\mathcal{X}_{n}, \mathscr{F}_{n}\right)$. Let $\left\{\mathbb{P}_{h, n}, h \in H\right\}$ be a collection of probability measures on $\left(\mathcal{X}_{n}, \mathcal{F}_{n}\right)$. We say that $\left\{\mathbb{P}_{h, n}, h \in H\right\}$ converges to $\left\{P_{h, \infty}, h \in H\right\}$ in the sense of limits of experiments, if for every finite subset $I$ of $H$ and $h_{0} \in H$ we have

$$
\begin{equation*}
\left(\frac{d \mathbb{P}_{h, n}}{d \mathbb{P}_{h_{0}, n}}(\mathbf{X})\right)_{h \in I} \xrightarrow{d, \mathbb{P}_{h_{0}, n}}\left(\frac{d \mathbb{P}_{h, \infty}}{d \mathbb{P}_{h_{0}, \infty}}(Y)\right)_{h \in I} . \tag{3.6}
\end{equation*}
$$

The RHS of (3.6) is also called the likelihood ratio process with base $h_{0}$. We will use the short hand 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 $\mathbb{P}_{h, \infty}=N\left(h, \tau^{2}\right)$ for some $\tau>0$, then we say the collection of experiments $\left\{\mathbb{P}_{h, n}, h \in H\right\}$ is locally asymptotically normal, or LAN. For examples of both LAN and non LAN experiments, see [48, Chapters 7, 9]).

### 3.2.4 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 $\hat{\theta}_{n}^{M L E}$ is defined as

$$
\arg \sup _{\theta \in \mathbb{R}}\left\{\frac{\theta}{2} \mathbf{X}^{T} Q_{n} \mathbf{X}-Z_{n}\left(\theta, Q_{n}\right)\right\},
$$

provided the supremum is attained uniquely. Since the function in the above display is strictly concave in $\theta$, it follows that $\hat{\theta}_{n}^{M L E}$, if it exists, is the unique solution to the equation

$$
\frac{1}{2} \mathbf{X}^{T} Q_{n} \mathbf{X}=Z_{n}^{\prime}\left(\theta, Q_{n}\right)=\left.\frac{\partial Z_{n}\left(\theta, Q_{n}\right)}{\partial \theta}\right|_{\theta=\hat{\theta}_{n}^{M L E}}=\frac{1}{2} \mathbb{E}_{\mathbb{P}_{\theta, Q_{n}}} \mathbf{X}^{T} Q_{n} \mathbf{X}
$$

in $\theta \in \mathbb{R}$. In the special case of the Curie-Weiss model, the asymptotics of $\hat{\theta}_{n}^{M L E}$ was studied in [17], 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 $\hat{\theta}_{n}^{M L E}$ 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 $Z_{n}\left(\theta, Q_{n}\right)$ is computationally intractable. To bypass this, Besag introduced the maximum pseudo-likelihood estimator ([5, 4]) for spatial interaction models. Below we define the maximum pseudo-likelihood estimator $\hat{\theta}_{n}^{M P L E}$.

Given $\mathbf{X} \sim \mathbb{P}_{\theta, Q_{n}}$, we have

$$
\begin{equation*}
\mathbb{P}_{\theta, Q_{n}}\left(X_{i}=x_{i} \mid X_{j} \text { for all } j \neq i\right):=\frac{\exp \left(\theta t_{i} x_{i}\right)}{\exp \left(\theta t_{i}\right)+\exp \left(-\theta t_{i}\right)}, \tag{3.7}
\end{equation*}
$$

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:

$$
P L_{n}(\theta):=\prod_{i=1}^{n} \mathbb{P}_{\theta, Q_{n}}\left(X_{i} \mid X_{j} \text { for all } j \neq i\right)=\frac{\exp \left(\theta \sum_{i=1}^{n} X_{i} t_{i}\right)}{2^{n} \prod_{i=1}^{n} \cosh \left(\theta t_{i}\right)} .
$$

The maximum pseudo-likelihood estimator $\hat{\theta}_{n}^{M P L E}$ is defined as

$$
\arg \sup _{\theta \in \mathbb{R}} \log P L_{n}(\theta)=\arg \sup _{\theta \in \mathbb{R}}\left\{\theta \sum_{i=1}^{n} X_{i} t_{i}-\sum_{i=1}^{n} \log \cosh \left(\theta t_{i}\right)\right\}
$$

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

$$
\begin{equation*}
\mathbf{X}^{T} Q_{n} \mathbf{X}=\sum_{i=1}^{n} t_{i} \tanh \left(\theta t_{i}\right) \tag{3.8}
\end{equation*}
$$

The above equation (3.8) does not involve the intractable function $Z_{n}\left(\theta, Q_{n}\right)$, 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 pseudolikelihood estimator for Ising models was established in [7, 14, 26], but the question of asymptotic distribution remained open.

Proposition 3.3.1 gives an exact characterization for the existence of the MLE and the MPLE, and shows that the conditions hold with probability tending to 1 .

### 3.2.5 Hypothesis Testing for $\theta$

Given $\mathbf{X} \sim \mathbb{P}_{\theta, Q_{n}}$ for some $\theta>0$, suppose we want to test

$$
\begin{equation*}
\mathcal{H}_{0}: \theta=\theta_{0} \quad \text { vs } \quad \mathcal{H}_{1}: \theta>\theta_{n} \tag{3.9}
\end{equation*}
$$

for a positive real number sequence $\left\{\theta_{n}\right\}_{n=1}^{\infty}$, 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 $\overline{\mathbf{X}}$ denote the sample mean of $\mathbf{X}$, and let

$$
\psi_{n}(\mathbf{X})= \begin{cases}1 & \text { If } n \overline{\mathbf{X}}^{2}>K_{n}(\alpha)  \tag{3.10}\\ 0 & \text { Otherwise }\end{cases}
$$

where $K_{n}(\alpha)$ is chosen such that $\psi_{n}$ has level $\alpha$. Let $\beta_{M S}$ 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 $\mathbf{X}^{T} Q_{n} \mathbf{X}$, and is given by

$$
\psi_{n}(\mathbf{X})= \begin{cases}1 & \text { If } \mathbf{X}^{T} Q_{n} \mathbf{X}>K_{n}(\alpha)  \tag{3.11}\\ 0 & \text { Otherwise }\end{cases}
$$

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 $\hat{\theta}_{n}^{M L E}$. Let $\beta_{N P}$ denote the limiting power of the above test (provided the limit exists).

- Pseudo-likelihood Test (PL-test) With $\hat{\theta}_{n}^{M P L E}$ denoting the pseudo-likelihood estimator, define the pseudo-likelihood test by setting

$$
\psi_{n}(\mathbf{X})= \begin{cases}1 & \text { If } \hat{\theta}_{n}^{M P L E}>K_{n}(\alpha)  \tag{3.12}\\ 0 & \text { Otherwise }\end{cases}
$$

where $K_{n}(\alpha)$ is chosen such that the above test has level $\alpha$. Let $\beta_{P L}$ denote the limiting
power of the above test (provided the limit exists).

Remark 3.2.2. 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.

### 3.2.6 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 [19, Lemma 1.1]:

Proposition 3.2.1. For any $\theta>0$ consider the fixed point equation

$$
\begin{equation*}
w(\theta, x)=0, \quad \text { where } \quad w(\theta, x):=x-\tanh (\theta x) \tag{3.13}
\end{equation*}
$$

(a) If $\theta>1$, then (3.13) 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 (3.13) 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 (3.13) has a unique root $m(\theta)=0$, and $\left.\frac{\partial w(\theta, x)}{\partial x}\right|_{x=0}>0$.

Definition 3.2.1. Using Proposition 3.2.1, 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 3.2.1. 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 $(0, \infty)$ into three distinct domains:

- Low Temperature Regime: $\Theta_{1}:=(1, \infty)$;
- Critical point: $\Theta_{2}=1$;
- High Temperature Regime: $\Theta_{3}:=(0,1)$.

The nomenclature of the three domains is inspired from statistical physics terminology (see for e.g. [23]).

For any $\theta \in \Theta_{1} \cup \Theta_{3}$, define the positive real $\sigma^{2}(\theta)$ by setting

$$
\begin{equation*}
\sigma^{2}(\theta):=\frac{1-m^{2}(\theta)}{1-\theta\left(1-m^{2}(\theta)\right)} \tag{3.14}
\end{equation*}
$$

Note that $\sigma^{2}(\theta)$ is well defined by Proposition 3.2.1, as $1-\left(1-m^{2}(\theta)\right) \theta=w_{\theta}^{\prime}(m(\theta))>0$. In particular, we have $\sigma^{2}(\theta)=\frac{1}{1-\theta}$ if $\theta \in \Theta_{3}$.

Definition 3.2.2. Given $h \in \mathbb{R}$ and $\theta_{0}>0$, define a positive sequence $\left\{\theta_{n}\right\}_{n \geq 1}$ (depending on $h, \theta_{0}$ ) by setting

$$
\begin{array}{rlr}
\theta_{n} & :=\theta_{0}+\frac{h}{\sqrt{n}} \quad \text { if } \theta_{0} \in \Theta_{1} \cup \Theta_{2},  \tag{3.15}\\
& =\theta_{0}+h & \text { if } \theta \in \Theta_{3} .
\end{array}
$$

We will omit the dependence of $\theta_{0}, h$, since it will be clear from the context.

Definition 3.2.3. 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

$$
\Psi_{\zeta}(p):=\inf \left\{t \in \mathbb{R}: F_{\zeta}(t) \geq p\right\}
$$

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 3.2.4. Let $\kappa:=\gamma^{2}-\|f\|_{2}^{2}$, where $\gamma, f$ are as in (3.4) and (3.5) respectively.

Let $W^{*} \sim N(0,2 \kappa)$ and $\left\{Y_{j}\right\}_{j \geq 2} \stackrel{\text { iid }}{\sim} \chi_{1}^{2}$ be mutually independent. For any $\theta \in \Theta$, define two
random variables $S_{\theta}$ and $T_{\theta}$ by setting

$$
\begin{align*}
& S_{\theta}:=\left(1-m^{2}(\theta)\right)\left[\sum_{j=2}^{\infty} \lambda_{j}\left(\frac{Y_{j}}{1-\theta\left(1-m^{2}(\theta)\right) \lambda_{j}}-1\right)-1+\left(1-m^{2}(\theta)\right) \theta \kappa+W^{*}\right]  \tag{3.16}\\
& T_{\theta}:=\left(1-m^{2}(\theta)\right)\left[\sum_{j=2}^{\infty} \frac{\lambda_{j}^{2} Y_{j}}{1-\theta\left(1-m^{2}(\theta)\right) \lambda_{j}}+\kappa\right] \tag{3.17}
\end{align*}
$$

where $m(\theta)$ is as in definition 3.2.1. Here the infinite sums in the limiting distributions converge in $L_{2}$ (see Lemma 3.6.2).

We now state the results for each of the domains separately.

## 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 3.2.2. Suppose $\mathbf{X} \sim \mathbb{P}_{\theta, Q_{n}}$ with $Q_{n}$ satisfying (3.2), (3.3), (3.4) and (3.5) for some $C_{W}, \kappa \in$ $(0, \infty)$ and $f \in \mathcal{W}$. Then with $R\left(\theta_{0}\right):=m^{2}\left(\theta_{0}\right) \sigma^{2}\left(\theta_{0}\right)$, the following conclusions hold:
(a) We have

$$
\left\{\mathbb{P}_{\theta_{0}+h n^{-1 / 2}, Q_{n}}\right\}_{h \in \mathbb{R}} \xrightarrow{\operatorname{Exp}}\left\{N\left(h, R\left(\theta_{0}\right)^{-1}\right)\right\}_{h \in \mathbb{R}} .
$$

(b) The MLE $\hat{\theta}_{n}^{M L E}$ and the MPLE $\hat{\theta}_{n}^{M P L E}$ (as defined in section 3.2.4) exist with probability tending to 1 , and have a common asymptotic distribution, given by

$$
\sqrt{n}\left(\tilde{\theta}_{n}^{M L E}-\theta_{0}\right) \xrightarrow{d} N\left(0, R\left(\theta_{0}\right)^{-1}\right), \quad \sqrt{n}\left(\hat{\theta}_{n}^{M P L E}-\theta_{0}\right) \xrightarrow{d} N\left(0, R\left(\theta_{0}\right)^{-1}\right) .
$$

(c) With $\theta_{n}=\theta_{0}+\frac{h}{\sqrt{n}}$ for some $h>0$, and $\beta_{M S}, \beta_{N P}$ and $\beta_{P L}$ as defined in section 3.2.5, we have

$$
\beta_{N P}=\beta_{P L}=\beta_{M S}=\mathbb{P}\left(N(0,1)>z_{\alpha}-h \sqrt{R\left(\theta_{0}\right)}\right),
$$

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

Remark 3.2.3. 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 $\hat{\theta}_{n}^{M L E}$ is asymptotically optimal. Perhaps surprisingly, part (b) above shows that $\hat{\theta}_{n}^{M P L E}$ (which requires significantly less computational resources) is also asymptotically optimal. Carrying this through, it is shown in part (c) that the tests based on $\hat{\theta}_{n}^{M L E}$ and $\hat{\theta}_{n}^{M P L E}$ 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 $Q_{n}$.

## 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 3.2.5. Let $\left\{\mathbb{H}_{h}(\cdot), h \in \mathbb{R}\right\}$ be a family of probability distributions on $\mathbb{R}$ parametrized by $h$, with density function

$$
\begin{equation*}
p_{h}(u)=\exp \left(-\frac{1}{12} u^{4}+\frac{1}{2} h u^{2}-F(h)\right) . \tag{3.18}
\end{equation*}
$$

Here

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

is the log normalizing constant, which makes $p_{h}($.$) into a density.$

Let $W^{*} \sim N(0,2 \kappa), U_{1, h} \sim \mathbb{H}_{h}$, and $\left(S_{1}, T_{1}\right)$ be mutually independent, where $\left(S_{1}, T_{1}\right)$ are as defined in (3.16) and (3.17) respectively. Set

$$
\begin{equation*}
V_{1, h}=\frac{1}{3} U_{1, h}^{2}+\frac{S_{1}-T_{1}}{U_{1, h}^{2}} . \tag{3.19}
\end{equation*}
$$

Theorem 3.2.3. Suppose $\mathbf{X} \sim \mathbb{P}_{\theta, Q_{n}}$ with $Q_{n}$ satisfying (3.2), (3.3), (3.4) and (3.5) for some $C_{W}, \kappa \in$ $(0, \infty)$ and $f \in \mathcal{W}$. Then the following conclusions hold:
(a) We have

$$
\begin{equation*}
\left\{\mathbb{P}_{\theta_{0}+h n^{-1 / 2}, Q_{n}}\right\}_{h \in \mathbb{R}} \xrightarrow{\operatorname{Exp}}\left\{\mathbb{H}_{h}\right\}_{h \in \mathbb{R}} . \tag{3.20}
\end{equation*}
$$

(b) The MLE $\hat{\theta}_{n}^{M L E}$ and MPLE $\hat{\theta}_{n}^{M P L E}$ as defined in section 3.2.4 exist with probability tending to 1 , and satisfy

$$
\begin{gather*}
\mathbb{P}_{1, Q_{n}}\left(\sqrt{n}\left(\hat{\theta}_{n}^{M L E}-1\right) \leq h\right) \longrightarrow \mathbb{P}\left(U_{1,0}^{2} \leq \mathbb{E} U_{1, h}^{2}\right),  \tag{3.21}\\
\sqrt{n}\left(\hat{\theta}_{n}^{M P L E}-1\right) \xrightarrow{d, \mathbb{P}_{1, Q_{n}}} V_{1,0}, \tag{3.22}
\end{gather*}
$$

(c) With $\theta_{n}=1+\frac{h}{\sqrt{n}}$ for some $h>0$, and $\beta_{M S}, \beta_{N P}$ and $\beta_{P L}$ as defined in section 3.2.5, we have

$$
\begin{gather*}
\beta_{N P}=\beta_{M S}=2 \mathbb{P}\left(U_{1, h}>\Psi_{\mathbb{H}, 0}(1-\alpha / 2)\right),  \tag{3.23}\\
\beta_{P L}=\mathbb{P}\left(V_{1, h}>\Psi_{V_{1,0}}(1-\alpha)\right) . \tag{3.24}
\end{gather*}
$$

Remark 3.2.4. Thus, similar to Theorem 3.2.2, the contiguous alternatives obtained in Theorem 3.2.3 in the critical regime is also of size $O\left(\frac{1}{\sqrt{n}}\right)$. 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 continuous to be asymptotically optimal, and thus provides a computationally efficient $Q_{n}$ agnostic solution to our testing problem. We will show in section 3.2.7 that $\beta_{M S}=\beta_{P L}$ for the Curie-Weiss model.

## High temperature regime $\theta \in \Theta_{1}$

When $\theta_{0} \in(0,1)$ (high temperature regime), the behavior is again different. In this regime the contiguous alternatives are of the form $\left\{\mathbb{P}_{\theta, Q_{n}}\right\}_{\theta \in(0,1)}$, and consistent estimation of the parameter $\theta$ is no longer possible. In this regime we do not study asymptotic distribution of the estimators, but only look at the asymptotic performance of the tests regarding $\theta$. To describe the limit experiments, and limiting behavior of tests, we require the following definitions:

Definition 3.2.6. For any $\theta_{0} \in(0,1)$ and $j \in \mathbb{Z}_{+}:=\{1,2, \ldots\}$ we define a probability measure $v_{\theta_{0}, j}$ on $\mathbb{R}$, by setting

$$
v_{\theta_{0}, j}:=N\left(0, \frac{1}{1-\lambda_{j} \theta_{0}}\right) \text { for } j \geq 1, \quad v_{\theta_{0}, 0}:=N\left(\sqrt{\frac{\kappa}{2}} \theta_{0}, 1\right)
$$

Let $v_{\theta_{0}}=\bigotimes_{j \in \mathbb{Z}_{+}} v_{\theta, j}$ be the product probability measure on $\mathbb{R}^{\mathbb{Z}^{+}}$.
Definition 3.2.7. Let $W^{*} \sim N(0,2 \kappa)$ and $\left\{Y_{i}\right\}_{i \geq 1} \stackrel{i i d}{\sim} \chi_{1}^{2}$ be mutually independent. For $\theta \in \Theta_{3}$, define

$$
\begin{align*}
U_{\theta} & :=\sum_{j=1}^{\infty} \lambda_{j}\left(\frac{Y_{j}}{1-\theta \lambda_{j}}-1\right)+\theta \kappa+W^{*}  \tag{3.25}\\
V_{\theta} & :=\frac{U_{\theta}}{\sum_{j=1}^{\infty} \frac{\lambda_{j}^{2}}{1-\theta \lambda_{j}} Y_{j}+\kappa} . \tag{3.26}
\end{align*}
$$

The infinite series in the above definitions converge in $L_{2}$ (see Lemma 3.6.2).

Theorem 3.2.4. Suppose $\mathbf{X} \sim \mathbb{P}_{\theta_{0}, Q_{n}}$ with $Q_{n}$ satisfying (3.2), (3.3), (3.4) and (3.5) for some $C_{W}, \kappa \in(0, \infty)$ and $f \in \mathcal{W}$. Then the following conclusions hold:
(a) We have

$$
\begin{equation*}
\left\{\mathbb{P}_{\theta_{0}+h, Q_{n}}\right\}_{h \in\left(-\theta_{0}, 1-\theta_{0}\right)} \xrightarrow{\operatorname{Exp}}\left\{v_{\theta_{0}+h}\right\}_{h \in\left(-\theta_{0}, 1-\theta_{0}\right)} . \tag{3.27}
\end{equation*}
$$

Here
(b) With $\theta_{n}=\theta_{0}+h$ for some $h>0$, and $\beta_{M S}, \beta_{N P}$ and $\beta_{P L}$ as defined in section 3.2.5, we have

$$
\begin{align*}
& \beta_{N P}=\mathbb{P}\left(U_{\theta_{0}+h}>\Psi_{U_{\theta_{0}}}(1-\alpha)\right),  \tag{3.28}\\
& \beta_{P L}=\mathbb{P}\left(V_{\theta_{0}+h}>\Psi_{V_{\theta_{0}}}(1-\alpha)\right),  \tag{3.29}\\
& \beta_{M S}=2 \mathbb{P}\left(N(0,1)>z_{\alpha / 2} \sqrt{\frac{1-\theta_{0}-h}{1-\theta_{0}}}\right) . \tag{3.30}
\end{align*}
$$

Remark 3.2.5. The non-existence of consistent estimators for $\theta$ in $\Theta_{3}=(0,1)$ for the above class of Ising models was first shown in [7, Theorem 2.3]. The description of the limiting experiment in this case is much more complicated, and requires us to work on probability measures on $\mathbb{R}^{\mathbb{Z}^{+}}$, and needs detailed knowledge of the eigenvalues of $f$. On the other hand, in the other two regimes $\Theta_{1}$ and $\Theta_{2}$ did not require any knowledge of finer properties of $f$. Thus the limiting experiment is universal for a sequence of Ising models on dense regular graphs in regimes $\Theta_{1} \cup \Theta_{2}$, but not in $\Theta_{3}$. Also, in general it seems unclear as to which of the two asymptotic powers $\beta_{M S}$ and $\beta_{P L}$ are larger in this case, as their formulas are not very amenable.

### 3.2.7 Examples

To illustrate our results, we will now apply our main theorems and simulation results on two examples, the Curie-Weiss model (where $G_{n}$ is the complete graph), and the Ising model on the complete bi-partite graph $K_{n / 2, n / 2}$.

Definition 3.2.8. - Suppose $\mathbf{X} \sim \mathbb{P}_{\theta, Q_{n}}(\cdot)$ for some $\theta>0$, where

$$
\begin{equation*}
Q_{n}(i, j):=\frac{1}{n} \text { if } i \neq j \tag{3.31}
\end{equation*}
$$

Thus $Q_{n}$ is the scaled adjacency graph of the complete graph $K_{n}$, and in this case $\mathbb{P}_{\theta, Q_{n}}(\cdot)$ is called the Curie-Weiss model. The limiting graphon for $f_{n Q_{n}}$ is the constant function $f=1$, with only one non-zero eigenvalue $\lambda_{1}=1$.

- Suppose $\mathbf{X} \sim \mathbb{P}_{\theta, Q_{n}}(\cdot)$ for some $\theta>0$, where $n$ is even, and

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

Thus $Q_{n}$ is the scaled adjacency matrix of the complete bipartite graph $K_{n / 2, n / 2}$. We will refer to $\mathbb{P}_{\theta, Q_{n}}(\cdot)$ as the bipartite Ising model. The limiting graphon for $f_{n Q_{n}}$ is the piecewise constant function $f$ given by

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

This function has two non zero eigenvalues, $\lambda_{1}=1$ and $\lambda_{2}=-1$.

Table 3.2.7 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 $N\left(0, R\left(\theta_{0}\right)^{-1}\right)$, for both the Curie-Weiss model and the bipartite Ising model (see Theorem 3.2.2 part (b)). In fact, the same universal limit continues to hold for any sequence of graphs $G_{n}$ satisfying (3.2), (3.3), (3.4) and (3.5). At the critical regime $\Theta_{2}$, the asymptotic distribution of the two estimators are not the same, for either the CurieWeiss model (see figure 3.1a, or the bipartite Ising model (see figure 3.1b). The simulations in figures 3.1a and 3.1b use the limiting distribution obtained in Theorem 3.2.3 part (b). In the high temperature regime all estimators (and hence MLE and MPLE) are both inconsistent.

| 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 |
| High Temperature | not consistent | not consistent |

In a similar manner, table 3.2 .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 3.2a). In the high temperature regime, all tests have different power for the bipartite Ising model, with the PL test performing the worst (see figure $3.2 b)$.

| Table 2: Aymptotics powers of tests |  |  |
| :--- | :--- | :--- |
| Regimes of $\theta$ | Curie-Weiss Model | Bipartite Ising Model |
| Low Temperature | $\beta_{N P}=\beta_{M S}=\beta_{P L}$ | $\beta_{N P}=\beta_{M S}=\beta_{P L}$ |
| Critical Point | $\beta_{N P}=\beta_{M S}=\beta_{P L}$ | $\beta_{N P}=\beta_{M S}>\beta_{P L}$ |
| High Temperature | $\beta_{N P}=\beta_{M S}=\beta_{P L}$ | $\beta_{N P}>\beta_{M S}>\beta_{P L}$ |

### 3.2.8 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, is universal and non LAN in the critical regime, and is non-universal and non LAN in the high temperature 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


Figure 3.1: 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.


Figure 3.2: Both panels show the testing powers versus h under biparite Ising model, at the critical regime and the high temperature regime, respectively.
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 nonzero 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.

### 3.2.9 Outline

The rest of this paper is organized as follows:
In section 3.3, we give the proofs of our the main theorems. The proofs of the theorems in section 3.3 rely on key lemmas about the Ising model, the proofs of which are deferred to appendix A (section 3.5). The proofs of these lemmas, in turn, depend on a precise understanding of quadratic forms under the Curie-Weiss model. The proofs of these results on the Curie-Weiss model are deferred to appendix B (section 3.6). Some auxiliary results are proved in appendix C (section 3.7).

### 3.3 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 3.3.1. Let $\mathbf{X} \sim \mathbb{P}_{\theta_{n}, Q_{n}}$, where $\theta_{0}>0, h \in \mathbb{R}$, and $\theta_{n}$ is as defined in (3.15). Assume that the matrix $Q_{n}$ satisfies (3.2), (3.3), (3.4) and (3.5) 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 (3.16) and (3.17) respectively. Then, setting $B_{n}:=Q_{n}-\frac{1}{n} \mathbf{1 1}^{T}$, the following conclusions hold:
(a) If $\theta_{0} \in \Theta_{1}$, then conditional on the set $\overline{\mathbf{X}}>0$ we have

$$
\left[\sqrt{n}\left(\overline{\mathbf{X}}-m\left(\theta_{n}\right)\right), \mathbf{X}^{T} B_{n} \mathbf{X}, \mathbf{X}^{T} B_{n}^{2} \mathbf{X}\right] \xrightarrow{d}\left[W_{\theta_{0}}, S_{\theta_{0}}, T_{\theta_{0}}\right] .
$$

where $m($.$) is as in definition 3.2.1, and W_{\theta_{0}} \sim N\left(0, \sigma^{2}\left(\theta_{0}\right)\right)$ is independent of $\left(S_{\theta_{0}}, T_{\theta_{0}}\right)$.
(b) If $\theta_{0} \in \Theta_{2}$, then

$$
\left[n^{1 / 4} \overline{\mathbf{X}}, \mathbf{X}^{T} B_{n} \mathbf{X}, \mathbf{X}^{T} B_{n}^{2} \mathbf{X}\right] \xrightarrow{d}\left[U_{h}, S_{\theta_{0}}, T_{\theta_{0}}\right],
$$

where $U_{1, h} \sim \mathbb{H}_{h}$ (see (3.18)) is independent of $\left(S_{\theta_{0}}, T_{\theta_{0}}\right)$.
(c) If $\theta_{0} \in \Theta_{3}$, then

$$
\left[\sqrt{n} \overline{\mathbf{X}}, \mathbf{X}^{T} B_{n} \mathbf{X}, \mathbf{X}^{T} B_{n}^{2} \mathbf{X}\right] \xrightarrow{d}\left[W_{\theta_{0}}, S_{\theta_{0}}, T_{\theta_{0}}\right],
$$

where $W_{\theta_{0}} \sim N\left(0, \sigma^{2}\left(\theta_{0}\right)\right)$ is independent of $\left(S_{\theta_{0}}, T_{\theta_{0}}\right)$.

Our second lemma gives very precise asymptotics for the normalizing constant of Ising models.
Lemma 3.3.2. . Let $\theta_{0}>0, h \in \mathbb{R}$, and let $\theta_{n}$ be as defined in (3.15). Assume that the matrix $Q_{n}$ satisfies (3.2), (3.3), (3.4) and (3.5) for some $C_{W}, \kappa \in(0, \infty)$ and $f \in \mathcal{W}$.
(a) If $\theta_{0}>1$ then we have

$$
\lim _{n \rightarrow \infty}\left\{Z_{n}\left(\theta_{n}, Q_{n}\right)-Z_{n}\left(\theta_{0}, Q_{n}\right)-\frac{1}{2} \sqrt{n} h m^{2}\left(\theta_{0}\right)\right\}=\frac{R\left(\theta_{0}\right) h^{2}}{2}
$$

(b) If $\theta_{0}=1$, then we have

$$
\lim _{n \rightarrow \infty}\left\{Z_{n}\left(\theta_{n}, Q_{n}\right)-Z_{n}\left(\theta_{0}, Q_{n}\right)\right\}=F(h)-F(0)
$$

where $F(h)$ is as defined in (3.18).
(c) If $\theta_{0} \in(0,1)$ and $\theta_{0}+h \in(0,1)$ then we have

$$
\lim _{n \rightarrow \infty}\left\{Z_{n}\left(\theta_{n}, Q_{n}\right)-Z_{n}\left(\theta_{n}, Q_{n}\right)\right\}=\frac{\kappa}{4} h^{2}+\frac{\kappa \theta_{0} h}{2}-\frac{1}{2} \sum_{j=1}^{\infty}\left(\lambda_{j}-\log \frac{1-\theta_{0} \lambda_{j}}{1-\left(\theta_{0}+h\right) \lambda_{j}}\right)
$$

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

Proposition 3.3.1. (a) Let

$$
a_{n}:=\min _{\mathbf{x} \in\{-1,1\}^{n}} \mathbf{x}^{T} Q_{n} \mathbf{x} \leq \max _{\mathbf{x} \in\{-1,1\}^{n}} \mathbf{x}^{T} Q_{n} \mathbf{x}=: b_{n}
$$

Then the MLE exists in $\mathbb{R}$ iff

$$
a_{n}<\mathbf{X}^{T} Q_{n} \mathbf{X}<b_{n} .
$$

(b) In particular, the MLE exists with probability tending to 1 for all $\theta_{0} \in \Theta$.
(c) Let

$$
S=S(\mathbf{X}):=\left\{i: t_{i} \neq 0\right\} .
$$

Then the MPLE exists in $\mathbb{R}$ iff neither of the following two events happen:

- $\mathbf{X}_{i}=1$ for all $i \in S$.
- $\mathbf{X}_{i}=-1$ for all $i \in S$.
(d) In particular, the MPLE exists with probability tending to 1 for all $\theta_{0} \in \Theta$.

Our final result is a calculus proposition which computes derivatives of the function $m$ (.) defined in 3.2.1.

Proposition 3.3.2. Let $m():.(1, \infty) \mapsto(0,1)$ be as in definition 3.2.1. Then with $\theta_{n}=\theta_{0}+\frac{h}{\sqrt{n}}$ for some $\theta_{0} \in \Theta_{3}$ and $h>0$ we have

$$
\lim _{n \rightarrow \infty} \frac{m\left(\theta_{n}\right)-m\left(\theta_{0}\right)}{\frac{h}{\sqrt{n}}}=m\left(\theta_{0}\right) \sigma^{2}\left(\theta_{0}\right)
$$

### 3.3.1 Proof of Theorem 3.2.2

Before we begin, we use Lemma 3.3.1 parts (a), (b), (c) to conclude that in all regimes of $\left(\theta_{0}, h\right)$ we have $\mathbf{X}^{T} B_{n} \mathbf{X}=O_{p}(1)$ under $\mathbb{P}_{\theta_{0}+h n^{-1 / 2}, Q_{n}}$, and so

$$
\begin{equation*}
\mathbf{X}^{T} Q_{n} \mathbf{X}=n \overline{\mathbf{X}}^{2}+\mathbf{X}^{T} B_{n} \mathbf{X}=n \overline{\mathbf{X}}^{2}+O_{p}(1) . \tag{3.32}
\end{equation*}
$$

Proof of Theorem 3.2.2.

## Part (a):

Let $I=\left\{h_{1}, \ldots, h_{k}\right\}$ 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

$$
\begin{equation*}
\log \frac{d \mathbb{P}_{\theta_{n, i}, Q_{n}}(\mathbf{X})}{d \mathbb{P}_{\theta_{0}, Q_{n}}(\mathbf{X})}=Z_{n}\left(\theta_{0}, Q_{n}\right)-Z_{n}\left(\theta_{n, i}, Q_{n}\right)+\frac{h_{i} \sqrt{n} m^{2}\left(\theta_{0}\right)}{2}+\frac{h_{i}}{2}\left(\frac{\mathbf{X}^{T} Q_{n} \mathbf{X}}{\sqrt{n}}-\sqrt{n} m^{2}\left(\theta_{0}\right)\right) \tag{3.33}
\end{equation*}
$$

By part (a) of Lemma 3.3.2 we have

$$
Z_{n}\left(\theta_{n, i}, Q_{n}\right)-Z_{n}\left(\theta_{0}, Q_{n}\right)-\frac{h_{i} \sqrt{n} m^{2}\left(\theta_{0}\right)}{2} \rightarrow \frac{R\left(\theta_{0}\right) h_{i}^{2}}{2}
$$

Also, using (3.32), under $\mathbb{P}_{\theta_{0}, Q_{n}}$ we have

$$
\begin{equation*}
\frac{1}{\sqrt{n}} \mathbf{X}^{T} Q_{n} \mathbf{X}-\sqrt{n} m^{2}\left(\theta_{0}\right)=\sqrt{n}\left(\overline{\mathbf{X}}^{2}-m^{2}\left(\theta_{0}\right)\right)+O_{p}\left(\frac{1}{\sqrt{n}}\right) \xrightarrow{d} 2 m\left(\theta_{0}\right) W_{\theta_{0}}, \tag{3.34}
\end{equation*}
$$

where the last step we use part (a) of Lemma 3.3.1 to conclude that under $\mathbb{P}_{\theta_{0}, Q_{n}}$ we have

$$
\begin{equation*}
\sqrt{n}\left(\overline{\mathbf{X}}^{2}-m^{2}\left(\theta_{0}\right)\right)=\sqrt{n}\left(\overline{\mathbf{X}}-m\left(\theta_{0}\right)\right)\left(\overline{\mathbf{X}}+m\left(\theta_{0}\right)\right) \xrightarrow{d} 2 m\left(\theta_{0}\right) W_{\theta_{0}}, \tag{3.35}
\end{equation*}
$$

where $W_{\theta_{0}} \sim N\left(0, \sigma^{2}\left(\theta_{0}\right)\right)$. The above calculation gives that under $\mathbb{P}_{\theta_{0}, Q_{n}}$ we have

$$
\left[\log \frac{d \mathbb{P}_{\theta_{n, i}, Q_{n}}(\mathbf{X})}{d \mathbb{P}_{\theta_{0}, Q_{n}}(\mathbf{X})}\right]_{1 \leq i \leq k} \xrightarrow{d}\left[-\frac{R\left(\theta_{0}\right) h_{i}^{2}}{2}+m\left(\theta_{0}\right) h_{i} W_{\theta_{0}}\right]_{1 \leq i \leq k} .
$$

Also note that if $W \sim N\left(0, R\left(\theta_{0}\right)^{-1}\right)$ then

$$
\begin{aligned}
{\left[\log \frac{d N\left(h_{i}, R\left(\theta_{0}\right)^{-1}\right)}{d N\left(0, R\left(\theta_{0}\right)^{-1}\right)}(W)\right]_{1 \leq i \leq k} } & =\left[-\frac{R\left(\theta_{0}\right) h_{i}^{2}}{2}+R\left(\theta_{0}\right) W h_{i}\right]_{1 \leq i \leq k} \\
& \stackrel{d}{=}\left[-\frac{R\left(\theta_{0}\right) h_{i}^{2}}{2}+m\left(\theta_{0}\right) h_{i} W_{\theta_{0}}\right]_{1 \leq i \leq k}
\end{aligned}
$$

It then follows from the last two displays that under $\mathbb{P}_{\theta_{0}, Q_{n}}$ we have

$$
\left[\log \frac{d \mathbb{P}_{\theta_{n, i}, Q_{n}}(\mathbf{X})}{d \mathbb{P}_{\theta_{0}, Q_{n}}(\mathbf{X})}\right]_{1 \leq i \leq k} \xrightarrow{d}\left[\log \frac{d N\left(h_{i}, R\left(\theta_{0}\right)^{-1}\right)}{d N\left(0, R\left(\theta_{0}\right)^{-1}\right)}(W)\right]_{1 \leq i \leq k},
$$

which verifies convergence of experiments.

## Part (b):

## - MLE

Suppose $\mathbf{X} \sim \mathbb{P}_{\theta, Q_{n}}$. Using Proposition 3.3.1 part (b), it follows that the MLE exists with probability tending to 1 . Also, the MLE $\hat{\theta}_{n}^{M L E}$ satisfies the equation

$$
Z_{n}^{\prime}\left(\hat{\theta}_{n}^{M L E}, Q_{n}\right)=\frac{1}{2} \mathbf{X}^{T} Q_{n} \mathbf{X}
$$

Proceeding to find the limit distribution of $\hat{\theta}_{n}^{M L E}$, for any $h \in \mathbb{R}$ setting $\theta_{n}:=\theta_{0}+h / \sqrt{n}$ we have

$$
\begin{align*}
& \mathbb{P}_{\theta_{0}, Q_{n}}\left(\sqrt{n}\left(\hat{\theta}_{n}^{M L E}-\theta_{0}\right) \leq h\right) \\
= & \mathbb{P}_{\theta_{0}, Q_{n}}\left(\mathbf{X}^{T} Q_{n} \mathbf{X} \leq 2 Z_{n}^{\prime}\left(\theta_{n}\right)\right) \\
= & \mathbb{P}_{\theta_{0}, Q_{n}}\left(\frac{1}{\sqrt{n}} \mathbf{X}^{T} Q_{n} \mathbf{X}-\sqrt{n} m^{2}\left(\theta_{0}\right) \leq \frac{2 Z_{n}^{\prime}\left(\theta_{n}, Q_{n}\right)}{\sqrt{n}}-\sqrt{n} m^{2}\left(\theta_{0}\right)\right) . \tag{3.36}
\end{align*}
$$

We now claim that

$$
\begin{equation*}
\frac{2 Z_{n}^{\prime}\left(\theta_{n}, Q_{n}\right)}{\sqrt{n}}-\sqrt{n} m^{2}\left(\theta_{0}\right)=2 h R\left(\theta_{0}\right)+o(1) \tag{3.37}
\end{equation*}
$$

Given (3.37), using (3.34) and (3.36) we get

$$
\begin{aligned}
\mathbb{P}_{\theta_{0}, Q_{n}}\left(\sqrt{n}\left(\hat{\theta}_{n}^{M L E}-\theta_{0}\right) \leq h\right)= & =\mathbb{P}_{\theta_{0}, Q_{n}}\left(\frac{1}{\sqrt{n}} \mathbf{X}^{T} Q_{n} \mathbf{X}-\sqrt{n} m^{2}\left(\theta_{0}\right) \leq 2 h R\left(\theta_{0}\right)+o(1)\right) \\
& \longrightarrow \mathbb{P}\left(N\left(0, R\left(\theta_{0}\right)^{-1}\right) \leq h\right),
\end{aligned}
$$

which verifies asymptotic distribution for the MLE.

Proceeding to verify (3.37), note that the LHS of the display in part (a) of Lemma 3.3.2 is convex in $h$, and converges point-wise to a convex function which is differentiable everywhere. It follows that the derivatives of the functions also converge, which gives (3.37).

## - MPLE

It follows from [7, Cor 3.1 (b)] that if $\theta_{0}>1$, the MPLE $\hat{\theta}_{n}^{M P L E}$ exists with probability going to 1 , and is $\sqrt{n}$ consistent (existence without consistency also follows from Proposition 3.3.1). It thus suffices to prove asymptotic normality. To this effect, we show the more general result that for any $h \in \mathbb{R}$, setting $\theta_{n}:=\theta_{0}+\frac{h}{\sqrt{n}}$, under $\mathbb{P}_{\theta_{n}, Q_{n}}$ we have

$$
\begin{equation*}
\sqrt{n}\left(\hat{\theta}_{n}^{M P L E}-\theta_{0}\right) \xrightarrow{d} N\left(h, R\left(\theta_{0}\right)^{-1}\right) . \tag{3.38}
\end{equation*}
$$

The claimed limiting distribution for $\hat{\theta}_{n}^{M P L E}$ follows form this, on setting $h=0$.
Proceeding to prove (3.38), using (3.8), we have

$$
\begin{aligned}
\mathbf{X}^{T} Q_{n} \mathbf{X} & =\sum_{i=1}^{n} t_{i} \tanh \left(\hat{\theta}_{n}^{M P L E} t_{i}\right) \\
& =\sum_{i=1}^{n} t_{i} \tanh \left(\theta_{0} t_{i}\right)+\sum_{i=1}^{n} t_{i}^{2} \operatorname{sech}^{2}\left(\xi_{n} t_{i}\right)\left(\hat{\theta}_{n}^{M P L E}-\theta_{0}\right),
\end{aligned}
$$

where $\xi_{n}$ lies between $\theta_{0}$ and $\hat{\theta}_{n}^{M P L E}$. The last display gives

$$
\begin{equation*}
\sqrt{n}\left(\hat{\theta}_{n}^{M P L E}-\theta_{0}\right)=\frac{\frac{1}{\sqrt{n}}\left(\mathbf{X}^{T} Q_{n} \mathbf{X}-\sum_{i=1}^{n} t_{i} \tanh \left(\theta_{0} t_{i}\right)\right)}{\frac{1}{n} \sum_{i=1}^{n} t_{i}^{2} \operatorname{sech}^{2}\left(\xi_{n} t_{i}\right)} \tag{3.39}
\end{equation*}
$$

For analyzing the RHS of (3.39), use part (a) to note that $\mathbb{P}_{\theta_{0}, Q_{n}}$ we have

$$
\log \frac{d \mathbb{P}_{\theta_{n}, Q_{n}}}{d \mathbb{P}_{\theta_{0}, Q_{n}}}(\mathbf{X}) \xrightarrow{d} N\left(-\frac{h^{2} R\left(\theta_{0}\right)}{2}, h^{2} R\left(\theta_{0}\right)\right) .
$$

Then, using Le Cam's third lemma, it follows that the measures $\mathbb{P}_{\theta_{0}, Q_{n}}$ and $\mathbb{P}_{\theta_{n}, Q_{n}}$ are mutually contiguous. Along with [19, Lem 2.1 part (a)], this gives that under $\mathbb{P}_{\theta_{n}, Q_{n}}$, we have

$$
\left(\sum_{i=1}^{n}\left[t_{i}-m\left(\theta_{0}\right)\right]^{2} \mid \overline{\mathbf{X}}>0\right)=O_{p}(1)
$$

On the set $\overline{\mathbf{X}}>0$, under $\mathbb{P}_{\theta_{n}, Q_{n}}$ this gives

$$
\begin{aligned}
& \sum_{i=1}^{n} t_{i} \tanh \left(\theta_{0} t_{i}\right) \\
= & \sum_{i=1}^{n}\left[m\left(\theta_{0}\right) \tanh \left(\theta_{0} m\left(\theta_{0}\right)\right)+m\left(\theta_{0}\right)\left(1+\theta_{0}\left(1-m^{2}\left(\theta_{0}\right)\right)\right)\left(t_{i}-m\left(\theta_{0}\right)\right)\right]+O_{p}\left(\sum_{i=1}^{n}\left(t_{i}-m\left(\theta_{0}\right)\right)^{2}\right) \\
= & n m^{2}\left(\theta_{0}\right)+n m\left(\theta_{0}\right)\left(1+\theta_{0}\left(1-m^{2}\left(\theta_{0}\right)\right)\right)\left(\overline{\mathbf{X}}-m\left(\theta_{0}\right)\right)+O_{p}(1) .
\end{aligned}
$$

Since (3.32) gives $\mathbf{X}^{T} Q_{n} \mathbf{X}=n \overline{\mathbf{X}}^{2}+O_{p}(1)$, the above display implies that under $\mathbb{P}_{\theta_{n}, Q_{n}}$,

$$
\begin{aligned}
& \mathbf{X}^{T} Q_{n} \mathbf{X}-\sum_{i=1}^{n} t_{i} \tanh \left(\theta_{0} t_{i}\right) \\
= & n\left(\overline{\mathbf{X}}^{2}-m^{2}\left(\theta_{0}\right)\right)-n m\left(\theta_{0}\right)\left(1+\theta_{0}\left(1-m^{2}\left(\theta_{0}\right)\right)\right)\left(\overline{\mathbf{X}}-m\left(\theta_{0}\right)\right)+O_{p}(1) \\
= & n\left(\overline{\mathbf{X}}-m\left(\theta_{0}\right)\right)\left[\overline{\mathbf{X}}-\theta_{0} m\left(\theta_{0}\right)\left(1-m^{2}\left(\theta_{0}\right)\right)\right]+O_{p}(1) .
\end{aligned}
$$

Also, under $\mathbb{P}_{\theta_{n}, Q_{n}}$ we have

$$
\begin{equation*}
\sqrt{n}\left(\overline{\mathbf{X}}-m\left(\theta_{0}\right)\right)=\sqrt{n}\left(\overline{\mathbf{X}}-m\left(\theta_{n}\right)\right)+\sqrt{n}\left(m\left(\theta_{n}\right)-m\left(\theta_{0}\right)\right) \xrightarrow{d} N\left(m\left(\theta_{0}\right) \sigma^{2}\left(\theta_{0}\right) h, \sigma^{2}\right), \tag{3.40}
\end{equation*}
$$

where the last step uses part (a) of Lemma 3.3.1 along with Proposition 3.3.2. The last two displays give that under $\mathbb{P}_{\theta_{n}, Q_{n}}$,
$\frac{1}{\sqrt{n}}\left(\mathbf{X}^{T} Q_{n} \mathbf{X}-\sum_{i=1}^{n} t_{i} \tanh \left(\theta_{0} t_{i}\right)\right) \xrightarrow{d}\left(1-\theta_{0}\left(1-m^{2}\left(\theta_{0}\right)\right)\right) N\left(h m^{2}\left(\theta_{0}\right) \sigma^{2}\left(\theta_{0}\right), m^{2}\left(\theta_{0}\right) \sigma^{2}\left(\theta_{0}\right)\right)$.

Also, using contiguity of the two measures $\mathbb{P}_{\theta_{0}, Q_{n}}$ and $\mathbb{P}_{\theta_{n}, Q_{n}}$ along with consistency of $\hat{\theta}_{n}^{M P L E}$ gives $\xi_{n} \xrightarrow{p} \theta_{0}$ under both measures, and so

$$
\frac{1}{n} \sum_{i=1}^{n} t_{i}^{2} \operatorname{sech}^{2}\left(\xi_{n} t_{i}\right) \xrightarrow{p} m^{2}\left(\theta_{0}\right) \operatorname{sech}^{2}\left(\theta_{0} m\left(\theta_{0}\right)\right)=m^{2}\left(\theta_{0}\right)\left(1-m^{2}\left(\theta_{0}\right)\right)
$$

Combining the last two displays along with (3.39) we get that under $\mathbb{P}_{\theta_{n}, Q_{n}}$,

$$
\begin{aligned}
\sqrt{n}\left(\hat{\theta}_{n}^{M P L E}-\theta_{0}\right) & \stackrel{d}{\rightarrow}\left[\frac{1-\theta_{0}\left(1-m^{2}\left(\theta_{0}\right)\right)}{m^{2}\left(\theta_{0}\right)\left(1-m^{2}\left(\theta_{0}\right)\right)}\right] N\left(h m^{2}\left(\theta_{0}\right) \sigma^{2}\left(\theta_{0}\right), m^{2}\left(\theta_{0}\right) \sigma^{2}\left(\theta_{0}\right)\right) \\
& =\frac{1}{m^{2}\left(\theta_{0}\right) \sigma^{2}\left(\theta_{0}\right)} N\left(h m^{2}\left(\theta_{0}\right) \sigma^{2}\left(\theta_{0}\right), m^{2}\left(\theta_{0}\right) \sigma^{2}\left(\theta_{0}\right)\right),
\end{aligned}
$$

which is equivalent to (3.38).

## Part (c):

- MS Test

If $\mathbf{X} \sim \mathbb{P}_{\theta_{0}, Q_{n}}$, then (3.35) gives

$$
K_{n}(\alpha)=n m^{2}\left(\theta_{0}\right)+2 z_{\alpha} \sqrt{n R\left(\theta_{0}\right)}+o(\sqrt{n}) .
$$

Also, using (3.40), under $\mathbb{P}_{\theta_{n}, Q_{n}}$ we have

$$
\sqrt{n}\left(\overline{\mathbf{X}}^{2}-m^{2}\left(\theta_{0}\right)\right) \xrightarrow{d} 2 m\left(\theta_{0}\right) N\left(m\left(\theta_{0}\right) \sigma^{2}\left(\theta_{0}\right) h, \sigma^{2}\right)=N\left(2 m^{2}\left(\theta_{0}\right) \sigma^{2}\left(\theta_{0}\right) h, 4 m^{2}\left(\theta_{0}\right) \sigma^{2}\left(\theta_{0}\right)\right) .
$$

Using the last two displays, we have

$$
\begin{aligned}
\mathbb{P}_{\theta_{n}, Q_{n}}\left(n \overline{\mathbf{X}}^{2}>K_{n}(\alpha)\right) & =\mathbb{P}_{\theta_{n}, Q_{n}}\left(\sqrt{n}\left(\overline{\mathbf{X}}^{2}-m^{2}\left(\theta_{0}\right)\right)>2 z_{\alpha} \sqrt{R\left(\theta_{0}\right)}\right)+o(1) \\
& =\mathbb{P}\left(N\left(2 R\left(\theta_{0}\right) h, 4 R\left(\theta_{0}\right)\right)>2 z_{\alpha} \sqrt{R\left(\theta_{0}\right)}\right)+o(1) \\
& =\mathbb{P}\left(N\left(h \sqrt{R\left(\theta_{0}\right)}, 1\right)>z_{\alpha}\right)+o(1) .
\end{aligned}
$$

Thus we have $\beta_{M S}=1-\Phi\left(z_{\alpha}-h \sqrt{R\left(\theta_{0}\right)}\right)$, as claimed.

## - NP Test

Note that

$$
\mathbf{X}^{T} Q_{n} \mathbf{X}=n \overline{\mathbf{X}}^{2}+O_{p}(1)
$$

using Lemma 3.3.1 part (a) under $\mathbb{P}_{\theta_{0}, Q_{n}}$ and $\mathbb{P}_{\theta_{n}, Q_{n}}$. Thus $\mathbf{X}^{T} Q_{n} \mathbf{X}$ has the same asymptotic distribution as $n \overline{\mathbf{X}}^{2}$, under both null and the alternative. This gives $\beta_{N P}=\beta_{M S}$, as desired.

## - PL Test

Using the limit distribution of $\hat{\theta}_{n}^{M P L E}$ in part (b) (invoke (3.38) under $\mathbb{P}_{\theta_{0}, Q_{n}}$ ) gives

$$
K_{n}(\alpha)=\theta_{0}+\frac{z_{\alpha}}{\sqrt{n R\left(\theta_{0}\right)}}+o\left(\frac{1}{\sqrt{n}}\right) .
$$

Thus we have

$$
\begin{aligned}
\mathbb{P}_{\theta_{n}, Q_{n}}\left(\hat{\theta}_{n}^{M P L E}>K_{n}(\alpha)\right) & =\mathbb{P}_{\theta_{n}, Q_{n}}\left(\sqrt{n}\left(\hat{\theta}_{n}^{M P L E}-\theta_{0}\right)>\frac{z_{\alpha}}{\sqrt{R\left(\theta_{0}\right)}}\right)+o(1) \\
& =\mathbb{P}\left(N\left(h, \frac{1}{R\left(\theta_{0}\right)}\right)>\frac{z_{\alpha}}{\sqrt{R\left(\theta_{0}\right)}}\right)+o(1),
\end{aligned}
$$

where the last step again uses (3.38) under $\mathbb{P}_{\theta_{n}, Q_{n}}$. The last term in the display above converges to

$$
\mathbb{P}\left(N\left(h \sqrt{R\left(\theta_{0}\right)}, 1\right)>z_{\alpha}\right)=1-\Phi\left(z_{\alpha}-h \sqrt{R\left(\theta_{0}\right)}\right),
$$

thus giving the same formula for $\beta_{P L}$.

### 3.3.2 Proof of Theorem 3.2.3

## Proof of Theorem 3.2.3.

## Part (a):

As in the proof of Theorem 3.2.2, let $I:=\left\{h_{1}, \ldots, h_{k}\right\}$ with $\left\{h_{1}<h_{2}<\ldots<h_{k}\right\}$ 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 (3.33). To this effect, use Lemma 3.3.2 part (b) to get

$$
Z_{n}\left(\theta_{n, i}, Q_{n}\right)-Z_{n}\left(\theta_{0}, Q_{n}\right) \rightarrow F(h)-F(0) .
$$

Proceeding to analyze the second term in the RHS of (3.33), using (3.32), under $\mathbb{P}_{\theta_{0}, Q_{n}}$ we have

$$
\begin{equation*}
\frac{\mathbf{X}^{T} Q_{n} \mathbf{X}}{\sqrt{n}}=\sqrt{n} \overline{\mathbf{X}}^{2}+O_{p}\left(\frac{1}{\sqrt{n}}\right) \xrightarrow{d} U_{1,0}^{2}, \tag{3.41}
\end{equation*}
$$

where the last step uses part (b) of Lemma 3.3.1. Combining the last two displays along with
(3.33), under $\mathbb{P}_{\theta_{0}, Q_{n}}$ we have

$$
\left[\log \frac{d \mathbb{P}_{\theta_{n, i}, Q_{n}}}{d \mathbb{P}_{\theta_{0}, Q_{n}}}(\mathbf{X})\right] \xrightarrow{d}\left[-F\left(h_{i}\right)+F(0)+\frac{h_{i}}{2} U_{1,0}^{2}\right]_{1 \leq i \leq k}
$$

Also, we have

$$
\log \frac{d \mathbb{H}_{h_{i}}}{d \mathbb{H}_{0}}(u)=\frac{1}{2} h_{i} u-F\left(h_{i}\right)+F(0) .
$$

Combining the above two displays, under $\mathbb{P}_{\theta_{0}, Q_{n}}$ we have

$$
\left[\log \frac{d \mathbb{P}_{\theta_{n, i}, Q_{n}}}{d \mathbb{P}_{\theta_{0}, Q_{n}}}(\mathbf{X})\right] \xrightarrow{d}\left[\frac{d \mathbb{H}_{h_{i}}}{d \mathbb{H}_{0}}\left(U_{1,0}\right)\right]_{1 \leq i \leq k},
$$

where $U_{0} \sim \mathbb{H}_{0}$. This verifies the desired convergence of experiments.

## Part (b):

## - MLE

Existence of MLE follows from Proposition 3.3.1 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 3.3.2 with respect to $h$ we get

$$
\frac{2 Z_{n}^{\prime}\left(\theta_{n}, Q_{n}\right)}{\sqrt{n}}=\mathbb{E} U_{1, h}^{2}+o(1)
$$

Consequently, using calculations similar to the proof of Theorem 3.2.2 part (b) we have

$$
\begin{aligned}
\mathbb{P}_{1, Q_{n}}\left(\sqrt{n}\left(\tilde{\theta}_{n}^{M L E}-1\right) \leq h\right) & =\mathbb{P}_{1, Q_{n}}\left(\mathbf{X}^{T} Q_{n} \mathbf{X} \leq 2 Z_{n}^{\prime}\left(\theta_{n}, Q_{n}\right)\right) \\
& =\mathbb{P}_{1, Q_{n}}\left(\frac{1}{\sqrt{n}} \mathbf{X}^{T} Q_{n} \mathbf{X} \leq \mathbb{E} U_{1, h}^{2}+o(1)\right) \\
& =\mathbb{P}\left(U_{1,0}^{2} \leq \mathbb{E} U_{1, h}^{2}\right)+o(1),
\end{aligned}
$$

which derives the limiting distribution of the MLE.

## - MPLE

Existence of the MPLE follows from Proposition 3.3.1 part (d). Proceeding to show consistency, use (3.8) to note that on the set $\hat{\theta}_{n}^{M P L E}>1+\delta$ under $\mathbb{P}_{1, Q_{n}}$ we have

$$
\begin{aligned}
\mathbf{X}^{T} Q_{n} \mathbf{X} & =\sum_{i=1}^{n} t_{i} \tanh \left(\hat{\theta}_{n}^{M P L E} t_{i}\right) \\
& \geq \sum_{i=1}^{n} t_{i} \tanh \left((1+\delta) t_{i}\right) \\
& =n \overline{\mathbf{X}} \tanh ((1+\delta) \overline{\mathbf{X}})+O_{p}\left(\sum_{i=1}^{n}\left(t_{i}-\overline{\mathbf{X}}\right)^{2}\right) .
\end{aligned}
$$

Using Lemma 3.3.1 part (b), under $\mathbb{P}_{1, Q_{n}}$ we get

$$
\mathbf{X}^{T} B_{n} \mathbf{X}=O_{p}(1), \quad \sum_{i=1}^{n}\left(t_{i}-\mathbf{X}\right)^{2}=\mathbf{X}^{T} B_{n}^{2} \mathbf{X}=O_{p}(1)
$$

Combining the last two displays, under $\mathbb{P}_{1, Q_{n}}$ we have

$$
\sqrt{n} \overline{\mathbf{X}}^{2} \geq \sqrt{n} \overline{\mathbf{X}} \tanh ((1+\delta) \overline{\mathbf{X}})+O_{p}\left(\frac{1}{\sqrt{n}}\right) .
$$

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 3.3.1 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 $\mathbb{P}_{1, Q_{n}}\left(\hat{\theta}_{n}^{M P L E} \geq 1+\delta\right) \rightarrow 0$. A similar proof gives $\mathbb{P}_{1, Q_{n}}\left(\hat{\theta}_{n}^{M P L E} \leq 1-\delta\right) \rightarrow 0$, and so $\hat{\theta}_{n}^{M P L E} \xrightarrow{p} 1$.

Proceeding to find the limiting distribution of $\hat{\theta}_{n}^{M P L E}$, 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

$$
\begin{equation*}
\sqrt{n}\left(\hat{\theta}_{n}^{M P L E}-1\right) \xrightarrow{d} V_{1, h} . \tag{3.42}
\end{equation*}
$$

As before, the desired limiting distribution for $\sqrt{n}\left(\hat{\theta}_{n}^{M P L E}-1\right)$ under $\mathbb{P}_{1, Q_{n}}$ then follows on taking $h=0$.

For proving (3.42), 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 3.2.2 part (b)), and so $\hat{\theta}_{n}^{M P L E} \xrightarrow{p} 1$ under $\mathbb{P}_{\theta_{n}, Q_{n}}$ as well. Further, using (3.8), under $\mathbb{P}_{\theta_{n}, Q_{n}}$ we have

$$
\begin{aligned}
& \sum_{i=1}^{n} t_{i} \tanh \left(\hat{\theta}_{n}^{M P L E} t_{i}\right) \\
= & \sum_{i=1}^{n} t_{i}\left[\tanh \left(t_{i}\right)+\left(\hat{\theta}_{n}^{M P L E}-1\right) t_{i} \operatorname{sech}^{2}\left(t_{i}\right)+O_{p}\left(\left(\hat{\theta}_{n}^{M P L E}-1\right)^{2}\left|t_{i}\right|^{3}\right)\right] .
\end{aligned}
$$

Under $\mathbb{P}_{\theta_{n}, Q_{n}}$, this gives

$$
\begin{equation*}
\hat{\theta}_{n}^{M P L E}-1=\frac{\mathbf{X}^{T} Q_{n} \mathbf{X}-\sum_{i=1}^{n} t_{i} \tanh \left(t_{i}\right)}{\sum_{i=1}^{n} t_{i}^{2} \operatorname{sech}^{2}\left(t_{i}\right)+O_{p}\left(\left(\hat{\theta}_{n}^{M P L E}-1\right) \sum_{i=1}^{n} t_{i}^{4}\right)} . \tag{3.43}
\end{equation*}
$$

For analyzing the numerator in (3.43), under $\mathbb{P}_{\theta_{n}, Q_{n}}$ we have

$$
\begin{aligned}
& \mathbf{X}^{T} Q_{n} \mathbf{X}-\sum_{i=1}^{n} t_{i} \tanh \left(t_{i}\right) \\
= & n \overline{\mathbf{X}}^{2}+\mathbf{X}^{T} B_{n} \mathbf{X}-n \overline{\mathbf{X}} \tanh (\overline{\mathbf{X}})-\sum_{i=1}^{n}\left(t_{i}-\overline{\mathbf{X}}\right)^{2}+O_{p}\left(\sum_{i=1}^{n}\left|t_{i}-\overline{\mathbf{X}}\right|^{3}\right) \\
= & \frac{n \overline{\mathbf{X}}^{4}}{3}+\mathbf{X}^{T} B_{n} \mathbf{X}-\mathbf{X}^{T} B_{n}^{2} \mathbf{X}+O_{p}\left(n \sqrt{\frac{\log n}{n}}{ }^{3}\right)+O_{p}\left(n^{1-\frac{6}{4}}\right) \\
= & \frac{n \overline{\mathbf{X}}^{4}}{3}+\mathbf{X}^{T} B_{n} \mathbf{X}-\mathbf{X}^{T} B_{n}^{2} \mathbf{X}+o_{p}(1),
\end{aligned}
$$

where in the last but one step we use Lemma 3.3.1 part (b) and [19, Lem 2.4], along with mutual contiguity shown above, to get that under $\mathbb{P}_{\theta_{n}, Q_{n}}$ we have

$$
|\overline{\mathbf{X}}|=O_{p}\left(n^{-1 / 4}\right), \quad \max _{i \in[n]}\left|t_{i}-\overline{\mathbf{X}}\right|=O_{p}\left(\sqrt{\frac{\log n}{n}}\right) .
$$

For the denominator in (3.43), under $\mathbb{P}_{\theta_{n}, Q_{n}}$ we have

$$
\begin{aligned}
& \sum_{i=1}^{n} t_{i}^{2} \operatorname{sech}^{2}\left(t_{i}\right)+O_{p}\left(\left(\hat{\theta}_{n}^{M P L E}-1\right) \sum_{i=1}^{n} t_{i}^{4}\right) \\
= & \sum_{i=1}^{n} t_{i}^{2}+O_{p}\left(\sum_{i=1}^{n} t_{i}^{4}\right) \\
= & n \overline{\mathbf{X}}^{2}+\sum_{i=1}^{n}\left(t_{i}-\overline{\mathbf{X}}\right)^{2}+O_{p}\left(\sum_{i=1}^{n}\left(t_{i}-\overline{\mathbf{X}}\right)^{4}+n \overline{\mathbf{X}}^{4}\right) \\
= & n \overline{\mathbf{X}}^{2}+O_{p}(1),
\end{aligned}
$$

where the last step uses Lemma 3.3.1 part (a), and mutual contiguity. Combining the above two displays along with (3.43), under $\mathbb{P}_{\theta_{n}, Q_{n}}$ we get

$$
\begin{aligned}
\sqrt{n}\left(\hat{\theta}_{n}^{M P L E}-1\right) & \stackrel{d}{=} \frac{\frac{n \overline{\mathbf{N}}^{4}}{3}+\mathbf{X}^{T} B_{n} \mathbf{X}-\mathbf{X}^{T} B_{n}^{2} \mathbf{X}+o_{p}(1)}{\sqrt{n} \overline{\mathbf{X}}^{2}+o_{p}(1)} \\
& \xrightarrow{d} \frac{\frac{U_{1, h}^{4}}{3}+S_{1}-T_{1}}{U_{1, h}^{2}}
\end{aligned}
$$

where we again use part (b) of Lemma 3.3.1. Recalling the formula of $V_{1, h}$ we have verified (3.42), and this completes part (b).

## Part (c):

## - MS Test

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

$$
\alpha=\mathbb{P}_{1, Q_{n}}\left(n \overline{\mathbf{X}}^{2}>K_{n}(\alpha)\right)=2 \mathbb{P}_{1, Q_{n}}\left(\overline{\mathbf{X}}>\sqrt{K_{n}(\alpha)}\right) .
$$

Using this, along with the limit distribution $n^{1 / 4} \overline{\mathbf{X}} \xrightarrow{d} U_{1,0}$ under $\mathbb{P}_{1, Q_{n}}$ (see part (b) of Lemma 3.3.1) gives

$$
\sqrt{K_{n}(\alpha)}=n^{1 / 4} \Psi_{\mathbb{H}_{0}}(1-\alpha / 2)+o\left(n^{1 / 4}\right) .
$$

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

$$
\begin{aligned}
2 \mathbb{P}_{\theta_{n}}\left(\sqrt{n} \overline{\mathbf{X}}>K_{n}(\alpha)\right) & =2 \mathbb{P}_{\theta_{n}, Q_{n}}\left(n^{1 / 4} \overline{\mathbf{X}}>\Psi_{\mathbb{H}_{0}}(1-\alpha / 2)+o(1)\right) \\
& =2 \mathbb{P}\left(U_{1, h}>\Psi_{\mathbb{H}_{0}}(1-\alpha / 2)\right)+o(1),
\end{aligned}
$$

where the last line again uses part (b) of Lemma 3.3.1. Thus we have $\beta_{M S}=2 \mathbb{P}\left(U_{1, h}>\right.$ $\left.\Psi_{\mathbb{H}_{0}}(1-\alpha / 2)\right)$, as desired.

## - NP Test

As in the proof of Theorem 3.2.2, we have $\mathbf{X}^{T} Q_{n} \mathbf{X}=n \overline{\mathbf{X}}^{2}+O_{p}(1)$ under both $\mathbb{P}_{1, Q_{n}}$ and $\mathbb{P}_{\theta_{n}, Q_{n}}$, using (3.32) and mutual contiguity. Thus we get $\beta_{N P}=\beta_{M S}$ as before.

## - PL Test

Using (3.42), under $\mathbb{P}_{1, Q_{n}}$ we have

$$
\sqrt{n}\left(\hat{\theta}_{n}^{M P L E}-1\right) \xrightarrow{d} V_{1,0}, \text { which gives } K_{n}(\alpha)=1+\frac{\Psi_{V_{1,0}}(\alpha)}{\sqrt{n}}+o\left(\frac{1}{\sqrt{n}}\right) .
$$

Then the asymptotic power is given by

$$
\mathbb{P}_{\theta_{n}, Q_{n}}\left(\hat{\theta}_{n}>K_{n}(\alpha)\right)=\mathbb{P}\left(V_{1, h}>\Psi_{V_{1,0}}(\alpha)\right)+o(1),
$$

where the last step again uses (3.42). This shows that $\beta_{P L}=\mathbb{P}\left(V_{1, h}>\Psi_{V_{1,0}}(\alpha)\right)$, as desired.

### 3.3.3 Proof of Theorem 3.2.4

## Part (a):

As in the proof of Theorem 3.2.2, let $I:=\left\{h_{1}, \ldots, h_{k}\right\} \subset\left(-\theta_{0}, 1-\theta_{0}\right)$ with $\left\{h_{1}<h_{2}<\ldots<\right.$ $\left.h_{k}\right\}$ for some positive integer $k$, and let $\theta_{i}:=\theta_{0}+h_{i} \in(0,1)$ for $1 \leq i \leq k$. It thus suffices to
analyze the terms in the RHS of (3.33). To this effect, first use part (c) of Lemma 3.3.1 to note that under $\mathbb{P}_{\theta_{0}, Q_{n}}$ we have

$$
\begin{align*}
& \mathbf{X}^{T} Q_{n} \mathbf{X}=n \overline{\mathbf{X}}^{2}+\mathbf{X}^{T} B_{n} \mathbf{X} \xrightarrow{d} W_{\theta_{0}}^{2}+S_{\theta_{0}} \\
& \stackrel{d}{=} \sum_{j=1}^{\infty} \lambda_{j}\left(\frac{Y_{j}}{1-\theta_{0} \lambda_{j}}-1\right)+\theta_{0} \kappa+W^{*}=U_{\theta_{0}}, \tag{3.44}
\end{align*}
$$

where $\left\{Y_{j}\right\}_{j \geq 1} \stackrel{i i d}{\sim} \chi_{1}^{2}$ are mutually independent, $U_{\theta_{0}}$ is as defined (3.25), and we use the fact that $\lambda_{1}=1$ (as $Q_{n}$ satisfies (3.2), and using Proposition 3.7.1 part (b)). Thus combining the last display along with (3.33) and part (c) of Lemma 3.3.2, under $\mathbb{P}_{\theta_{0}, Q_{n}}$ we have

$$
\log \frac{d \mathbb{P}_{\theta_{i}, Q_{n}}}{d \mathbb{P}_{\theta_{0}, Q_{n}}}(\mathbf{X}) \xrightarrow{d} \frac{h_{i}}{2} U_{\theta_{0}}-\frac{\kappa}{4} h_{i}^{2}-\frac{\kappa \theta_{0} h_{i}}{2}+\frac{1}{2} \sum_{j=1}^{\infty}\left(\lambda_{j} h_{i}-\log \left(\frac{1-\theta_{0} \lambda_{j}}{1-\theta_{i} \lambda_{j}}\right)\right) .
$$

To show the desired convergence of experiments, with $\mathbf{Z}:=\left(Z_{j}\right)_{j \in \mathbb{Z}_{+}} \in \mathbb{R}^{\mathbb{Z}_{+}}$we need to show that under $v_{\theta_{0}}$ we have

$$
\begin{equation*}
\left[\frac{d v_{\theta_{i}}}{d v_{\theta_{0}}}(\mathbf{Z})\right]_{1 \leq i \leq k} \stackrel{d}{=}\left[\frac{h_{i}}{2} S\left(\theta_{0}\right)-\frac{\kappa}{4} h_{i}^{2}-\frac{\kappa \theta_{0} h_{i}}{2}+\frac{1}{2} \sum_{j=1}^{\infty}\left(\lambda_{j} h_{i}-\log \left(\frac{1-\theta_{0} \lambda_{j}}{1-\theta_{i} \lambda_{j}}\right)\right]_{1 \leq i \leq k}\right. \tag{3.45}
\end{equation*}
$$

where $v_{\theta}$ is as in Definition 3.2.6. To this effect, first note that for any $j \geq 1$ the log Hellinger affinity between the centered Gaussian distributions $v_{\theta_{i}, j}$ and $v_{\theta_{0}, j}$ is given by

$$
\log \int_{\mathbb{R}} \sqrt{\frac{d v_{\theta_{i}, j}}{d x} \frac{d v_{\theta_{i}, j}}{d x}} d x=\log \frac{2\left(1-\theta_{0} \lambda_{j}+\theta_{i} \lambda_{j}+\theta_{0} \theta_{i} \lambda_{j}^{2}\right)}{\sqrt{\left(1-\theta_{0} \lambda_{j}\right)\left(1-\theta_{i} \lambda_{j}\right)}\left(2-\theta_{0} \lambda_{j}-\theta_{i} \lambda_{j}\right)}=O\left(\lambda_{j}^{2}\right),
$$

where the last equality uses the estimate $\log (1+x)=x+O\left(x^{2}\right)$ for all $|x| \leq 1-\delta$, for any $\delta>0$. Since $\sum_{j=1}^{\infty} \lambda_{j}^{2}<\infty$ (by Proposition 3.7.1 part (a)), the LHS of the display above is summable in $j$, for every $i \in[k]$. It then follows by Kakutani's theorem [22, Thm 4.3.8] that the probability measures $\otimes_{j \in \mathbb{Z}_{+}} v_{\theta_{0}, j}$ and $\otimes_{j \in \mathbb{Z}_{+}} v_{\theta_{i}, j}$ are mutually absolutely continuous, and further under $v_{\theta_{0}}$ we
have

$$
\begin{equation*}
\left(\prod_{j=0}^{J} \frac{d v_{\theta_{i}, j}}{d v_{\theta_{0}, j}}\right)_{1 \leq i \leq k} \xrightarrow{d}\left(\prod_{j \in \mathbb{Z}_{+}} \frac{d v_{\theta_{i}, j}}{d v_{\theta_{0}, j}}\right)_{1 \leq i \leq k} . \tag{3.46}
\end{equation*}
$$

Now, for $j=0$ we have

$$
\log \frac{d v_{\theta_{i}, 0}}{d v_{\theta_{0}, 0}}\left(Z_{0}\right)=h_{i} \sqrt{\frac{\kappa}{2}}\left(Z_{0}-\theta_{0} \sqrt{\frac{\kappa}{2}}\right)-\frac{\kappa}{4} h_{i}^{2} \stackrel{d}{=}-\frac{\kappa}{4} h_{i}^{2}+\frac{h_{i}}{2} W^{*}
$$

where $Z_{0} \sim N\left(\theta_{0} \sqrt{\frac{\kappa}{2}}, 1\right)$ and $W^{*} \sim N(0,2 \kappa)$. And for $j \geq 1$ we have

$$
\log \frac{d v_{\theta_{i}, j}}{d v_{\theta_{0}, j}}\left(Z_{j}\right) \stackrel{d}{=} \frac{1}{2} \log \left(\frac{1-\theta_{i} \lambda_{j}}{1-\theta_{0} \lambda_{j}}\right)+\frac{h_{i}}{2} \lambda_{j} Z_{j}^{2} \stackrel{d}{=} \frac{1}{2} \log \left(\frac{1-\theta_{i} \lambda_{j}}{1-\theta_{0} \lambda_{j}}\right)+\frac{1}{2\left(1-\theta_{0} \lambda_{j}\right)} h_{i} \lambda_{j} Y_{j},
$$

where $Z_{j} \sim N\left(0, \frac{1}{1-\theta_{0} \lambda_{j}}\right)$ and $Y_{j} \sim \chi_{1}^{2}$. Combining the last two displays, under $v_{\theta_{0}}$ we get

$$
\begin{aligned}
\log \frac{d \bigotimes_{j=0}^{J} v_{\theta_{i}, j}}{d \bigotimes_{j=0}^{J} v_{\theta_{0}, j}}\left(Z_{j}, 0 \leq j \leq J\right) \stackrel{d}{=} & -\frac{\kappa}{4} h_{i}^{2}+\frac{h_{i}}{2}\left(\sum_{j=1}^{J} \lambda_{j}\left(\frac{Y_{j}}{1-\theta_{0} \lambda_{j}}-1\right)+W^{*}\right) \\
& +\frac{1}{2}\left(\sum_{j=1}^{J} \lambda_{j} h_{i}-\log \left(\frac{1-\theta_{0} \lambda_{j}}{1-\theta_{i} \lambda_{j}}\right)\right)
\end{aligned}
$$

Letting $J \rightarrow \infty$ and using (3.46), under $v_{\theta_{0}}$ we get

$$
\begin{aligned}
{\left[\begin{array}{rl}
d \bigotimes_{j \in \mathbb{Z}} \\
\log v_{\theta_{i}, j} \\
d \bigotimes_{j \in \mathbb{Z}_{+}} v_{\theta_{0}, j}
\end{array}(\mathbf{Z})\right]_{1 \leq i \leq k} \xrightarrow{d} } & {\left[-\frac{\kappa}{4} h_{i}^{2}+\frac{h_{i}}{2}\left(\sum_{j=1}^{\infty} \lambda_{j}\left(\frac{Y_{j}}{1-\theta_{0} \lambda_{j}}-1\right)+W^{*}\right)\right.} \\
& \left.+\frac{1}{2}\left(\sum_{j=1}^{J} \lambda_{j} h_{i}-\log \left(\frac{1-\theta_{0} \lambda_{j}}{1-\theta_{i} \lambda_{j}}\right)\right)\right]_{1 \leq i \leq k},
\end{aligned}
$$

where we again use the fact $\sum_{j=1}^{\infty} \lambda_{j}^{2}<\infty$ (from Proposition 3.7.1 part (a)). From this, the desired conclusion (3.45) follows on using Lemma 3.3.2 part (c), along with (3.33). This verifies concludes
the proof of Theorem 3.2.4 part (a).

## Part (b):

## - PL Test

We first investigate Pseudo-Likelihood test (3.12). Proposition 3.3.1 part (d) shows that the MPLE exists with probability tending to 1 . We now claim that $\hat{\theta}_{n}^{M P L E}=O_{p}(1)$ under $\mathbb{P}_{\theta_{0}, Q_{n}}$. Indeed, the MPLE $\hat{\theta}_{n}^{M P L E}$ satisfies

$$
\begin{align*}
\mathbf{X}^{T} Q_{n} \mathbf{X} & =\sum_{i=1}^{n} t_{i} \tanh \left(\hat{\theta}_{n}^{M P L E} t_{i}\right) \\
& =\sum_{i=1}^{n} t_{i} \tanh \left(\theta_{0} t_{i}\right)+\sum_{i=1}^{n} t_{i}^{2} \int_{\theta_{0}}^{\hat{\theta}_{n}^{M P L E}} \operatorname{sech}^{2}\left(\xi t_{i}\right) d \xi . \tag{3.47}
\end{align*}
$$

Also by [19, Lemma 2.3], under $\mathbb{P}_{\theta_{0}, Q_{n}}$ we have

$$
\begin{equation*}
\limsup _{M \rightarrow \infty} \limsup _{n \rightarrow \infty} \mathbb{P}_{\theta_{0}, Q_{n}}\left(\max _{1 \leq i \leq n}\left|t_{i}\right|>M \sqrt{\frac{\log n}{n}}\right)=0 \tag{3.48}
\end{equation*}
$$

Thus, fixing $K, M>0$ on the set

$$
\left\{\max _{1 \leq i \leq n}\left|t_{i}\right| \leq M \sqrt{\frac{\log n}{n}},\left|\hat{\theta}_{n}^{M P L E}\right|>K\right\},
$$

we have

$$
\begin{aligned}
\left|\mathbf{X}^{T} Q_{n} \mathbf{X}-\sum_{i=1}^{n} t_{i} \tanh \left(\theta_{0} t_{i}\right)\right| & =\left|\sum_{i=1}^{n} t_{i}^{2} \int_{\theta_{0}}^{\hat{\theta}_{n}^{M P L E}} \operatorname{sech}^{2}\left(\xi t_{i}\right) d \xi\right| \\
& \geq\left|\sum_{i=1}^{n} t_{i}^{2} \int_{\theta_{0}}^{K} \operatorname{sech}^{2}\left(M \xi \sqrt{\frac{\log n}{n}}\right) d \xi\right| \\
& \geq\left(K-\theta_{0}\right) \operatorname{sech}^{2}\left(M K \sqrt{\frac{\log n}{n}}\right) \sum_{i=1}^{n} t_{i}^{2}
\end{aligned}
$$

Also we have

$$
\sum_{i=1}^{n} t_{i}^{2}=\mathbf{X}^{T} Q_{n}^{2} \mathbf{X}=n \overline{\mathbf{X}}^{2}+\mathbf{X}^{T} B_{n}^{2} \mathbf{X}, \quad \max _{1 \leq i \leq n}\left|t_{i}\right|=O_{p}\left(\sqrt{\frac{\log n}{n}}\right)
$$

(the second conclusion follows from (3.48)). Combining the last two displays, under $\mathbb{P}_{\theta_{0}, Q_{n}}$ we get

$$
\begin{aligned}
\left(K-\theta_{0}\right) \operatorname{sech}^{2}\left(M K \sqrt{\frac{\log n}{n}}\right) & \leq \frac{\left|\mathbf{X}^{T} Q_{n} \mathbf{X}-\sum_{i=1}^{n} t_{i} \tanh \left(\theta_{0} t_{i}\right)\right|}{\sum_{i=1}^{n} t_{i}^{2}} \\
& =\frac{\left|\mathbf{X}^{T} Q_{n} \mathbf{X}-\theta_{0} \sum_{i=1}^{n} t_{i}^{2}-O_{p}\left(\sum_{i=1}^{n} t_{i}^{4}\right)\right|}{\sum_{i=1}^{n} t_{i}^{2}} \\
& =\left|\frac{n \overline{\mathbf{X}}^{2}+\mathbf{X}^{T} B_{n} \mathbf{X}}{n \overline{\mathbf{X}}^{2}+\mathbf{X}^{T} B_{n}^{2} \mathbf{X}}-\theta_{0}\right|+o_{p}(1) .
\end{aligned}
$$

The RHS of the display above converges in distribution as $n \rightarrow \infty$, by part (c) of Lemma 3.3.1, whereas the LHS converges to $K-\theta_{0}$. Since $K>0$ is arbitrary, this gives

$$
\limsup _{K \rightarrow \infty} \limsup _{n \rightarrow \infty} \mathbb{P}_{\theta_{0}, Q_{n}}\left(\left|\hat{\theta}_{n}^{M P L E}\right|>K, \max _{1 \leq i \leq n}\left|t_{i}\right| \leq M \sqrt{\frac{\log n}{n}}\right)=0 .
$$

Along with (3.48), this gives

$$
\limsup _{K \rightarrow \infty} \limsup _{n \rightarrow \infty} \mathbb{P}_{\theta_{0}, Q_{n}}\left(\left|\hat{\theta}_{n}^{M P L E}\right|>K\right)=0,
$$

which is equivalent to $\hat{\theta}_{n}^{M P L E}=O_{p}(1)$. Using (3.47) now gives under $\mathbb{P}_{\theta_{0}, Q_{n}}$,

$$
\hat{\theta}_{n}^{M P L E}-\theta_{0}=\frac{\mathbf{X}^{T} Q_{n} \mathbf{X}-\sum_{i=1}^{n} t_{i} \tanh \left(\theta_{0} t_{i}\right)}{\sum_{i=1}^{n} t_{i}^{2} \operatorname{sech}^{2}\left(\xi_{n} t_{i}\right)}=\frac{n \overline{\mathbf{X}}^{2}+\mathbf{X}^{T} B_{n} \mathbf{X}}{n \overline{\mathbf{X}}^{2}+\mathbf{X}^{T} B_{n}^{2} \mathbf{X}}-\theta_{0}+o_{p}(1),
$$

for some $\xi_{n}$ lying between $\theta_{0}$ and $\hat{\theta}_{n}^{M P L E}$ (and hence $\left.O_{p}(1)\right)$. Using part (c) of Lemma 3.3.1 along with the above display, under $\mathbb{P}_{\theta_{0}, Q_{n}}$ we get

$$
\begin{align*}
\hat{\theta}_{n}^{M P L E}-\theta_{0} & \xrightarrow{d} \frac{W_{\theta_{0}}^{2}+S_{\theta_{0}}}{W_{\theta_{0}}^{2}+T_{\theta_{0}}} \\
& \stackrel{\frac{d}{1-\theta} Y_{1}+\sum_{j=2}^{\infty} \lambda_{j}\left(\frac{Y_{j}}{1-\theta \lambda_{j}}-1\right)-1+\theta \kappa+W^{*}}{\frac{1}{1-\theta} Y_{1}+\sum_{j=2}^{\infty} \frac{\lambda_{j}^{2} Y_{j}}{1-\theta \lambda_{j}}+\kappa}=V_{\theta_{0}}, \tag{3.49}
\end{align*}
$$

where $V_{\theta_{0}}$ is as defined in (3.26). Since $\theta_{0} \in(0,1)$ is arbitrary, the same argument above also shows that under $\mathbb{P}_{\theta_{0}+h, Q_{n}}$ we have $\hat{\theta}_{n}^{M P L E}-\theta_{0} \xrightarrow{d} V_{\theta_{0}+h}$. The desired formula for $\beta_{P L}$ follows from this.

## - NP Test

Using (3.44) it follows that for all $\theta \in(0,1)$, under $\mathbb{P}_{\theta, Q_{n}}$ we have $\mathbf{X}^{T} Q_{n} \mathbf{X} \xrightarrow{d} U_{\theta}$. The formula for $\beta_{N P}$ follows on using this with $\theta=\theta_{0}+h$ and $\theta=\theta_{0}$.

## - MS Test

Again invoking part (c) of Lemma 3.3.1, under $\mathbb{P}_{\theta, Q_{n}}$ with $\theta \in(0,1)$ we have

$$
n \overline{\mathbf{X}}^{2} \xrightarrow{d} W_{\theta}^{2} \stackrel{d}{=} \frac{Y_{1}}{1-\theta} .
$$

Thus we have

$$
\begin{aligned}
\alpha=\mathbb{P}_{\theta_{0}, Q_{n}}\left(n \overline{\mathbf{X}}^{2}>K_{n}(\alpha)\right) & =\mathbb{P}\left(Y_{1}>\left(1-\theta_{0}\right) K_{n}(\alpha)\right)+o(1) \\
& =2 \mathbb{P}\left(N(0,1)>\sqrt{\left(1-\theta_{0}\right) K_{n}(\alpha)}\right)+o(1) .
\end{aligned}
$$

This gives $K_{n}(\alpha)=\frac{z_{\alpha / 2}^{2}}{1-\theta_{0}}+o(1)$. In turn this gives

$$
\begin{aligned}
\mathbb{P}_{\theta_{0}+h, Q_{n}}\left(n \bar{X}^{2}\right. & \left.>K_{n}(\alpha)\right)=\mathbb{P}\left(\frac{Y_{1}}{1-\theta_{0}-h}>\frac{z_{\alpha / 2}^{2}}{1-\theta_{0}}\right)+o(1) \\
& =2 \mathbb{P}\left(N(0,1)>z_{\alpha / 2} \sqrt{\frac{1-\theta_{0}-h}{1-\theta_{0}}}\right)+o(1),
\end{aligned}
$$

thus verifying the desired formula for $\beta_{M S}$.

### 3.3.4 Proof of Proposition 3.3.1

(a) By definition of $a_{n}, b_{n}$ we have

$$
\lim _{\theta \rightarrow-\infty} \mathbb{E}_{\theta} \mathbf{X}^{\prime} Q_{n} \mathbf{X}=a_{n}, \quad \lim _{\theta \rightarrow \infty} \mathbb{E}_{\theta} \mathbf{X}^{T} Q_{n} \mathbf{X}=b_{n}
$$

Since the function

$$
\theta \mapsto \frac{\theta}{2} \mathbf{X}^{T} Q_{n} \mathbf{X}-Z_{n}\left(\theta, Q_{n}\right)
$$

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

$$
\mathbf{X}^{\prime} Q_{n} \mathbf{X}=\mathbb{E}_{\theta} \mathbf{X}^{\prime} Q_{n} \mathbf{X}
$$

has a real solution, which holds iff $a_{n}<\mathbf{X}^{T} Q_{n} \mathbf{X}<b_{n}$, as desired.
(b) This is immediate from part (a), and on noting that $\mathbf{X}^{T} Q_{n} \mathbf{X}$ has a continuous limiting distribution in all regimes, as shown in the proofs above (in particular, see (3.34), (3.41), (3.44) for domains $\Theta_{1}, \Theta_{2}, \Theta_{3}$ respectively).
(c) Using (3.8), the existence of MPLE is equivalent to the existence of a real valued root of the equation

$$
\begin{equation*}
\mathbf{X}^{T} Q_{n} \mathbf{X}=\sum_{i=1}^{n} t_{i} \tanh \left(\theta t_{i}\right) \tag{3.50}
\end{equation*}
$$

Taking limits as $\theta \rightarrow \pm \infty$, we get

$$
\lim _{\theta \rightarrow-\infty} \sum_{i=1}^{n} t_{i} \tanh \left(\theta t_{i}\right)=-\sum_{i=1}^{n}\left|t_{i}\right|, \quad \lim _{\theta \rightarrow \infty} \sum_{i=1}^{n} t_{i} \tanh \left(\theta t_{i}\right)=\sum_{i=1}^{n}\left|t_{i}\right| .
$$

Thus the existence of MPLE holds iff

$$
-\sum_{i=1}^{n}\left|t_{i}\right|<\mathbf{X}^{T} Q_{n} \mathbf{X}<\sum_{i=1}^{n}\left|t_{i}\right| .
$$

Suppose $\mathbf{X}^{T} Q_{n} \mathbf{X}=\sum_{i=1}^{n} t_{i}$. This happens iff $X_{i}=1$ for all $i \in S(\mathbf{X})$. Similarly we have

$$
\mathbf{X}^{T} Q_{n} \mathbf{X}=-\sum_{i=1}^{n} t_{i} \Leftrightarrow X_{i}=-1 \text { for all } i \in S(\mathbf{X})
$$

The conclusion of part (c) follows from this.
(d) By symmetry, we only show that

$$
\mathbb{P}_{\theta_{0}, Q_{n}}\left(X_{i}=1, i \in S\right) \rightarrow 0
$$

To this end, note that by mutual contiguity it suffices to show the result under the CurieWeiss model. To this effect, we first claim that for any positive sequence $\left\{\varepsilon_{n}\right\}_{n \geq 1}$ converging to 0 and constant $C>0$ free of $n$, we have

$$
\begin{equation*}
\lim _{n \rightarrow \infty} \sup _{\mathbf{a} \in \mathbb{R}^{n} /\{0\}:\|\mathbf{a}\|_{\infty} \leq \varepsilon_{n}\|\mathbf{a}\|_{2}} \mathbb{P}\left(\sum_{i=1}^{n} a_{i} \xi_{i}=0\right)=0, \tag{3.51}
\end{equation*}
$$

where $\left\{\xi_{i}\right\}_{1 \leq i \leq n}$ are iid random variables such that $\operatorname{Var}\left(\xi_{1}\right) \neq 0, \mathbb{E}\left|\xi_{1}\right|^{3} \leq C$.
Given this claim, we first complete the proof of part (d). Let $\phi_{n}$ be the auxiliary variable introduced in Proposition. Then we have

$$
\mathbb{P}_{\theta_{0}, \mathrm{CW}}\left(t_{i}=0 \text { for some } \mathrm{i}, 1 \leq i \leq n\right)=\mathbb{E P}_{\theta_{0}, \mathrm{CW}}\left(t_{i}=0 \text { for some } \mathrm{i}, 1 \leq i \leq n \mid \phi_{n}\right) .
$$

Given $\phi_{n}, t_{i}$ is a weighted sum of iid random random variables, with $\mathbb{E}\left(\left|X_{1}\right|^{3} \mid \phi_{n}\right) \leq 1$. Also, we have

$$
\left|Q_{n}(i, j)\right| \leq \frac{C_{w}}{n}, \quad \sqrt{\sum_{j=1}^{n} Q_{n}(i, j)^{2}}=\frac{1}{\sqrt{d_{n}}}
$$

and so we can take $C=1, \varepsilon_{n}=\frac{C^{\prime}}{\sqrt{n}}$ for some suitable constant $C^{\prime}$ free of $n$. Thus we have

$$
\mathbb{P}_{\theta_{0}, \mathrm{CW}}\left(t_{i}=0 \text { for some } \mathrm{i}, 1 \leq i \leq n \mid \phi_{n}\right) \xrightarrow{p} 0 \Rightarrow \lim _{n \rightarrow \infty} \mathbb{P}_{\theta_{0}, \mathrm{CW}}\left(t_{i}=0 \text { for some } i\right)=0 .
$$

To complete the proof, it suffices to show that $\mathbb{P}_{\theta_{0}, \mathrm{CW}}(\overline{\mathbf{X}}=\mathbf{1}) \rightarrow 0$. But this is immediate from the weak law of $\overline{\mathbf{X}}$ derived in Lemma 3.3.1 (and mutual contiguity of $\mathbb{P}_{\theta_{0}, Q_{n}}$ and $\mathbb{P}_{\theta_{0}, \mathrm{CW}}$ ).

It thus remains to verify the claim. But this follows on setting $\mu_{n}:=\mathbb{E} \xi_{1}, \tau_{n}^{2}=\operatorname{Var}\left(\xi_{1}\right)$ and noting that

$$
\mathbb{P}\left(\sum_{i=1}^{n} a_{i} \xi_{i}=0\right)=\mathbb{P}\left(\frac{\sum_{i=1}^{n} a_{i}\left(\xi_{i}-\mu_{n}\right)}{\tau_{n}\|\mathbf{a}\|_{2}^{\mathbf{2}}}=-\frac{\mu_{n} \sum_{i=1}^{n} a_{i}}{\tau_{n}\|\mathbf{a}\|_{2}}\right),
$$

and the fact that

$$
\frac{\sum_{i=1}^{n} a_{i}\left(\xi_{i}-\mu_{n}\right)}{\tau_{n}\|\mathbf{a}\|_{2}^{2}} \xrightarrow{d} N(0,1)
$$

by the Lyapunov CLT.

### 3.4 Acknowledgements

We thank Nabarun Deb and Rajarshi Mukherjee for helpful comments throughout this project. We also thank Richard Nickl for suggesting this problem. The second author gratefully acknowledges the support of NSF (DMS-2113414) for partial support during this research.

### 3.5 Appendix A:Proofs of main lemmas

As it turns out, our proof technique relies on a very precise understanding of what happens under the Curie-Weiss model. Denote the Curie-Weiss model by $\mathbb{P}_{\theta, \mathrm{Cw}}$. Let $Q_{n}$ is given as in
(3.31). A simple calculation shows that the probability mass function of the Curie-Weiss model is given by

$$
\mathbb{P}_{\theta, \mathrm{CW}}(\mathbf{X}=\mathbf{x})=\exp \left(\frac{n \theta \bar{x}^{2}}{2}-Z_{n}(\theta, \mathrm{CW})\right)
$$

The following two lemmas summarize the estimates that we need from the Curie-Weiss model. The proof of these lemmas are deferred to section 3.6.

Lemma 3.5.1. For any $h \in \mathbb{R}$ and $\theta_{0}>0$, let $\theta_{n}=\theta_{n}\left(\theta_{0}, h\right)$ be as defined in (3.15). Let $\mathbf{X} \sim$ $\mathbb{P}_{\theta_{n}, \mathrm{CW}}$, where $Q_{n}$ is a sequence of matrices which satisfy (3.2), (3.3), (3.4) and (3.5). Set $B_{n}=$ $Q_{n}-\frac{1}{n} \mathbf{1 1}^{\mathbf{T}}$ as before. Also, let $W_{\theta_{0}} \sim N\left(0, \sigma^{2}\left(\theta_{0}\right)\right), U_{1, h} \sim \mathbb{H}_{h}$ (see (3.18)) be independent of $\left(S_{0}, T_{0}\right)$ (see (3.16) and (3.17) respectively). Then the following conclusions hold under $\mathbb{P}_{\theta_{n}, \mathrm{CW}}$ :
(a) Low Temperature Regime: Suppose $\theta_{0} \in \Theta_{1}$.
(i) With $m$ (.) as in definition 3.2.1, we have

$$
\left(\sqrt{n}\left(\overline{\mathbf{X}}-m\left(\theta_{n}\right)\right), \mathbf{X}^{T} B_{n} \mathbf{X}, \mathbf{X}^{T} B_{n}^{2} \mathbf{X}\right) \xrightarrow{d}\left(W_{\theta_{0}},\left(1-m^{2}\left(\theta_{0}\right)\right) S_{0},\left(1-m^{2}\left(\theta_{0}\right)\right) T_{0}\right) .
$$

(ii) Further we have

$$
\lim _{n \rightarrow \infty}\left\{Z_{n}\left(\theta_{n}, \mathrm{CW}\right)-Z_{n}\left(\theta_{0}, \mathrm{CW}\right)-\frac{\sqrt{n} m^{2}\left(\theta_{0}\right)}{2}\right\}=\frac{R\left(\theta_{0}\right) h^{2}}{2}
$$

where $R\left(\theta_{0}\right)$ is as defined in Theorem 3.2.2.
(b) Critical Point: Suppose $\theta_{0} \in \Theta_{2}$.
(i) We have

$$
\left(n^{1 / 4} \overline{\mathbf{X}}, \mathbf{X}^{T} B_{n} \mathbf{X}, \mathbf{X}^{T} B_{n}^{2} \mathbf{X}\right) \xrightarrow{d}\left(U_{1, h}, S_{0}, T_{0}\right) .
$$

(ii) Further, we have

$$
\lim _{n \rightarrow \infty}\left\{Z_{n}\left(1+h_{n}\right)-Z_{n}(1)\right\}=F(h)-F(0),
$$

where $F($.$) is as defined in (3.18).$
(c) High Temperature Regime: Suppose $\theta_{0} \in \Theta_{3}$.
(i) We have

$$
\left(\sqrt{n} \overline{\mathbf{X}}, \mathbf{X}^{T} B_{n} \mathbf{X}, \mathbf{X}^{T} B_{n}^{2} \mathbf{X}\right) \xrightarrow{d}\left(W_{1}, S_{0}, T_{0}\right) .
$$

(ii) Further we have

$$
\lim _{n \rightarrow \infty}\left\{Z_{n}\left(\theta_{0}+h, \mathrm{CW}\right)-Z_{n}\left(\theta_{0}, \mathrm{CW}\right)\right\}=\frac{1}{2}\left[\log \left(1-\theta_{0}\right)-\log \left(1-\theta_{0}-h\right)\right]
$$

Lemma 3.5.2. Suppose the matrix $Q_{n}$ satisfies (3.2), (3.3), (3.4) and (3.5) for some $C_{W}, \kappa \in(0, \infty)$ and $f \in \mathcal{W}$. Let $\theta_{0}>0, h \in \mathbb{R}$, and $\theta_{n}$ be as defined in (3.15). Then the following conclusions hold in all the three regimes $\Theta_{1} \cup \Theta_{2} \cup \Theta_{3}$.
(a)

$$
\lim _{n \rightarrow \infty} Z_{n}\left(\theta_{n}, Q_{n}\right)-Z_{n}\left(\theta_{n}\right)=C\left(\theta_{0}\right)
$$

where

$$
C\left(\theta_{0}\right):=-\frac{1}{2} \theta\left(1-m^{2}\right)+\frac{\kappa}{4} \theta_{0}^{2}\left(1-m^{2}\right)^{2}+\frac{1}{2} \sum_{i=2}^{\infty}\left[\log \left(1-\theta_{0}\left(1-m^{2}\right) \lambda_{i}\right)+\lambda_{i} \theta_{0}\left(1-m^{2}\right)\right] .
$$

(b) The probability measures $\mathbb{P}_{\theta_{n}, Q_{n}}$ and $\mathbb{P}_{\theta_{n}, \mathrm{CW}}$ are mutually contiguous.

Using Lemma 3.5.1 and Lemma 3.5.2, we now prove the lemmas used to prove the main results in section 3.3.

### 3.5.1 Proof of Lemma 3.3.1

To begin, note that in all regimes of $\theta$, the following hold:

- The log likelihood ratio

$$
\begin{equation*}
\log \frac{d \mathbb{P}_{\theta_{n}, Q_{n}}}{d \mathbb{P}_{\theta_{n}, \mathrm{CW}}}(\mathbf{X})=\frac{\theta_{n}}{2} \mathbf{X}^{T} B_{n} \mathbf{X}-Z_{n}\left(\theta_{n}, Q_{n}\right)+Z_{n}\left(\theta_{n}, \mathrm{CW}\right) \tag{3.52}
\end{equation*}
$$

is a function of $\mathbf{X}^{T} B_{n} \mathbf{X}$.

- The asymptotic non degenerate limiting distribution of $\overline{\mathbf{X}}$ is jointly independent of $\mathbf{X}^{T} B_{n} \mathbf{X}$ and $\mathbf{X}^{T} B_{n}^{2} \mathbf{X}$ (this follows from Lemma 3.5.1).
- The two measures $\mathbb{P}_{\theta_{n}, Q_{n}}$ and $\mathbb{P}_{\theta_{n}, \mathrm{CW}}$ are mutually contiguous (this follows from Lemma 3.5.2 part (b)).

It thus follows from Le-Cam's third lemma that the asymptotic non degenerate distribution of $\overline{\mathbf{X}}$ under $\mathbb{P}_{\theta_{n}, Q_{n}}$ and $\mathbb{P}_{\theta_{n}, \mathrm{CW}}$ are the same, and is asymptotically independent of the joint distribution of $\left(\mathbf{X}^{T} B_{n} \mathbf{X}, \mathbf{X}^{T} B_{n}^{2} \mathbf{X}\right)$ under both models.

To complete the proof of Lemma 3.3.1, it then suffices to show that in all the regimes of $\theta_{0} \in \Theta$, under $\mathbb{P}_{\theta_{n}, Q_{n}}$ we have

$$
\begin{equation*}
\left(\mathbf{X}^{T} B_{n} \mathbf{X}, \mathbf{X}^{T} B_{n}^{2} \mathbf{X}\right) \xrightarrow{d}\left(S_{\theta_{0}}, T_{\theta_{0}}\right) . \tag{3.53}
\end{equation*}
$$

To show this, first note that under $\mathbb{P}_{\theta_{n}, \mathrm{CW}}$ we have

$$
\begin{aligned}
& {\left[\mathbf{X}^{T} B_{n} \mathbf{X}, \mathbf{X}^{T} B_{n}^{2} \mathbf{X}, \log \frac{d \mathbb{P}_{\theta_{n}}, Q_{n}}{d \mathbb{P}_{\theta_{n}, \mathrm{CW}}}(\mathbf{X})\right] } \\
= & {\left[\mathbf{X}^{T} B_{n} \mathbf{X}, \mathbf{X}^{T} B_{n}^{2} \mathbf{X}, \frac{\theta_{n}}{2} \mathbf{X}^{T} B_{n} \mathbf{X}-Z_{n}\left(\theta_{n}, Q_{n}\right)+Z_{n}\left(\theta_{n}, \mathrm{CW}\right)\right] } \\
\stackrel{d}{\rightarrow} & {\left[\left(1-m^{2}\left(\theta_{0}\right)\right) S_{0},\left(1-m^{2}\left(\theta_{0}\right)\right) T_{0}, \frac{\theta_{0}\left(1-m^{2}\left(\theta_{0}\right)\right)}{2} S_{0}-C\left(\theta_{0}\right)\right], }
\end{aligned}
$$

and we have used (3.52) in the first step, and Lemma 3.5.1 and Lemma 3.5.2 part (a) in the second step (and $C\left(\theta_{0}\right)$ is defined in Lemma 3.5.2 part (a)). Using mutual contiguity, it follows that under
$\mathbb{P}_{\theta_{n}, Q_{n}}$, we have

$$
\left(\mathbf{X}^{T} B_{n} \mathbf{X}, \mathbf{X}^{T} B_{n}^{2} \mathbf{X}\right) \xrightarrow{d}\left(S^{\prime}, T^{\prime}\right),
$$

where $\left(S^{\prime}, T^{\prime}\right)$ is a bi-variate random vector with characteristic function

$$
\begin{aligned}
& \mathbb{E} e^{i\left(s S^{\prime}+t T^{\prime}\right)} \\
& =\mathbb{E} \exp \left\{\left(1-m^{2}\left(\theta_{0}\right)\right) i\left(s S_{0}+t T_{0}\right)+\frac{\theta_{0}\left(1-m^{2}\left(\theta_{0}\right)\right)}{2} S_{0}-C\left(\theta_{0}\right)\right\} \\
& =e^{-C\left(\theta_{0}\right)} \mathbb{E} \exp \left\{\left(1-m^{2}\left(\theta_{0}\right)\right)\left(i s+\frac{\theta_{0}}{2}\right) S_{0}+i t T_{0}\right\} \\
& =e^{-C\left(\theta_{0}\right)} \mathbb{E} \exp \left\{\left(1-m^{2}\left(\theta_{0}\right)\right)\left(i s+\frac{\theta_{0}}{2}\right)\left(\sum_{j=2}^{\infty} \lambda_{j}\left(Y_{j}-1\right)-1+W^{*}\right)+i t\left(1-m^{2}\left(\theta_{0}\right)\right)\left(\sum_{j=2}^{\infty} \lambda_{j}^{2} Y_{j}+\kappa\right)\right\} \\
& =e^{-C\left(\theta_{0}\right)} \exp \left\{-\left(1-m^{2}\left(\theta_{0}\right)\right)\left(i s+\frac{\theta_{0}}{2}\right)+i \kappa t\left(1-m^{2}\left(\theta_{0}\right)\right)\right\} \mathbb{E} \exp \left\{\left(1-m^{2}\left(\theta_{0}\right)\right)\left(i s+\frac{\theta_{0}}{2}\right) W^{*}\right\} \\
& \prod_{j=2}^{\infty} \exp \left\{-\left(1-m^{2}\left(\theta_{0}\right)\right)\left(i s+\frac{\theta_{0}}{2}\right) \lambda_{j}\right\} \mathbb{E} \exp \left\{\left(1-m^{2}\left(\theta_{0}\right)\right)\left(i s+\frac{\theta_{0}}{2}\right) \lambda_{j} Y_{j}+\left(1-m^{2}\left(\theta_{0}\right)\right) t \lambda_{j}^{2} Y_{j}\right\} \\
& =e^{-C\left(\theta_{0}\right)} \exp \left\{\left(1-m^{2}\left(\theta_{0}\right)\right)\left(i t \kappa-i s-\frac{\theta_{0}}{2}\right)+\left(1-m^{2}\left(\theta_{0}\right)\right)^{2}\left(i s+\frac{\theta_{0}}{2}\right)^{2} \kappa\right\} \\
& \prod_{j=2}^{\infty} \frac{\exp \left\{-\left(1-m^{2}\left(\theta_{0}\right)\right)\left(i s+\frac{\theta_{0}}{2}\right) \lambda_{j}\right\}}{\sqrt{1-\left(1-m^{2}\left(\theta_{0}\right)\right)\left(\theta_{0} \lambda_{j}+2 i \lambda_{j} s+2 i \lambda_{j}^{2} t\right)}} \\
& =e^{\left(1-m^{2}\left(\theta_{0}\right)\right)(i t \kappa-i s)+\left(1-m^{2}\left(\theta_{0}\right)\right)^{2}\left(i \theta_{0} s \kappa-s^{2} \kappa\right)} \prod_{j=2}^{\infty} \frac{e^{-\left(1-m^{2}\left(\theta_{0}\right)\right)\left(i s+\frac{\theta_{0}}{2}\right) \lambda_{j}}}{\sqrt{1-\left(1-m^{2}\left(\theta_{0}\right)\right)\left(1-2 i\left(1-m^{2}\left(\theta_{0}\right)\right) \frac{\lambda_{j} s+\lambda_{j}^{2} t}{1-\theta_{0}\left(1-m^{2}\left(\theta_{0}\right)\right) \lambda_{j}}\right)}},
\end{aligned}
$$

where in the last step we use the formula for $C\left(\theta_{0}\right)$ from Lemma 3.5.2 part (a). The last display above can be checked to be the joint characteristic function of

$$
\begin{gathered}
\left(1-m^{2}\left(\theta_{0}\right)\right)\left[\sum_{j=2}^{\infty} \lambda_{j}\left(\frac{Y_{j}}{1-\theta_{0}\left(1-m^{2}\left(\theta_{0}\right)\right) \lambda_{j}}-1\right)-1+\left(1-m^{2}\left(\theta_{0}\right)\right) \theta_{0} \kappa+W^{*},\right. \\
\\
\left.\sum_{j=2}^{\infty} \frac{\lambda_{j}^{2} Y_{j}}{1-\theta_{0}\left(1-m^{2}\left(\theta_{0}\right)\right) \lambda_{j}}+\kappa\right]
\end{gathered}
$$

and so the proof of (3.53) is complete.

### 3.5.2 Proof of Lemma 3.3.2

This is immediate on combining Lemma 3.5.1 part (a)(ii), part (b)(ii), part (c)(ii), along with Lemma 3.5.2 part (a).

### 3.6 Appendix B: Proofs of lemmas on Curie-Weiss models

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 [39,51]. We omit the proof.

Proposition 3.6.1. [51, Lem 3] Given $\mathbf{X} \sim \mathbb{P}_{\theta, \mathrm{CW}}$, let $\phi_{n}$ be a real valued random variable defined by

$$
\begin{equation*}
\phi_{n} \sim N\left(\overline{\mathbf{X}}, \frac{1}{n \theta}\right) \tag{3.54}
\end{equation*}
$$

Then the following conclusions hold:
(a) Given $\phi_{n}$, the random variables $\left(X_{1}, \ldots, X_{n}\right)$ are IID, with

$$
\begin{equation*}
\mathbb{P}_{\theta, \mathrm{CW}}\left(X_{j}=1 \mid \phi_{n}\right)=\frac{\exp \left(\theta \phi_{n}\right)}{\exp \left(\theta \phi_{n}\right)+\exp \left(-\theta \phi_{n}\right)} \tag{3.55}
\end{equation*}
$$

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

$$
\begin{equation*}
f_{\theta, n}\left(\phi_{n}\right)=\exp \left(-\frac{1}{2} n \theta \phi_{n}^{2}+n \log \cosh \left(\theta \phi_{n}\right)\right) . \tag{3.56}
\end{equation*}
$$

Definition 3.6.1. Let $F_{n, \theta}$ denote the distribution of $\phi_{n}$, as defined in Proposition 3.6.1.

We first state a lemma about $F_{n, \theta}$, which is the main ingredient to prove Lemma 3.3.1 and Lemma 3.3.2.

Lemma 3.6.1. Fix $\theta_{0}>0, h \in \mathbb{R}$, and let $\theta_{n}$ be as defined in (3.15). Let $\phi_{n} \sim F_{n, \theta_{n}}$, where $F_{n, \theta}$ is as in definition 3.6.1.
(a) If $\theta_{0} \in \Theta_{3}$, then

$$
\sqrt{n} \phi_{n} \rightarrow N\left(0, \frac{1}{\theta_{0}-\theta_{0}^{2}}\right)
$$

in distribution, and in moments.
(b) If $\theta_{0} \in \Theta_{2}$, then

$$
n^{1 / 4} \phi_{n} \rightarrow \mathbb{H}_{h}
$$

in distribution, and in moments, where $H_{h}$ is as defined in (3.18).
(c) If $\theta_{0} \in \Theta_{1}$, then conditional on $\phi_{n}>0$ we have

$$
\sqrt{n}\left(\phi_{n}-m\left(\theta_{n}\right)\right) \rightarrow N\left(0, \frac{1}{\theta_{0}-\left(1-m^{2}\left(\theta_{0}\right)\right) \theta_{0}^{2}}\right)
$$

in distribution, and in moments, where $m($.$) is as in definition 3.2.1.$

Our second lemma characterizes the limit distribution of quadratic forms of IID random variables (by comparing it to quadratic forms of Gaussians).

Lemma 3.6.2. Suppose $\mathbf{Z}:=\left(Z_{i}\right)_{1 \leq i \leq n}$ 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 (3.2), (3.3), (3.4) and (3.5) for some $C_{W}, \kappa \in(0, \infty)$ and $f \in \mathcal{W}$. Then with $B_{n}=Q_{n}-\frac{1}{n} \mathbf{1 1}^{T}$ we have

$$
\left[\sqrt{n} \overline{\mathbf{Z}}, \mathbf{Z}^{T} B_{n} \mathbf{Z}, \mathbf{Z}^{T} B_{n}^{2} \mathbf{Z}\right] \stackrel{d}{\rightarrow}\left[\tau W_{0}, \tau^{2}\left(\sum_{j=2}^{\infty} \lambda_{j}\left(Y_{j}-1\right)-1+W^{*}\right), \tau^{2}\left(\sum_{j=2}^{\infty} \lambda_{j}^{2} Y_{j}+\kappa\right)\right],
$$

where

$$
W_{0} \sim N(0,1), \quad W^{*} \sim N(0,2 \kappa), \quad\left\{Y_{j}\right\}_{j \geq 2} \stackrel{i i d}{\sim} \chi_{1}^{2}
$$

are mutually independent. Here the infinite sums in the limiting distributions converge in $L_{2}$.
The final result that we need is an elementary calculus result.

Proposition 3.6.2. Suppose $g_{n}($.$) is a sequence of functions defined on a compact interval [a, b]$. Assume the following:

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

Then $g_{n}$ converges to $g_{\infty}$ uniformly on $[a, b]$.

The proof of these results are deferred to the appendix C 3.7.

### 3.6.1 Proof of Lemma 3.5.1

(a) Before we begin the proof, we point out that Lemma 3.6.1 part (c) implies that by symmetry,

$$
\phi_{n} \xrightarrow{d} \frac{1}{2}\left(\delta_{m\left(\theta_{0}\right)}+\delta_{-m\left(\theta_{0}\right)}\right),
$$

and Proposition 3.6.1 implies that

$$
\left|\phi_{n}-\overline{\mathbf{X}}\right| \xrightarrow{p} 0 .
$$

The last two displays together imply

$$
\mathbb{P}_{n, \theta_{n}}\left(\phi_{n}<0 \mid \overline{\mathbf{X}}>0\right) \rightarrow 0,
$$

and so without loss of generality we can interchange between the conditioning events $\overline{\mathbf{X}}>0$ and $\phi_{n}>0$. Also, conditional on $\phi_{n}>0$ we have $\phi_{n} \xrightarrow{p} m\left(\theta_{0}\right)$, from Lemma 3.6.1 part (c).
(i) Using Proposition 3.6.1, conditioning on $\phi_{n}$, the random variables $\left(X_{1}, \ldots, X_{n}\right)$ are IID, with

$$
\mathbb{E}\left(X_{1} \mid \phi_{n}\right)=\tanh \left(\theta_{n} \phi_{n}\right)=: \mu_{n}, \quad \operatorname{Var}\left(X_{1} \mid \phi_{n}\right)=\operatorname{sech}^{2}\left(\theta_{n} \phi_{n}\right) .
$$

Noting that $B_{n} \mathbf{1}=\mathbf{0}$, setting $\mu_{n}:=\mu_{n} \mathbf{1}$ one can write

$$
\begin{align*}
& {\left[\sqrt{n}\left(\overline{\mathbf{X}}-m\left(\theta_{n}\right)\right), \mathbf{X}^{T} B_{n} \mathbf{X}, \mathbf{X}^{T} B_{n}^{2} \mathbf{X}\right] } \\
= & {\left[\sqrt{n}\left(\overline{\mathbf{X}}-\mu_{n}\right),\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right)^{T} B_{n}\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right),\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right)^{T} B_{n}^{2}\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right)\right] }  \tag{3.57}\\
+ & {\left[\sqrt{n}\left(\mu_{n}-m\left(\theta_{n}\right)\right), 0,0\right] . }
\end{align*}
$$

By Lemma 3.6.2, conditioning on $\phi_{n}$, on the event $\phi_{n}>0$ we get

$$
\begin{align*}
& {\left[\sqrt{n}\left(\overline{\mathbf{X}}-\mu_{n}\right),\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right)^{T} B_{n}\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right),\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right)^{T} B_{n}^{2}\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right)\right] }  \tag{3.58}\\
& \stackrel{d}{\rightarrow}\left[\tau W_{0}, \tau^{2} S_{0}, \tau^{2} T_{0}\right],
\end{align*}
$$

where

$$
\operatorname{sech}\left(\theta_{n} \phi_{n}\right) \xrightarrow{p} \operatorname{sech}\left(\theta_{0} m\left(\theta_{0}\right)\right)=\sqrt{1-m^{2}\left(\theta_{0}\right)}=: \tau
$$

Proceeding to analyze the second term in the RHS of (3.57), a one term Taylor's expansion then gives that

$$
\begin{aligned}
\sqrt{n}\left(\mu_{n}-m\left(\theta_{n}\right)\right) & =\sqrt{n}\left(\tanh \left(\theta_{n} \phi_{n}\right)-\tanh \left(\theta_{n} m\left(\theta_{n}\right)\right)\right) \\
& =\sqrt{n}\left(\phi_{n}-m\left(\theta_{n}\right)\right) \theta_{n} \operatorname{sech}^{2}\left(\theta_{n} \xi_{n}\right),
\end{aligned}
$$

where $\xi_{n}$ lies between $\phi_{n}$ and $m\left(\theta_{n}\right)$, and hence converges to $m\left(\theta_{0}\right)$ in probability. Along with the above display, this gives that conditional on $\phi_{n}>0$ we have

$$
\begin{aligned}
\sqrt{n}\left(\mu_{n}-m\left(\theta_{n}\right)\right) & \stackrel{d}{\rightarrow} \theta_{0} \operatorname{sech}^{2}\left(\theta_{0} m\left(\theta_{0}\right)\right) N\left(0, \frac{1}{\theta_{0}-\left(1-m^{2}\left(\theta_{0}\right)\right) \theta_{0}^{2}}\right) \\
& =N\left(0, \frac{\theta_{0}\left(1-m^{2}\left(\theta_{0}\right)\right)^{2}}{1-\left(1-m^{2}\left(\theta_{0}\right)\right) \theta_{0}}\right) .
\end{aligned}
$$

Combining the last display along with (3.57) and (3.58), the conclusion of part (a)
follows, on noting that

$$
1-m^{2}\left(\theta_{0}\right)+\frac{\theta_{0}\left(1-m^{2}\left(\theta_{0}\right)\right)^{2}}{1-\left(1-m^{2}\left(\theta_{0}\right)\right) \theta_{0}}=\frac{1-m^{2}\left(\theta_{0}\right)}{1-\theta_{0}\left(1-m^{2}\left(\theta_{0}\right)\right)}=\sigma^{2}\left(\theta_{0}\right)
$$

(ii) To begin, use Proposition 3.6.1 to get

$$
\begin{equation*}
\phi_{n}=\overline{\mathbf{X}}+\frac{W_{0}}{\sqrt{n \theta_{n}}} \tag{3.59}
\end{equation*}
$$

where $W_{0} \sim N(0,1)$ is independent of $\mathbf{X}$. This gives

$$
\begin{equation*}
\sqrt{n}\left(\overline{\mathbf{X}}-m\left(\theta_{n}\right)\right)=\sqrt{n}\left(\phi_{n}-m\left(\theta_{n}\right)\right)-\frac{W_{0}}{\sqrt{n \theta_{n}}} . \tag{3.60}
\end{equation*}
$$

which along with part (c) of Lemma 3.6.1 shows that conditional on $\overline{\mathbf{X}}>0$ we have

$$
\sqrt{n}\left(\overline{\mathbf{X}}-m\left(\theta_{n}\right)\right) \xrightarrow{d} N\left(0, \frac{1-m^{2}\left(\theta_{0}\right)}{1-\theta_{0}\left(1-m^{2}\left(\theta_{0}\right)\right)}\right)=N\left(0, \sigma^{2}\left(\theta_{0}\right)\right) .
$$

Thus in turn implies that unconditionally, we have

$$
\sqrt{n}\left(\overline{\mathbf{X}}^{2}-m^{2}\left(\theta_{n}\right)\right) \xrightarrow{d} N\left(0,4 \sigma^{2}\left(\theta_{0}\right) m^{2}\left(\theta_{0}\right)\right) .
$$

Using Proposition 3.3.2, we then have

$$
\frac{\sqrt{n}}{2}\left(\overline{\mathbf{X}}^{2}-m^{2}\left(\theta_{0}\right)\right) \xrightarrow{d} N\left(m^{2}\left(\theta_{0}\right) \sigma^{2}\left(\theta_{0}\right) h, m^{2}\left(\theta_{0}\right) \sigma^{2}\left(\theta_{0}\right)\right),
$$

which gives

$$
\begin{aligned}
\frac{Z_{n}^{\prime}\left(\theta_{0}+h / \sqrt{n}\right)}{\sqrt{n}}-\frac{\sqrt{n} m^{2}\left(\theta_{0}\right)}{2} & =\frac{\sqrt{n}}{2}\left(\mathbb{E}_{\mathbb{P}_{\theta_{n}, \mathrm{CW}}} \overline{\mathbf{X}}^{2}-m^{2}\left(\theta_{0}\right)\right) \\
& \rightarrow \mathbb{E} N\left(m^{2}\left(\theta_{0}\right) \sigma^{2}\left(\theta_{0}\right) h, m^{2}\left(\theta_{0}\right) \sigma^{2}\left(\theta_{0}\right)\right)^{2} \\
& =m^{2}\left(\theta_{0}\right) \sigma^{2}\left(\theta_{0}\right) h=R\left(\theta_{0}\right) h .
\end{aligned}
$$

In the line above, we have used the fact that $\sqrt{n}\left(\overline{\mathbf{X}}^{2}-m^{2}\left(\theta_{0}\right)\right)$ is uniformly integrable. But this follows on using (3.60), along with the fact that $\sqrt{n}\left(\phi_{n}-m\left(\theta_{n}\right)\right)$ is uniformly integrable (from Lemma 3.6.1 part (c)).

The convergence in the above display holds for all $h$ fixed. Integrating both sides over the interval $[0, h]$ we get

$$
Z_{n}\left(\theta_{0}+h / \sqrt{n}\right)-Z_{n}\left(\theta_{0}, \mathrm{CW}\right)-\frac{\sqrt{n} h m^{2}\left(\theta_{0}\right)}{2} \rightarrow \frac{R\left(\theta_{0}\right) h^{2}}{2}
$$

as desired. In the last convergence above, we use the fact that the function $h \mapsto$ $\frac{Z_{n}^{\prime}\left(\theta_{0}+h / \sqrt{n}\right)}{\sqrt{n}}$ is monotone, and hence converges uniformly over compact sets, by Proposition 3.6.2. This completes the proof of part (a).
(b) (i) Again using calculations similar to (3.57), we get

$$
\begin{aligned}
& {\left[n^{1 / 4} \overline{\mathbf{X}}, \mathbf{X}^{T} B_{n} \mathbf{X}, \mathbf{X}^{T} B_{n}^{2} \mathbf{X}\right] } \\
= & {\left[n^{1 / 4}\left(\overline{\mathbf{X}}-\mu_{n}\right),\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right)^{T} B_{n}\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right),\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right)^{T} B_{n}^{2}\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right)\right]+\left[n^{1 / 4} \mu_{n}, 0,0\right] }
\end{aligned}
$$

Conditioning on $\phi_{n}$, using Proposition 3.6.1 and Lemma 3.6.2 we have

$$
\left[n^{1 / 4}\left(\overline{\mathbf{X}}-\mu_{n}\right),\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right)^{T} B_{n}\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right),\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right)^{T} B_{n}^{2}\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right)\right] \xrightarrow{d}\left[0, S_{0}, T_{0}\right],
$$

where we use the fact

$$
\operatorname{Var}\left(X_{1} \mid \phi\right)=\operatorname{sech}^{2}\left(\theta_{n} \phi_{n}\right) \xrightarrow{p} 1 .
$$

Also, an application of delta theorem along with Lemma 3.6.1 part (b) gives

$$
n^{1 / 4} \mu_{n}=n^{1 / 4} \tanh \left(\phi_{n}\right) \xrightarrow{d} U_{1 . h} .
$$

Combining the above, it follows that

$$
\left[n^{1 / 4} \overline{\mathbf{X}}, \mathbf{X}^{T} B_{n} \mathbf{X}, \mathbf{X}^{T} B_{n}^{2} \mathbf{X}\right] \xrightarrow{d}\left[U_{1, h}, S_{0}, T_{0}\right]
$$

as desired.
(ii) Using (3.59) we get

$$
n^{1 / 4} \overline{\mathbf{X}}=n^{1 / 4} \phi_{n}-n^{1 / 4} \frac{W_{0}}{\sqrt{n \theta_{0}}} \xrightarrow{d} U_{1, h},
$$

where we have used Lemma 3.6.1 part (b). This in turn gives,

$$
\sqrt{n} \overline{\mathbf{X}}^{2} \xrightarrow{d} U_{1, h}^{2} .
$$

and so

$$
\frac{Z_{n}^{\prime}\left(\theta_{n}\right)}{\sqrt{n}}=\frac{\sqrt{n}}{2} \mathbb{E}_{\mathbb{P}_{\theta_{n}, \mathrm{CW}}} \overline{\mathbf{X}}^{2} \rightarrow \frac{1}{2} \mathbb{E} U_{1, h}^{2}=F^{\prime}(h) .
$$

Again in the last step we use the fact that $\sqrt{n} \bar{X}^{2}$ is uniformly integrable, which follows from (3.59) and Lemma 3.6.1 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 3.6.2.
(c) (i) Using calculations similar to (3.57), we get

$$
\begin{aligned}
& {\left[\sqrt{n} \overline{\mathbf{X}}, \mathbf{X}^{T} B_{n} \mathbf{X}, \mathbf{X}^{T} B_{n}^{2} \mathbf{X}\right] } \\
= & {\left[\sqrt{n}\left(\overline{\mathbf{X}}-\mu_{n}\right),\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right)^{T} B_{n}\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right),\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right)^{T} B_{n}^{2}\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right)\right]+\left[\sqrt{n} \mu_{n}, 0,0\right] . }
\end{aligned}
$$

Conditioning on $\phi_{n}$, using Proposition 3.6.1 and Lemma 3.6.2 we have

$$
\left[\sqrt{n}\left(\overline{\mathbf{X}}-\mu_{n}\right),\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right)^{T} B_{n}\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right),\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right)^{T} B_{n}^{2}\left(\mathbf{X}-\boldsymbol{\mu}_{n}\right)\right] \xrightarrow{d}\left[W_{0}, S_{0}, T_{0}\right]
$$

where $W_{0} \sim N(0,1)$ is independent of $\left(S_{0}, T_{0}\right)$ In the above display, we again use

$$
\operatorname{Var}\left(X_{1} \mid \phi\right)=\operatorname{sech}^{2}\left(\theta_{n} \phi_{n}\right) \xrightarrow{p} 1 .
$$

Also, an application of delta theorem along with Lemma 3.6.1 part (b) gives

$$
\sqrt{n} \mu_{n}=\sqrt{n} \tanh \left(\theta_{n} \phi_{n}\right) \xrightarrow{d} \theta_{0} N\left(0, \frac{1}{\theta_{0}\left(1-\theta_{0}\right)}\right)=N\left(0, \frac{\theta_{0}}{1-\theta_{0}}\right) .
$$

Combining the above, the desired conclusion follows on noting that

$$
\frac{\theta_{0}}{1-\theta_{0}}+1=\frac{1}{1-\theta_{0}}=\sigma^{2}\left(\theta_{0}\right)
$$

(ii) As before, using (3.59) along with Lemma 3.6.1 part (a) we get

$$
\sqrt{n} \overline{\mathbf{X}} \xrightarrow{d} N\left(0, \frac{1}{1-\theta_{0}}\right), \text { which gives } n \overline{\mathbf{X}}^{2} \xrightarrow{d} \frac{\chi_{1}^{2}}{1-\theta_{0}} .
$$

The desired conclusion follows from this using similar calculations as above, and using uniform integrability along with Lemma 3.6.1 part (a) and Proposition 3.6.2.

### 3.6.2 Proof of Lemma 3.5.2

(a) To begin, note that

$$
\begin{aligned}
\exp \left(Z_{n}\left(\theta_{n}, Q_{n}\right)-Z_{n}\left(\theta_{n}, \mathrm{CW}\right)\right) & =\frac{\sum_{\mathbf{x} \in\{-1,1\}^{n}} e^{\frac{\theta_{n}}{2} \mathbf{X}^{T} Q_{n} \mathbf{X}}}{\sum_{\mathbf{x} \in\{-1,1\}^{n}} e^{\frac{\theta_{n}}{2} n \overline{\mathbf{X}}^{2}}} \\
& =\mathbb{E}_{\theta_{n}, \mathrm{CW}} e^{\frac{\theta_{n}}{2} \mathbf{X}^{T} B_{n} \mathbf{X}}
\end{aligned}
$$

It follows from Lemma 3.5.1 that if $\mathbf{X} \sim \mathbb{P}_{\theta_{n}, \mathrm{CW}}$, then for all $\theta_{0} \in \Theta_{0}$ we have

$$
\mathbf{X}^{T} B_{n} \mathbf{X} \xrightarrow{d}\left(1-m^{2}\left(\theta_{0}\right)\right) S_{0} \Rightarrow e^{\frac{\theta_{n}}{2}} \mathbf{X}^{T} B_{n} \mathbf{X} \xrightarrow{d} e^{\frac{\theta_{0}\left(1-m^{2}\left(\theta_{0}\right)\right)}{2} S_{0}} .
$$

Assume now that there exists $\delta>0$ such that

$$
\begin{equation*}
\mathbb{E}_{\theta_{n}, \mathrm{CW}} e^{\frac{(1+\delta) \theta_{n}}{2}} \mathbf{X}^{T} B_{n} \mathbf{X} \tag{3.61}
\end{equation*}
$$

Uniform integrability then gives

$$
\mathbb{E}_{\theta_{n}, \mathrm{CW}} e^{\frac{\theta_{n}}{2} \mathbf{X}^{T} B_{n} \mathbf{X}} \rightarrow \mathbb{E} e^{\frac{\theta_{0}\left(1-m^{2}\left(\theta_{0}\right)\right)}{2} S_{0}},
$$

which equals $C\left(\theta_{0}\right)$ using the formula for $S_{0}$ (see (3.16)).

It thus remains to verify (3.61). To this effect, note that

$$
\mathbb{E}_{\theta_{n}, \mathrm{CW}} e^{\frac{\theta_{n}}{2} \mathbf{X}^{T} B_{n} \mathbf{x}}=\mathbb{E}_{\theta_{n}, \mathrm{CW}}\left(\left.e^{\left.\frac{\theta_{n}}{2} \mathbf{X}-\mu_{n}\right)^{T} B_{n}\left(\mathbf{x}-\mu_{n}\right)} \right\rvert\, \phi_{n}\right),
$$

where $\boldsymbol{\mu}_{n}=\mu_{n} \mathbf{1}$ with $\mu_{n}=\tanh \left(\theta_{n} \phi_{n}\right)$, as in the proof of Lemma 3.5.1. Invoking Lemma 3.6.1, we have that given $\phi_{n}$ the random variables $\left(X_{1}, \ldots, X_{n}\right)$ are IID with mean $\mu_{n}$. Also,
setting

$$
\begin{array}{rlrl}
s_{\mu} & :=\frac{2 \mu}{\log (1+\mu)-\log (1-\mu)} & \text { if } \mu \neq 0, \\
& = & 1 & \\
\text { if } \mu=0
\end{array}
$$

we have that $s$. is a strictly positive continuous even function, with $s_{\mu_{n}} \xrightarrow{p} s_{m\left(\theta_{0}\right)}=\frac{1}{\theta_{0}}$ in all regimes, where the last equality uses the fact that $m\left(\theta_{0}\right)=\tanh \left(\theta_{0} m\left(\theta_{0}\right)\right)$. Since

$$
\limsup _{n \rightarrow \infty} \lambda_{1}\left(B_{n}\right)=\lambda_{2}<1
$$

by (3.5), there exists $\delta>0$ such that on the set $\left|\left|\mu_{n}\right|-m\left(\theta_{0}\right)\right|>\delta$ we have

$$
\limsup _{n \rightarrow \infty} \theta_{n} \lambda_{1}\left(B_{n}\right) s_{\mu_{n}}<1
$$

Thus using [19, Prop 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 $\left|\left|\mu_{n}\right|-m\right|>\delta$ we have

$$
\log \mathbb{E}_{\theta_{n}, \mathrm{CW}}\left(\left.e^{\frac{\theta_{n}}{2}\left(\mathbf{X}-\mu_{n}\right)^{T} B_{n}\left(\mathbf{X}-\mu_{n}\right)} \right\rvert\, \phi_{n}\right) \leq C .
$$

To complete the proof of (3.61), it suffices to show that

$$
\limsup _{n \rightarrow \infty} \mathbb{E}_{\theta_{n}, \mathrm{CW}} e^{\frac{\theta_{n}}{2} \mathbf{X}^{T} B_{n} \mathbf{X}_{1}} 1\left\{\left|\mu_{n}\right|-m\left(\theta_{0}\right) \mid>\delta\right\}<\infty,
$$

for some $\delta>0$. But this follows from [19, Lem 4.2] on using

$$
N=n, \quad V_{N}=\theta_{n} \mathbf{X}^{T} B_{n} \mathbf{X} .
$$

The statement of this lemma does not include $\Theta_{3}=(0,1)$, but the proof applies verbatim in all regimes, and allows for $\theta=\theta_{n}$ depending on $n$.
(b) The likelihood ratio between $\mathbb{P}_{\theta_{n}, Q_{n}}$ and $\mathbb{P}_{\theta_{n}}$ is given by

$$
\frac{d \mathbb{P}_{\theta_{n}, Q_{n}}}{d \mathbb{P}_{\theta_{n}, \mathrm{CW}}}(\mathbf{X})=\exp \left(\frac{\theta_{n}}{2} \mathbf{X}^{T} B_{n} \mathbf{X}-Z_{n}\left(\theta_{n}, Q_{n}\right)+Z_{n}\left(\theta_{n}, \mathrm{CW}\right)\right)
$$

Using Lemma 3.5.1, for all $\theta_{0} \in \Theta$ we have

$$
\mathbf{X}^{T} B_{n} \mathbf{X} \xrightarrow{d}\left(1-m^{2}\left(\theta_{0}\right)\right) S_{0} .
$$

Also, using part (a) we have

$$
Z_{n}\left(\theta_{n}, Q_{n}\right)-Z_{n}\left(\theta_{n}, \mathrm{CW}\right) \rightarrow \log \mathbb{E} e^{\frac{\theta_{0}\left(1-m^{2}\left(\theta_{0}\right)\right)}{2} S_{0}}
$$

Combining, if $\mathbf{X} \sim \mathbb{P}_{\theta_{n}}, \mathrm{CW}$, then we have

$$
\frac{d \mathbb{P}_{\theta_{n}, Q_{n}}}{d \mathbb{P}_{\theta_{n}, \mathrm{CW}}}(\mathbf{X}) \xrightarrow{d} \frac{e^{\frac{\theta_{0}\left(1-m^{2}\left(\theta_{0}\right)\right)}{2}} S_{0}}{\mathbb{E} e^{\frac{\theta_{0}\left(1-m^{2}\left(\theta_{0}\right)\right)}{2}} S_{0}}
$$

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.

### 3.7 Appendix C: Proofs of supporting lemmas

### 3.7.1 Proof of Proposition 3.3.2

We prove the more general result

$$
\lim _{h \rightarrow 0} \frac{m\left(\theta_{0}+h\right)-m\left(\theta_{0}\right)}{h}=m\left(\theta_{0}\right) \sigma^{2}\left(\theta_{0}\right) .
$$

The desired conclusion then follows on replacing $h$ by $\frac{h}{\sqrt{n}}$, and letting $n \rightarrow \infty$. 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

$$
\frac{\partial w(\theta, m)}{\partial \theta}=1-\theta \operatorname{sech}^{2}(\theta m)
$$

By Proposition 3.3.1, we have $\theta\left(1-m^{2}(\theta)\right)<1$, 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

$$
m^{\prime}(\theta)=\operatorname{sech}^{2}(\theta m(\theta))\left[m(\theta)+\theta m^{\prime}(\theta)\right]=\left(1-m^{2}(\theta)\right)\left[m(\theta)+\theta m^{\prime}(\theta)\right]
$$

Solving for $m^{\prime}(\theta)$ gives

$$
m^{\prime}(\theta)=\frac{m(\theta)\left(1-m^{2}(\theta)\right.}{1-\theta\left(1-m^{2}(\theta)\right)}=m(\theta) \sigma^{2}(\theta)
$$

as desired.

### 3.7.2 Proof of Proposition 3.6.2

Let $\left\{t_{n}\right\}_{n \geq 1}$ be a real sequence in $[a, b]$ converging to $t_{\infty}$. We need to show that $g_{n}\left(t_{n}\right)$ converges to $g_{\infty}\left(t_{\infty}\right)$. To this effect, fixing $\delta>0$ arbitrary, for all $n$ large we have $\left|t_{n}-t_{\infty}\right|<\delta$. Using
the monotonicity of $\left\{g_{n}\right\}_{1 \leq n \leq \infty}$ gives

$$
g_{n}\left(t_{n}\right)-g_{\infty}\left(t_{\infty}\right) \leq g_{n}(t-\delta)-g_{\infty}\left(t_{\delta}\right) .
$$

Taking limits as $n \rightarrow \infty$ gives

$$
\limsup _{n \rightarrow \infty}\left\{g_{n}\left(t_{n}\right)-g_{\infty}\left(t_{\infty}\right)\right\} \leq g_{\infty}\left(t_{\infty}+\delta\right)=g_{\infty}(\delta)
$$

Since $\delta$ is arbitrary, letting $\delta \downarrow 0$ and using the continuity of $g_{\infty}($.$) gives$

$$
\limsup _{n \rightarrow \infty}\left\{g_{n}\left(t_{n}\right)-g_{\infty}\left(t_{\infty}\right)\right\} \leq 0 .
$$

A similar proof gives

$$
\liminf _{n \rightarrow \infty}\left\{g_{n}\left(t_{n}\right)-g_{\infty}\left(t_{\infty}\right)\right\} \geq 0
$$

The proof is complete by combining the last two displays.

### 3.7.3 Proof of Lemma 3.6.1

(a) Proof of part (a): High Temperature Regime $\Theta_{3}$. Recall from Proposition 3.6.1 that $\phi_{n}$ has a density proportional to

$$
\begin{equation*}
f_{\theta_{n}, n}(\phi)=\exp \left\{-n q_{\theta_{n}}(\phi)\right\}, \quad q_{\theta}(\phi)=\frac{1}{2} \theta \phi^{2}-\log \cosh (\theta \phi) . \tag{3.62}
\end{equation*}
$$

Differentiating twice we get

$$
\theta_{n} \geq q_{\theta_{n}}^{\prime \prime}(\phi) \geq \theta_{n}-\theta_{n}^{2}
$$

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

$$
c_{1} \leq q_{\theta_{n}}^{\prime \prime}(\phi) \leq c_{2} .
$$

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

$$
\mathbb{P}\left(\sqrt{n}\left|\phi_{n}\right|>K\right) \leq \frac{2 \int_{K}^{\infty} e^{-n q_{\theta_{n}}(\phi)} d \phi}{\int_{-\infty}^{\infty} e^{-n q_{\theta_{n}}(\phi)} d \phi} \leq \frac{2 \int_{K}^{\infty} e^{-c_{1} \phi^{2} / 2} d \phi}{\int_{-\infty}^{\infty} e^{-c_{2} \phi^{2} / 2} d \phi} \leq 2 \sqrt{\frac{c_{2}}{c_{1}}} \mathbb{P}\left(N\left(0, \frac{1}{c_{1}}\right)>K\right) .
$$

Thus we have $\sqrt{n} \phi_{n}=O_{p}(1)$, and further all moments of $\sqrt{n} \phi_{n}$ are bounded. Finally, since $q_{\theta_{n}}^{\prime \prime}(\phi) \rightarrow \theta_{0}-\theta_{0}^{2}$, it follows from standard calculus that for any $a, b$ fixed, standard calculus gives

$$
\sqrt{n} \int_{a / \sqrt{n}}^{b / \sqrt{n}} e^{-n q \theta_{n}(\phi)} d \phi \rightarrow \int_{a}^{b} e^{-\frac{\theta_{0}-\theta_{0}^{2}}{2} t^{2}} d t .
$$

Combining the above calculations, it follows that

$$
\sqrt{n} \phi_{n} \rightarrow N\left(0, \frac{1}{\theta_{0}-\theta_{0}^{2}}\right),
$$

in distribution and in moments.
(b) In this case we have

$$
q_{\theta_{n}}^{\prime}(0)=q_{\theta_{n}}^{\prime \prime \prime}(0)=0, \quad \sqrt{n} q_{\theta_{n}}^{\prime \prime}(0)=\sqrt{n}\left(\theta_{n}-\theta_{n}^{2}\right) \rightarrow-h, \quad c_{1}^{\prime} \leq \inf _{|\phi| \leq 2} q_{\theta_{n}}^{\prime \prime \prime \prime}(\phi) \leq \sup _{|\phi| \leq 2} q_{\theta_{n}}^{\prime \prime \prime \prime}(\phi) \leq c_{2}^{\prime},
$$

for some constants $c_{1}^{\prime}, c_{2}^{\prime}$ depending only on $h$. Also, using Proposition 3.6.1 we have

$$
\mathbb{P}\left(\left|\phi_{n}\right|>2\right) \leq \mathbb{P}\left(|N(0,1)|>\sqrt{n \theta_{n}}\right),
$$

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

Finally, fixing $a, b$, straight-forward calculus gives

$$
\int_{a / n^{1 / 4}}^{b / n^{1 / 4}} n^{1 / 4} e^{-n q_{\theta_{n}}(\phi)} d \phi \rightarrow \int_{a}^{b} e^{\frac{h}{2} \phi^{2}-\phi^{4} / 12} d \phi,
$$

where we use the fact that $q_{\theta_{n}}^{\prime \prime \prime \prime}(0) \rightarrow 2$. Combining, the desired limiting distribution follows. Uniform integrability also follows from the estimates on $q_{\theta_{n}}($.$) .$
(c) In this case we have $\mathbb{P}\left(\phi_{n}>0\right)=\mathbb{P}\left(\phi_{n}<0\right)=\frac{1}{2}$ 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 3.6.1 we have $q^{\prime \prime}(m(\theta))>0$, and so the function $\Psi:[0,2] \times[0,1]$ defined by

$$
\begin{aligned}
\Psi(x, \theta) & :=\frac{q_{\theta}(x)-q_{\theta}(m(\theta))}{(x-m(\theta))^{2}} \text { if } x \neq m(\theta), \\
& =\frac{q_{\theta}^{\prime \prime}(m(\theta))}{2} \text { if } x=m(\theta)
\end{aligned}
$$

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

$$
\frac{c_{1}}{2} \phi^{2} \leq q_{\theta_{n}}(\phi)-q_{\theta_{n}}\left(m\left(\theta_{n}\right)\right) \leq \frac{c_{2}}{2} \phi^{2} .
$$

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

$$
q_{\theta_{n}}^{\prime \prime}\left(m\left(\theta_{n}\right)\right) \xrightarrow{p} \theta_{0}-\theta_{0}^{2}\left(1-m^{2}\left(\theta_{0}\right)\right) .
$$

### 3.7.4 Proof of Lemma 3.6.2

We first state the following proposition connecting eigenvalues of the matrix $Q_{n}$ and eigenvalues of the limiting graphon $f$.

Proposition 3.7.1. Let $\left\{Q_{n}\right\}_{n=1}^{\infty}$ be a sequence of matrices satisfying (3.2), (3.3), (3.4) and (3.5)
for some $C_{W}, \kappa \in(0, \infty)$ and $f \in \mathcal{W}$. Let $\left\{\lambda_{j, n}\right\}_{j=1}^{n}$ denote the eigenvalues of $Q_{n}$ arranged in decreasing order of absolute value, and let $\left\{\lambda_{j}\right\}_{j \geq 1}$ be the eigenvalues of the operator $T_{f}$ as defined in section 3.2.2. Then the following conclusions hold:
(a)

$$
\sum_{j=1}^{\infty} \lambda_{j}^{2}=\int_{[0,1]^{2}} f(x, y)^{2} d x d y=\|f\|_{2}^{2}<\infty
$$

(b) For any $j \in \mathbb{N}$,

$$
\lim _{n \rightarrow \infty} \lambda_{j, n}=\lambda_{j} .
$$

(c) For any i $\geq 3$ we have

$$
\lim _{n \rightarrow \infty} \sum_{j=2}^{n} \lambda_{j, n}^{i}=\sum_{j=2}^{\infty} \lambda_{j}^{i} .
$$

The proof of Proposition 3.7.1 part (a) follows [38, Chapter 7.5], whereas part (b) and (c) follow from [38, Theorem 11.54].

The proof of Lemma 3.6.2 will be completed, once we can show the following two steps:
(a) Suppose $\left\{R_{i}\right\}_{1 \leq i \leq n}$ are IID $N(0,1)$. Then the desired conclusion holds.
(b) For any positive integers $a, b, c$ we have

$$
\mathbb{E}(\sqrt{n} \overline{\mathbf{Z}})^{a}\left(\mathbf{Z}^{T} B_{n} \mathbf{Z}\right)^{b}\left(\mathbf{Z}^{T} B_{n}^{2} \mathbf{Z}\right)^{c}-\mathbb{E}(\sqrt{n} \overline{\mathbf{R}})^{a}\left(\mathbf{R}^{T} B_{n} \mathbf{R}\right)^{b}\left(\mathbf{R}^{T} B_{n}^{2} \mathbf{R}\right)^{c} \rightarrow 0
$$

Proof of (a). Let

$$
Q_{n}=P^{T} \Lambda P=\sum_{i=1}^{n} \lambda_{i, n} \mathbf{p}_{i} \mathbf{p}_{i}^{T}
$$

be the spectral decomposition of $Q_{n}$, where the eigenvalues $\left\{\lambda_{i, n}\right\}_{1 \leq i \leq n}$ are arranged in decreasing order of absolute values. Thus we have $\lambda_{1, n}=1$, and $\mathbf{p}_{1}=\frac{1}{\sqrt{n}} \mathbf{1}$. Then setting $\widetilde{\mathbf{R}}:=P \mathbf{R}$ we have

$$
\left[\sqrt{n} \overline{\mathbf{R}}, \mathbf{R}^{T} B_{n} \mathbf{R}, \mathbf{R}^{T} B_{n}^{2} \mathbf{R}\right]=\left[\widetilde{R}_{1}, \sum_{i=2}^{n} \lambda_{i, n} \widetilde{R}_{i}^{2}, \sum_{i=2}^{n} \lambda_{i, n}^{2} \widetilde{R}_{i}^{2}\right] .
$$

Since $\widetilde{\mathbf{R}} \stackrel{d}{=} \mathbf{R}$, it suffices to find the limiting distribution of

$$
\left[R_{1}, \sum_{i=2}^{n} \lambda_{i, n} R_{i}^{2}, \sum_{i=2}^{n} \lambda_{i, n}^{2} R_{i}^{2}\right] .
$$

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

$$
\begin{align*}
& \log \mathbb{E} \exp \left\{s \sum_{j=2}^{n} \lambda_{j, n} R_{j}^{2}+t \sum_{j=2}^{n} \lambda_{j, n}^{2} R_{i}^{2}\right\} \\
= & -\frac{1}{2} \sum_{j=2}^{n} \log \left[1-2\left(\lambda_{j, n} s+\lambda_{j, n}^{2} t\right)\right] \\
= & \frac{1}{2} \sum_{j=2}^{n} \sum_{i=1}^{\infty} \frac{2^{i}\left(\lambda_{j, n} s+\lambda_{j, n}^{2} t\right)^{i}}{i} \\
= & {\left[-s+t \sum_{j=2}^{n} \lambda_{j, n}^{2}\right]+\frac{1}{2} \sum_{j=2}^{n} \sum_{i=2}^{\infty} \frac{2^{i}\left(\lambda_{j, n} s+\lambda_{j, n}^{2} t\right)^{i}}{i} } \\
= & {\left[-s+t \sum_{j=2}^{n} \lambda_{j, n}^{2}\right]+\frac{1}{2} \sum_{i=2}^{\infty} \sum_{j=2}^{n} \frac{2^{i}\left(\lambda_{j, n} s+\lambda_{j, n}^{2} t\right)^{i}}{i}, } \tag{3.63}
\end{align*}
$$

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

$$
\begin{equation*}
2^{i}\left(\lambda_{j, n} s+\lambda_{j, n}^{2} t\right)^{i} \leq 4^{i}\left|\lambda_{j, n}\right|^{i} 8^{-i} \leq 2^{-i} \lambda_{j, n}^{2} . \tag{3.64}
\end{equation*}
$$

For every fixed $i \geq 2$, uses Proposition 3.7.1 part (b) we have

$$
\sum_{j=2}^{n}\left(\lambda_{j, n} s+\lambda_{j, n}^{2} t\right)^{i} \rightarrow \sum_{j=2}^{\infty}\left(\lambda_{j} s+\lambda_{j}^{2} t\right)^{i}
$$

Also, using (3.64) we have

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

where

$$
\sum_{j=2}^{n} \lambda_{j, n}^{2} \rightarrow \sum_{j=2}^{\infty} \lambda_{j}^{2}<\infty
$$

by Proposition 3.7.1 part (c). Combining the last three displays along with dominated convergence theorem, the RHS of (3.63) converges to

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

This is the log moment generating function of

$$
\left(\sum_{j=2}^{\infty} \lambda_{j}\left(Y_{j}-1\right)-1+W^{*}, \sum_{j=2}^{\infty} \lambda_{j}^{2} Y_{j}+\kappa\right)
$$

The random variables in the RHS above converge in $L_{2}$, as

$$
\begin{aligned}
& \mathbb{E}\left[\sum_{j=k+1}^{\infty} \lambda_{j}\left(Y_{j}-1\right)\right]^{2}=2 \sum_{j=k+1}^{\infty} \lambda_{j}^{2} \xrightarrow{k \rightarrow \infty} 0, \\
& \mathbb{E}\left[\sum_{j=k+1}^{\infty} \lambda_{j}^{2} Y_{j}\right]^{2} \leq 3 \sum_{j=k+1}^{\infty} \lambda_{j}^{4}+\left(\sum_{j=k+1}^{\infty} \lambda_{j}^{2}\right)^{2} \xrightarrow{k \rightarrow \infty} 0 .
\end{aligned}
$$

The convergence in the above display uses Proposition 3.7.1 part (a). The proof of the lemma is complete.

Proof of $(b)$. To begin, use (3.3) to note that $\left|B_{n}(i, j)\right| \leq \frac{C_{W}}{n}$, and

$$
\left|B_{n}^{2}(i, j)\right| \leq \sum_{k=1}^{n}\left|B_{n}(i, k) B_{n}(k, j)\right| \leq \frac{C_{W}^{2}}{n} .
$$

Set $r:=\frac{a}{2}+b+c$, and let $\mathcal{S}(\ell, 2 r)$ denote the set of all positive integer solutions to the equation
$\sum_{i=1}^{\ell} \alpha_{i}=2 r$. Then we have

$$
\begin{align*}
& \left|\mathbb{E}(\sqrt{n} \mathbf{Z})^{a}\left(\mathbf{Z}^{T} B_{n} \mathbf{Z}\right)^{b}\left(\mathbf{Z}^{T} B_{n}^{2} \mathbf{Z}\right)^{c}-\mathbb{E}(\sqrt{n} \mathbf{R})^{a}\left(\mathbf{R}^{T} B_{n} \mathbf{R}\right)^{b}\left(\mathbf{R}^{T} B_{n}^{2} \mathbf{R}\right)^{c}\right| \\
\leq & n^{-r} C_{W}^{b+2 c} \sum_{\ell=1}^{2 r} n^{\ell} \sum_{\alpha \in \mathcal{S}(\ell, 2 r)}\left|\mathbb{E} \prod_{i=1}^{\ell} Z_{i}^{\alpha_{i}}-\mathbb{E} \prod_{i=1}^{\ell} R_{i}^{\alpha_{i}}\right| . \tag{3.65}
\end{align*}
$$

To bound the RHS of (3.65), we consider the following cases based on $\boldsymbol{\alpha}$

- There exists $i \in[\ell]$ such that $\alpha_{i}=1$

In this case we have

$$
\mathbb{E} \prod_{i=1}^{\ell} Z_{i}^{\alpha_{i}}=\mathbb{E} \prod_{i=1}^{\ell} R_{i}^{\alpha_{i}}=0 .
$$

- $\ell>r$

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

$$
2 r=\sum_{i=1}^{\ell} \alpha_{i} \geq 2 \ell
$$

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

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

a contradiction. Since $\mathbb{E} Z_{i}^{2}=\mathbb{E} R_{i}^{2}=1$, we have

$$
\mathbb{E} \prod_{i=1}^{\ell} Z_{i}^{\alpha_{i}}=\mathbb{E} \prod_{i=1}^{\ell} R_{i}^{\alpha_{i}}=1
$$

Combining the cases above, the RHS of (3.65) is bounded by

$$
n^{-r} C_{W}^{b+2 c} \sum_{\ell=1}^{r-1} n^{\ell} \sum_{\alpha \in \mathcal{S}(\ell, 2 r)}\left|\mathbb{E} \prod_{i=1}^{\ell} Z_{i}^{\alpha_{i}}-\mathbb{E} \prod_{i=1}^{\ell} R_{i}^{\alpha_{i}}\right|=O\left(\frac{1}{n}\right),
$$

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

## References

[1] C. J. Anderson, S. Wasserman, and B. Crouch, "A p* primer: Logit models for social networks," Social networks, vol. 21, no. 1, pp. 37-66, 1999.
[2] A. Basak and S. Mukherjee, "Universality of the mean-field for the potts model," Probability Theory and Related Fields, vol. 168, no. 3, pp. 557-600, 2017.
[3] Q. Berthet, P. Rigollet, and P. Srivastava, "Exact recovery in the ising blockmodel," The Annals of Statistics, vol. 47, no. 4, pp. 1805-1834, 2019.
[4] J Besag, "Statistical analysis of non-lattice data," Journal of the Royal Statistical Society: Series D (The Statistician), vol. 24, no. 3, pp. 179-195, 1975.
[5] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," Journal of the Royal Statistical Society: Series B (Methodological), vol. 36, no. 2, pp. 192-225, 1974.
[6] S. Bhamidi, J. M. Steele, and T. Zaman, "Twitter event networks and the superstar model," Annals of Applied Probability, vol. 25, no. 5, pp. 2462-2502, 2015.
[7] B. B. Bhattacharya and S. Mukherjee, "Inference in ising models," Bernoulli, vol. 24, no. 1, pp. 493-525, 2018.
[8] J. Blitzstein and P. Diaconis, "A sequential importance sampling algorithm for generating random graphs with prescribed degrees," Internet mathematics, vol. 6, no. 4, pp. 489-522, 2011.
[9] C. Borgs, J. T. Chayes, L Lovász, V. T. Sós, and K. Vesztergombi, "Convergent sequences of dense graphs i: Subgraph frequencies, metric properties and testing," Advances in Mathematics, vol. 219, no. 6, pp. 1801-1851, 2008.
[10] C. Borgs, J. T. Chayes, L Lovász, V. T. Sós, and K. Vesztergombi, "Convergent sequences of dense graphs ii. multiway cuts and statistical physics," Annals of Mathematics, pp. 151-219, 2012.
[11] M. Burnašev, "Minimax detection of an imperfectly known signal against a background of gaussian white noise," Teor. Veroyatn. Primen, vol. 24, pp. 106-118, 1979.
[12] S. Chatterjee, "Stein's method for concentration inequalities," arXiv preprint math/0604352, 2006.
[13] S. Chatterjee and S. Mukherjee, "Estimation in tournaments and graphs under monotonicity constraints," IEEE Transactions on Information Theory, vol. 65, no. 6, pp. 3525-3539, 2019.
[14] S. Chatterjee, "Estimation in spin glasses: A first step," The Annals of Statistics, vol. 35, no. 5, pp. 1931-1946, 2007.
[15] S. Chatterjee and P. Diaconis, "Estimating and understanding exponential random graph models," The Annals of Statistics, vol. 41, no. 5, pp. 2428-2461, 2013.
[16] S. Chatterjee, P. Diaconis, and A. Sly, "Random graphs with a given degree sequence," Annals of Applied Probability, vol. 21, no. 4, pp. 1400-1435, 2011.
[17] F. Comets and B. Gidas, "Asymptotics of maximum likelihood estimators for the curie-weiss model," The Annals of Statistics, pp. 557-578, 1991.
[18] N. Deb, R. Mukherjee, S. Mukherjee, and M. Yuan, "Detecting structured signals in ising models," arXiv preprint arXiv:2012.05784, 2020.
[19] N. Deb and S. Mukherjee, "Fluctuations in mean-field ising models," arXiv preprint arXiv:2005.00710, 2020.
[20] A. Dembo and A. Montanari, "Ising models on locally tree-like graphs," The Annals of Applied Probability, vol. 20, no. 2, pp. 565-592, 2010.
[21] A. Dembo and A. Montanari, "Gibbs measures and phase transitions on sparse random graphs," Brazilian Journal of Probability and Statistics, vol. 24, no. 2, pp. 137-211, 2010.
[22] R. Durrett, Probability: theory and examples. Cambridge university press, 2019, vol. 49.
[23] R. S. Ellis and C. M. Newman, "The statistics of curie-weiss models," Journal of Statistical Physics, vol. 19, no. 2, pp. 149-161, 1978.
[24] P. Erdős, A. Rényi, et al., "On the evolution of random graphs," Publ. Math. Inst. Hung. Acad. Sci, vol. 5, no. 1, pp. 17-60, 1960.
[25] O. Frank and D. Strauss, "Markov graphs," Journal of the American Statistical Association, vol. 81, no. 395, pp. 832-842, 1986.
[26] P. Ghosal and S. Mukherjee, "Joint estimation of parameters in ising model," The Annals of Statistics, vol. 48, no. 2, pp. 785-810, 2020.
[27] F. Götze, H. Sambale, and A. Sinulis, "Concentration inequalities for polynomials in $\alpha$-subexponential random variables," Electronic Journal of Probability, vol. 26, pp. 1-22, 2021.
[28] A. B. Harris, "Effect of random defects on the critical behaviour of ising models," Journal of Physics C: Solid State Physics, vol. 7, no. 9, p. 1671, 1974.
[29] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," Social networks, vol. 5, no. 2, pp. 109-137, 1983.
[30] P. W. Holland and S. Leinhardt, "An exponential family of probability distributions for directed graphs," Journal of the American Statistical Association, vol. 76, no. 373, pp. 33-50, 1981.
[31] Y. I. Ingster, "Minimax detection of a signal for $l_{n}$-balls," Mathematical Methods of Statistics, vol. 7, no. 4, pp. 401-428, 1998.
[32] Y. I. Ingster, "Minimax detection of a signal in $l_{p}$ metrics," Journal of Mathematical Sciences, vol. 68, no. 4, pp. 503-515, 1994.
[33] Y. Ingster and I. A. Suslina, Nonparametric goodness-of-fit testing under Gaussian models. Springer Science \& Business Media, 2012, vol. 169.
[34] E. Ising, "Beitrag zur theorie des ferromagnetismus," Zeitschrift für Physik, vol. 31, no. 1, pp. 253-258, 1925.
[35] Z. Kabluchko, M. Löwe, and K. Schubert, "Fluctuations of the magnetization for ising models on dense erdős-rényi random graphs,"Journal of Statistical Physics, vol. 177, no. 1, pp. 78-94, 2019.
[36] L. Le Cam et al., "Limits of experiments," in Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press Berkeley-Los Angeles, vol. 1, 1972, pp. 245-261.
[37] J. L. Lebowitz, "Ghs and other inequalities," Communications in Mathematical Physics, vol. 35, no. 2, pp. 87-92, 1974.
[38] L. Lovász, Large networks and graph limits. American Mathematical Soc., 2012, vol. 60.
[39] R. Mukherjee, S. Mukherjee, and S. Sen, "Detection thresholds for the $\beta$-model on sparse graphs," The Annals of Statistics, vol. 46, no. 3, pp. 1288-1317, 2018.
[40] R. Mukherjee, S. Mukherjee, and M. Yuan, "Global testing against sparse alternatives under ising models," The Annals of Statistics, vol. 46, no. 5, pp. 2062-2093, 2018.
[41] S. Mukherjee and Y. Xu, "Statistics of the two-star ergm," To appear, Bernoulli, 2021.
[42] L. Onsager, "Crystal statistics. i. a two-dimensional model with an order-disorder transition," Physical Review, vol. 65, no. 3-4, p. 117, 1944.
[43] J. Park and M. E. Newman, "Solution of the two-star model of a network," Physical Review $E$, vol. 70, no. 6, p. 066 146, 2004.
[44] A. Rinaldo, S. Petrović, and S. E. Fienberg, "Maximum lilkelihood estimation in the betamodel," The Annals of Statistics, vol. 41, no. 3, pp. 1085-1110, 2013.
[45] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, "An introduction to exponential random graph ( $p^{*}$ ) models for social networks," Social networks, vol. 29, no. 2, pp. 173-191, 2007.
[46] M. Schweinberger and J. Stewart, "Concentration and consistency results for canonical and curved exponential-family models of random graphs," The Annals of Statistics, vol. 48, no. 1, pp. 374-396, 2020.
[47] C. R. Shalizi and A. Rinaldo, "Consistency under sampling of exponential random graph models," Annals of statistics, vol. 41, no. 2, pp. 508-535, 2013.
[48] A. W. Van der Vaart, Asymptotic statistics. Cambridge university press, 2000, vol. 3.
[49] S. Wasserman and K. Faust, Social network analysis: Methods and applications. Cambridge university press, 1994, vol. 8.
[50] S. Wasserman and P. Pattison, "Logit models and logistic regressions for social networks: I. an introduction to markov graphs andp," Psychometrika, vol. 61, no. 3, pp. 401-425, 1996.
[51] M. Yuan, R. Liu, Y. Feng, and Z. Shang, "Testing community structures for hypergraphs," Tech. Rep., 2018.

