CONSISTENT ESTIMATION IN THE TWO STAR EXPONENTIAL RANDOM GRAPH MODEL

SUMIT MUKHERJEE†

ABSTRACT. This paper explores statistical properties of a particular Exponential Random Graph Model, the two star probability distribution on the space of simple graphs. Non degenerate limiting distributions for the number of edges is derived for all parameter domains, and is shown to have similar phase transition properties as the magnetization in the Curie-Weiss model of statistical physics. As a consequence estimates for both parameters are derived, which are consistent irrespective of the phase transition.

1. INTRODUCTION

Exponential Random Graph Models (ERGM) are a class of random graph models which have been studied extensively in social science literature. For a list of references see [AWC], [FS], [H], [HL], [MHH], [Newman], [PW], [RPKL], [Snijders], [Strauss], [SPRH], [WF], and the references therein.

The two star model is possibly the simplest example of a non Erdos Renyi ERGM. This model has been studied in [PN] in 2004. The model is formally defined below:

1.1. Definition of Two star model. For \( n \in \mathbb{N} \) be a positive integer, let \( \mathcal{X}_n \) denote the space of all simple graphs with vertices labeled \( [n] := \{1, 2, \cdots, n\} \). Since a simple graph is uniquely identified by its adjacency matrix, a graph can be identified with its adjacency matrix. Thus w.l.o.g. take \( \mathcal{X}_n \) to be the set of all symmetric \( n \times n \) matrices, with 0 on the diagonal elements and \( \{0/1\} \) on the off-diagonal elements. Set \( x_{ij} = 1 \) if an edge is present between vertices \( i \) and \( j \), and 0 otherwise. The two star model on \( \mathcal{X}_n \) has the probability mass function

\[
\frac{1}{Z_n(\beta)} \exp\left\{ \frac{\beta_2}{n-1} T(x) + \left( \beta_1 + \frac{\beta_2}{n-1} \right) E(x) \right\},
\]

where \( Z_n(\beta) \) is the unknown normalizing constant. The parameter space considered in this paper is \( \beta = (\beta_1, \beta_2) \in \mathbb{R} \times (0, \infty) \). Here \( E(x) \) is the number of edges in the graph \( x \), given by \( E(x) = \sum_{i<j} x_{ij} \), and \( T(x) \) is the number of two stars in \( x \), given by \( \sum_{i<j} \sum_{j<l,j,l \neq i} x_{ij}x_{ik} \).

The main difficulty in developing estimators for the parameter \( \beta \) in this model is that the normalizing constant \( Z_n(\beta) \) is not available in closed form. Explicit computation of the partition function takes time which is exponential time \( n \), and so the calculation of MLE becomes infeasible. Theoretical properties of other estimators such as MCMC (see [GT]) or pseudo-likelihood estimator ([B1],[B2]) are not well understood for ERGM’s in general. In 2008, [BSB] studied the mixing times of Glauber dynamics for ERGM models, showing that there are some regimes of
parameters where the mixing time is polynomial in \(n\), and other regimes where the mixing time is exponential in \(n\). In 2011 the limiting normalizing constant for ERGM models was computed in [CD, Theorem 4.1]. In terms of the two star model, the theorem states

\[
\lim_{n \to \infty} \frac{1}{n^2} \log Z_n(\beta) = \frac{1}{2} \sup_{0 < p < 1} \{ \beta_1 p + \beta_2 p^2 - p \log p + (1 - p) \log(1 - p) \}.
\]

[CD] also proves that if the optimization above is attained at a unique \(p\), the model generates data which are very close in the sense of cut-metric to those generated by an Erdos Renyi. On the other hand if the optimizing \(p\) is not unique, the graph can look like a mixture of Erdos Renyi. For a discussion on the cut metric, see [CD] and the references therein. Since an Erdos Renyi graph is characterized by one and one parameter only, this seems to suggest that the two model might be un-identifiable in the limit.

Building on ideas of [PN], in [M] the author gave an exact characterization of the different parameter domains, and explored some properties of the degrees in these domains. In particular [M] shows the following two results:

**Theorem.** [M, Theorem 4.1] If either \(\beta_1 + \beta_2 \neq 0\) or \(\beta_2 < 2\) then there exists a \(p_0\) in \((0, 1)\) such that

\[
\mathbb{P}_{n, \beta} \left( \max_{1 \leq i \leq n} \left| \frac{d_i}{n - 1} - p_0 \right| > \delta \right) \leq e^{-C(\delta)n},
\]

where \(C(\delta) > 0\).

**Theorem.** [M, Theorem 4.2] If \(\beta_1 + \beta_2 = 0\) and \(\beta_2 > 2\) then there exists \(p_1 > p_2\) in \((0, 1)\) such that

\[
\mathbb{P}_{n, \beta} \left( \max_{1 \leq i \leq n} \left| \frac{d_i}{n - 1} - p_j \right| > \delta \right) - \frac{1}{2} \leq e^{-C(\delta)n},
\]

where \(C(\delta) > 0\), and \(j = 1, 2\).

The theorem also gives predictions for \(p_j\) for \(j = 0, 1, 2\). This further illustrates the Erdos-Renyi (Theorem 4.1) and mixture of Erdos Renyi behavior (Theorem 4.2), and seems to corroborate the non identifiability of this model in the limit.

### 1.2. Main results of this paper.

The first result of this paper is Theorem 2.1, which characterizes the limiting distribution of the number of edges \(E(x)\) for different domains of parameters. The limiting behavior of the number of edges has similar phase transition properties as that of the Curie Weiss Ising model.

The second result is Theorem 2.2, which gives the weak limit of the sampling variance of the degrees. The limiting constant is not smooth in the parameters at the critical point, as observed in [PN] (see also remark 2.1).

Using the above two theorems, it is shown in Theorem 2.3 that consistent estimation of both the parameters is indeed possible in the two star model. As a consequence, it follows that even though the two star model looks like an Erdos Renyi mixture in the cut metric, the same convergence does not go through in total variation. However for a practical perspective see remark 2.2.

All the theorems are formally stated in section 2, which also recalls the partition of the parameter space done in [M, section 3]. Section 2 also gives the proof of Theorem 2.3. Section 3 recalls the construction of auxiliary variables, which was done in [M, section 2]. Theorems 2.1
and 2.2 are proved in section 3 using a series of lemmas. Section 4 is dedicated to proving these
lemmas. Section 5 confirms Theorem 2.1 via simulations using the auxiliary variable method of [M].

2. Definitions and statement of results

As in [M], the parameters $\beta := (\beta_1, \beta_2)$ are re-parametrized as

$$\theta_1 := \frac{\beta_1 + \beta_2}{2}, \theta_2 := \frac{\beta_2}{4}.$$ 

A new parameter $m = m(\theta)$, which will be required for stating the results, is defined below. The
definition of $m$ depends on the roots of the equation $t = \tanh(2\theta_2 t + \theta_1)$, and is different for every
domain.

- Uniqueness region

$$\Theta_1 := \left\{ \theta_1 = 0, 0 < \theta_2 < \frac{1}{2} \right\} \cup \left\{ \theta_1 > 0, \theta_2 > 0 \right\} \cup \left\{ \theta_1 < 0, \theta_2 > 0 \right\} := \Theta_{11} \cup \Theta_{12} \cup \Theta_{12}$$

For $\theta \in \Theta_{11}$ the function $t = \tanh(2\theta_2 t + \theta_1)$ has the unique root 0. In this domain set

$m = 0$.

For $\theta \in \Theta_{12}$ the function $t = \tanh(2\theta_2 t + \theta_1)$ has a unique positive root, which is defined
to be $m$.

For $\theta \in \Theta_{13}$ the function $t = \tanh(2\theta_2 t + \theta_1)$ has a unique negative root, which is defined
to be $m$.

- Non uniqueness region

$$\Theta_2 : \left\{ \theta_1 = 0, \theta_2 > \frac{1}{2} \right\}$$

In this domain the function $t = \tanh(2\theta_2 t + \theta_1)$ has exactly two non zero roots of equal magni-
dtude but opposite sign. In this domain set $m$ to be the unique positive root of the equation.

- Critical point

$$\Theta_3 := \left\{ \theta_1 = 0, \theta_2 = \frac{1}{2} \right\}$$

In this domain the function $t = \tanh(2\theta_2 t + \theta_1)$ has the unique root at 0, and so set $m = 0$.

The assertions about the roots of the equation $t = \tanh(2\theta_2 t + \theta_1)$ can be checked directly, or can
be checked from ([DM, Page 9]). It follows from their analysis that $m$ is continuous but not smooth
as a function of $(\theta_1, \theta_2)$ . In particular along the line $\theta_1 = 0$, $m = 0$ for $\theta_2 \leq \frac{1}{2}$, and $m > 0$ for
$\theta_2 > \frac{1}{2}$. Note in passing that $(m, \theta_2)$ is a re-parametrization of $(\theta_1, \theta_2)$. 
Definition 2.1. Let \((d_1(x), \cdots, d_n(x))\) denote the degree statistics of the graph \(x\). Also let

\[
\mu := \frac{2\theta_2 m (1 - m^2)}{[1 - \theta_2 (1 - m^2)][1 - 2\theta_2 (1 - m^2)]},
\]

\[
\tau_1 := \frac{2(1 - m^2)}{1 - 2\theta_2 (1 - m^2)},
\]

\[
\tau_2 := \frac{(1 - m^2)}{1 - \theta_2 (1 - m^2)}.
\]

The first theorem gives non-degenerate limit distributions for the number of edges.

Theorem 2.1. Let

\[
S_1 := \frac{1}{n-1} \left[ 2\bar{d} - (n-1) \right] = \frac{4}{n(n-1)} \left[ E(x) - \frac{1}{4} n(n-1) \right].
\]

Then

- For \(\theta \in \Theta_1\),
  \[
  (n-1)(S_1 - m) \xrightarrow{d} N(-\mu, \tau_1).
  \]

- For \(\theta \in \Theta_2\),
  \[
  (n-1)(S_1 - m) \xrightarrow{d} \frac{1}{2} \left[ N(\mu, \tau_1) + N(-\mu, \tau_1) \right].
  \]

The second theorem gives the weak limit distributions for the sample variance of the degree distribution.

Theorem 2.2. With

\[
S_2 := \frac{4}{(n-1)^2} \sum_{i=1}^{n} (d_i - \bar{d})^2,
\]

for all \(\theta \in \Theta_1 \cup \Theta_2\),

\[
S_2 \xrightarrow{p} \tau_2.
\]

Remark 2.1. Theorem 2.2 shows that the in probability limit of the sampling variance of the degrees exhibits a phase transition in \(\theta_2\) along the line \(\theta_1 = 0\), as was observed in [PN][Figure 2]. Indeed, note that \(m\) is not smooth near \(\theta_2 = \frac{1}{2}\), and consequently neither is \(\tau_2\).

With the help of Theorem 2.1 and Theorem 2.2, explicit estimators of \(\theta = (\theta_1, \theta_2)\) are constructed, which are shown to be consistent on \(\Theta_1 \cup \Theta_2\) in the next theorem.

Theorem 2.3. Define the estimators \((S_3, S_4)\) as follows:

\[
S_3 := \frac{S_1^2 + S_2 - 1}{S_2 - S_1 S_2}, \quad S_4 := \tanh^{-1}(S_1) - 2S_3 S_1.
\]

If \(\theta \in \Theta_1 \cup \Theta_2\),

\[
(S_3, S_4) \xrightarrow{p} (\theta_2, \theta_1).
\]

Proof. Theorem (2.1) gives that \(|S_1| \xrightarrow{d} |m|\). Along with Theorem (2.2) this gives

\[
S_3 \xrightarrow{p} \frac{\tau_2 + m^2 - 1}{\tau_2 (1 - m^2)}.
\]
which equals $\theta_2$ after a simple calculation.

Proceeding to show that $S_4$ is consistent for $\theta_1$, the argument splits into two case:

- $\theta \in \Theta_1$
  In this domain $S_1 \xrightarrow{p} m$, and so
  $$S_4 \xrightarrow{p} \tanh^{-1}(m) - 2\theta_2 m = \theta_1,$$
  since $m = \tanh(2m\theta_2 + \theta_1)$.

- $\theta \in \Theta_2$
  In this domain $S_1 \xrightarrow{d} \frac{1}{2} \delta_m + \frac{1}{2} \delta_{-m}$. Conditioning on the set $S_1 \xrightarrow{p} \pm m$, the same argument as before shows that $S_4 \xrightarrow{p} \theta_1$ thus completing the proof.

**Remark 2.2.** Theorem 2.1 says that the error in estimating $|m|$ by $|S_1|$ is $O(1/n)$. A more detailed analysis than Theorem 2.2 shows that the error in estimating $S_2$ by $\tau_2$ is $O(1/\sqrt{n})$. Since $\theta_1, \theta_2$ are functions of both $m$, the resulting errors $|S_3 - \theta_2|, |S_4 - \theta_1|$ are also $O(1/\sqrt{n})$. Thus even though the mean parameter $m$ can be approximated with good precision, estimation of the natural parameters of the exponential family is harder.

Another observation is that the function $m$ converges to $\pm 1$ exponentially fast in either of the parameters $\theta_1/\theta_2$, when the other parameter is kept fixed, and so for any reasonably large values of the parameters $m$ is close to $\pm 1$. Since $m \approx 1$ corresponds to near complete graphs and $m \approx -1$ corresponds to near empty graphs, for almost all reasonably large positive and negative values of the parameters the graph is either full or empty. In fact, if $\theta_1 = 0$ and $\theta_2$ is large, then the graph is a mixture of Erdos Renyi $(1 + m)/2 \approx 1$ and $(1 - m)/2 \approx 0$ with probability $1/2$, and so repeated sampling from this model will produce entirely different graphs. This also makes the problem of statistical estimation in this model difficult, as the model is somewhat unstable. This phenomenon of ERGM, commonly known as degeneracy, has already been observed in social science literature (see [H], [SPRH], [MHH]).

3. **Auxiliary variables.** This subsection recalls the construction of auxiliary variables as done in [M, section 2.2].

Transform the problem from $\{0, 1\}$ to $\{-1, 1\}$ by defining an $n \times n$ matrix $y$ as follows:

$$y_{ij} = y_{ji} := 2x_{ij} - 1 \in \{-1, 1\}, 1 \leq i < j \leq n; \quad y_{ii} = 0.$$

With $k_i := \sum_{j \neq i} y_{ij}$, let $(\phi_1, \cdots, \phi_n)$ have the following law:

Given $y$, $(\phi_i)$’s are mutually independent, with

$$\phi_i \sim \mathcal{N}\left(\frac{k_i}{n-1}, \frac{1}{(n-1)\theta_2}\right).$$

The above construction is equivalent to the following representation:

$$\phi_i := \frac{k_i}{n-1} + \frac{Z_i}{\sqrt{(n-1)\theta_2}}, Z_i \overset{i.i.d.}{\sim} \mathcal{N}(0, 1) \text{ independent of } \{k_i\}_{i=1}^n. \quad (3.1)$$
The marginal density of $\phi$ is given by the un-normalized density $f_n(.)$, where
\begin{equation}
\begin{aligned}
f_n(\phi) := e^{\sum_{i<j} p(\phi_i, \phi_j)}, \quad p(x, y) := & \frac{\theta_2}{2}(x^2 + y^2) - \log(\cosh(\theta_2(x + y) + \theta_1)).
\end{aligned}
\tag{3.2}
\end{equation}

3.2. Contiguous approximation lemma. The following lemma gives a general approximation scheme for un-normalized densities of the form $f_n(\phi)$. The proof of the Lemma is moved to the appendix.

**Definition 3.1.** By an un-normalized density (u.n.d) function in a $\sigma$ finite measure space $(\Omega, F, \mu)$, is meant a strictly positive measurable function $h$ which is integrable w.r.t. $\mu$, i.e. $\int_{\Omega} h d\mu < \infty$.

The probability induced by $h$ induced on $(\Omega, F)$ is defined to be $\mathbb{H}$, where $\mathbb{H}$ is given by
\begin{equation}
\begin{aligned}
\mathbb{H}(A) = \frac{\int_A h d\mu}{\int_{\Omega} h d\mu}.
\end{aligned}
\end{equation}

**Lemma 3.1.** Let $(\Omega_n, F_n, \mu_n)$ be a sequence of $\sigma$-finite measure spaces, and let $h_n(\cdot), g_n(\cdot)$ be two u.n.d. on $\Omega_n$. Define $L_n(\cdot) = \log \frac{h_n(\cdot)}{g_n(\cdot)}$, and let $\mathbb{H}_n, \mathbb{G}_n$ denote the probability measures induced by $h_n, g_n$ respectively. To be precise, if
\begin{equation}
\begin{aligned}
\int_{\Omega_n} h_n d\mu_n =: a_n, \int_{\Omega_n} g_n d\mu_n =: b_n,
\end{aligned}
\end{equation}
then for $A_n \in F_n$,
\begin{equation}
\begin{aligned}
\mathbb{H}_n(A_n) = \frac{\int_{A_n} h_n d\mu_n}{a_n}, \mathbb{G}_n(A_n) = \frac{\int_{A_n} g_n d\mu_n}{b_n}.
\end{aligned}
\end{equation}

If $L_n$ is $O_p(1)$ under both measures $\mathbb{H}_n, \mathbb{G}_n$, then the measures $\mathbb{H}_n, \mathbb{G}_n$ are mutually contiguous. In this case,
\begin{enumerate}[(a)]
\item if $$(X_n, L_n) \overset{d}{\rightarrow} N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_{12})$$
then $\mu_2^2 + \frac{1}{2}\sigma_2^2 = 0$, and
\begin{equation}
\begin{aligned}
X_n \overset{d}{\rightarrow} N(\mu_1 + \sigma_{12}, \sigma_1^2)
\end{aligned}
\end{equation}
\item if $L_n \overset{d}{\rightarrow} c$ where $c > 0$ is a constant, then $\frac{h_n(.)}{g_n(.)} \overset{d}{\rightarrow} 1$ under both $\mathbb{H}_n, \mathbb{G}_n$. Also in this case
\begin{equation}
\begin{aligned}
\|\mathbb{H}_n - \mathbb{G}_n\|_{TV} \rightarrow 0.
\end{aligned}
\end{equation}
\end{enumerate}

Using Lemma 3.1, the following Lemma is derived in section 4 by a detailed analysis of $f_n(.)$ in the two domains $\Theta_1$ and $\Theta_2$ separately.

**Lemma 3.2.** Set
\begin{equation}
\begin{aligned}
\eta_1 := \frac{1}{\theta_2[1 - 2\theta_2(1 - m^2)]}, \quad \eta_2 := \frac{1}{\theta_2[1 - \theta_2(1 - m^2)]}.
\end{aligned}
\end{equation}

If $\theta \in \Theta_1$ then
\begin{equation}
\begin{aligned}
n(\bar{\phi} - m) \overset{d}{\rightarrow} N(-\mu, \eta_1).
\end{aligned}
\tag{3.3}
\end{equation}

If $\theta \in \Theta_2$ then
\begin{equation}
\begin{aligned}
[n(\bar{\phi} - m)|\bar{\phi} > 0] \overset{d}{\rightarrow} N(-\mu, \eta_1), \tag{3.4}
\end{aligned}
\end{equation}
\begin{equation}
\begin{aligned}
[n(\bar{\phi} + m)|\bar{\phi} < 0] \overset{d}{\rightarrow} N(\mu, \eta_1). \tag{3.5}
\end{aligned}
\end{equation}

Further, the following results hold for both the domains:
\[ E_{\mathcal{P}_{n,\beta}} \sum_{i=1}^{n} (\phi_i - \bar{\phi})^2 = O(1), \] (3.6)

\[ \sum_{i=1}^{n} (\phi_i - \bar{\phi})^2 \xrightarrow{p} \eta_2. \] (3.7)

### 3.3. Proofs of Theorem 2.1 and Theorem 2.2.

**Proposition 3.1.** Let \((\Omega_n, \mathcal{F}_n, \mathbb{P}_n)\) be a sequence of probability spaces.

If there are random variables \((A_n, B_n)\) such that the following three conditions hold:

(i) \(A_n + B_n \xrightarrow{d} N(0, \sigma_1^2 + \sigma_2^2)\).

(ii) \(A_n\) and \(B_n\) are independent.

(iii) \(A_n \xrightarrow{d} N(0, \sigma_1^2)\).

Then

\[ B_n \xrightarrow{d} N(0, \sigma_2^2). \]

**Proof.** A direct calculation using characteristic functions gives

\[ \phi_{A_n+B_n}(t) = \phi_{A_n}(t)\phi_{B_n}(t) \]

with

\[ \phi_{A_n}(t) \to e^{-t^2\sigma_1^2/2}, \quad \phi_{A_n+B_n}(t) \to e^{-t^2(\sigma_1^2+\sigma_2^2)/2} \]

which readily implies

\[ \phi_{B_n}(t) \to e^{-t^2\sigma_2^2/2}, \]

completing the proof of the proposition.

\[ \square \]

**Proof of Theorem 2.1.** Using (3.1) gives

\[ (n-1)(\bar{\phi} - m) + \mu = (\bar{k} - (n-1)m + \mu) + \frac{\sqrt{n-1}}{\sqrt{\theta_2}} \bar{Z} \]

where \(\{Z_i\}_{i=1}^{n}\) are i.i.d. \(N(0,1)\). To complete the proof of (2.1) requires an application of Lemma 3.1 with

\[ A_n = \frac{\sqrt{n-1}}{\sqrt{\theta_2}} \bar{Z}, \quad B_n = \bar{k} - (n-1)m + \mu. \]

Condition (i) follows by (3.3), whereas (ii) and (iii) follow on noting that \(\eta_1 = \tau_1 + 1/\theta_2\). Thus Lemma 3.1 gives

\[ \bar{k} - (n-1)m \xrightarrow{d} N(-\mu, \tau_2), \]

which is a restatement of (2.1).

For (2.2), note that by [M, Lemma 4.3],

\[ \lim_{n \to \infty} \mathbb{P}_{n,\beta}(\phi_i > 0, 1 \leq i \leq n) = \frac{1}{2}. \]

Conditioned on the set \(\phi \in (0, \infty)^n\) a similar argument using (3.4) gives

\[ ((n-1)(S_1 - m)|\phi \in (0, \infty)^n) \xrightarrow{d} N(-\mu, \tau_1). \]
A similar argument using (3.5) gives

\[ ((n-1)(S_1-m)\phi \in (-\infty, 0)^n) \overset{d}{\rightarrow} N(\mu, \tau_1), \]

proving (2.2) and concluding the proof of Theorem 2.1.

\[ \square \]

**Proof of Theorem 2.2.** Using (3.1) gives

\[
\sum_{i=1}^{n} (\phi_i - \bar{\phi})^2 = A_n + B_n + C_n,
\]

where

\[
A_n := \frac{1}{(n-1)\theta_2} \sum_{i=1}^{n} (Z_i - \bar{Z})^2,
\]

\[
B_n := 2 \frac{1}{(n-1)^{3/2} \sqrt{\theta_2}} \sum_{i=1}^{n} (k_i - \bar{k})Z_i,
\]

\[
C_n := \frac{1}{(n-1)^2} \sum_{i=1}^{n} (k_i - \bar{k})^2.
\]

Note that \( \mathbb{E}_{\mathbb{P}_{\eta,\beta}} (B_n) = \mathbb{E}_{\mathbb{P}_{\eta,\beta}} \mathbb{E}_{\mathbb{P}_{\eta,\beta}} (B_n|y) = 0 \), and

\[
\text{Var}(B_n) = \mathbb{E}_{\mathbb{P}_{\eta,\beta}} \text{Var}(B_n|y) = \frac{4}{(n-1)^3 \theta_2} \mathbb{E}_{\mathbb{P}_{\eta,\beta}} \sum_{i=1}^{n} (k_i - \bar{k})^2 = O(1/n),
\]

where the last bound is a consequence of (3.6) and representation (3.1). It follows that \( B_n \overset{p}{\rightarrow} 0 \).

Also,

\[ A_n + B_n + C_n \overset{p}{\rightarrow} \eta_2, \quad A_n \overset{p}{\rightarrow} \frac{1}{\theta_2}, \]

where the first conclusion uses (3.7). Since \( \eta_2 = \tau_2 + 1/\theta_2 \), it follows that \( C_n \overset{p}{\rightarrow} \tau_2 \) which is a re-statement of Theorem 2.2.

\[ \square \]

4. The Two Domains

This section carries out domain specific analysis using the density \( f_n(\phi) \) of (3.2) to deduce Lemma 3.2.

The first Lemma is not domain specific, and in fact works even if \((\theta_1, \theta_2) = (0, 1/2)\), the critical point configuration.

**Lemma 4.1.** For \( \theta \in \Theta_1 \cup \Theta_2 \cup \Theta_3 \),

\[
\frac{1}{n} \sum_{e \in \mathcal{E}} [y_e - \tanh(\frac{\theta}{n-1}t_e(y) + \theta_1)] = O_p(1)
\]

where \( t_e(y) := \sum_{f \in N(e)} y_e \).

**Proof.** The proof of this claim is similar to the proof of [C, Lemma 1.2] using exchangeable pairs, and is not repeated here.

\[ \square \]
The next lemma gives a moment estimate. Recall from [M, (3.1)] the definition of $q(.)$.

**Definition 4.1.** Define $q : \mathbb{R} \mapsto \mathbb{R}$ by

$$q(t) := \frac{\theta_2}{4} t^2 - \log \cosh(\theta_2 t + \theta_1).$$

The relation between $p(.,.)$ and $q(.)$ is given by $p(x,y) = q(x+y) + \frac{\theta_1}{4}(x-y)^2$. 

**Lemma 4.2.** For $U \in \{(0,\infty), (-\infty, 0), \mathbb{R}\}$, suppose there exists $\phi_0 \in U$ such that $q(t)$ has a global minima on $U$ at $t = 2\phi_0$ with $q''(2\phi_0) > 0$. Then for any $l \in \mathbb{N}$, there exists $C_l < \infty$ such that

$$\mathbb{E}_{n,\beta}(|\phi_i - \phi_0|^l | \phi \in U^n) \leq \frac{C_l}{n^{l/2}}.$$

**Proof.** The proof of this lemma uses similar calculations to the proof of [M, Lemma 4.2] and is not repeated here. □

### 4.1. Uniqueness domain: $\theta \in \Theta_1$

From [M, Section 3] it follows that the conditions of Lemma 4.2 are satisfied with $\phi_0 = m$ and $U = \mathbb{R}$.

The next lemma is the first step for invoking Lemma 3.1. As a comment, to show that a sequence of random variables $X_n$ is $O_p(1)$, it is enough to show

$$\limsup_{n \to \infty} \mathbb{E}|X_n|^l < \infty$$

for some $l > 0$. For then Markov’s inequality gives

$$\limsup_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}(|Z_n| > K) \leq \limsup_{K \to \infty} \frac{1}{K^l} \limsup_{n \to \infty} \mathbb{E}|X_n|^l = 0,$$

proving tightness.

**Lemma 4.3.** (a) $\mathbb{E}_{p,n,\beta} \sum_{i=1}^{n} (\phi_i - m)^2 = O(1)$.

(b) $\sum_{i=1}^{n} (\phi_i - m) = O_p(1)$.

**Proof.** (a) An application of Lemma 4.2 with $l = 2$ gives the desired conclusion.

(b) A Taylor expansion of tanh gives

$$\frac{1}{n} \left| \sum_{e \in \mathcal{E}} (y_e - \tanh(\frac{\theta_2 t_e(y)}{n-1} + \theta_1) - \sum_{e \in \mathcal{E}} (y_e - m) + \sum_{e \in \mathcal{E}} 2\theta_2 (1-m^2) (\frac{t_e(y)}{2(n-1)} - m) \right| \leq \frac{C}{n} \sum_{e \in \mathcal{E}} \left( \frac{t_e(y)}{2(n-1)} - m \right)^2,$$

for some constant $C < \infty$. Using (3.1) it follows that the r.h.s. of (4.1) is bounded by

$$C' \left[ \sum_{i=1}^{n} (\phi_i - m)^2 + \frac{1}{(n-1)\theta_2} \sum_{i=1}^{n} Z_i^2 \right]$$

for some $C' < \infty$. By part (a) $\sum_{i=1}^{n} (\phi_i - m)^2 = O_p(1)$, and so the l.h.s. of (4.1) is $O_p(1)$.

Writing the l.h.s. of (4.1) as

$$\left| \frac{1}{n} \sum_{e \in \mathcal{E}} (y_e - \tanh(\frac{\theta_2 t_e(y)}{n-1} + \theta_1) - \frac{1-2\theta_2 (1-m^2)}{n} \sum_{e \in \mathcal{E}} (y_e - m) - \frac{2\theta_2 (1-m^2)}{n(n-1)} \sum_{e \in \mathcal{E}} y_e \right|,$$
it follows by Lemma 4.1 that \( \frac{1}{n} \sum_{e \in E} (y_e - m) \) is \( O_p(1) \). (This needs the fact that \( 2\theta_2(1 - m^2) < 1 \), but this is guaranteed by \( q''(2m) > 0 \).) By again using (3.1) the desired conclusion follows.

The next Lemma gives a mutually contiguous approximation for \( P_n \), which is the marginal distribution of \( \phi \) under \( P_{n,\beta} \). Since the same idea will be used in Domain 2, the calculations are done in detail.

**Lemma 4.4.** Let \( g_n(\phi) \) be defined by

\[
- \log g_n(\phi) := \frac{n(n - 1)}{2} p(m, m) + \frac{a_1 n}{2} \sum_{i=1}^{n} (\phi_i - m)^2 - \frac{a_2 n^2}{2} (\bar{\phi} - m)^2,
\]

where \( a_1 = \theta_2 - \theta_2^2(1 - m^2) \) and \( a_2 = \theta_2^2(1 - m^2) \). Then \( g_n \) is an u.n.d. (see Definition 3.1), and the two corresponding laws \( P_n \) and \( G_n \) are mutually absolutely contiguous.

**Proof.** Expanding \( p(x, y) \) by Taylor series around \((m, m)\) gives

\[
p(x, y) = p(m, m) + \frac{1}{2} [a_1 (x - m)^2 + a_1 (y - m)^2 - 2a_2 (x - m)(y - m)] + \frac{a_3}{3!} (x + y - 2m)^3
\]

\[
+ \frac{a_4}{4!} (x + y - 2m)^4 + R(x, y), \quad |R(x, y)| \leq C_5 (|x - m|^5 + |y - m|^5)
\]

where \( C_5 < \infty \), and the constants are given by

\[a_1 = \theta_2 - \theta_2^2(1 - m^2), \quad a_2 = \theta_2^2(1 - m^2), \quad a_3 := 2\theta_2^3 m(1 - m^2), \quad a_4 := 2\theta_2^4 (1 - m^2)(1 - 3m^2)\].

Since \( q''(2m) > 0 \) it follows that \( a_1 > a_2 > 0 \), and consequently \( g_n \) is integrable, and so an u.n.d. Using the above expansion, the first four terms in the expansion of \(- \log f_n(\phi)\) consists of the following terms:

\[
R_{1, f_n} := \sum_{i < j} \frac{a_1}{2} [(\phi_i - m)^2 + (\phi_j - m)^2] = \frac{a_1(n - 1)}{2} \sum_{i=1}^{n} (\phi_i - m)^2
\]

\[
R_{2, f_n} := -a_2 \sum_{i < j} (\phi_i - m)(\phi_j - m) = -\frac{a_2 n^2}{2} (\bar{\phi} - m)^2 + \frac{a_2 n}{2} \sum_{i=1}^{n} (\phi_i - \bar{\phi})^2 + \frac{na_2}{2} (\bar{\phi} - m)^2
\]

\[
R_{3, f_n} := \frac{a_3}{2} \sum_{i < j} (\phi_i + \phi_j - 2m)^3 = \frac{a_3}{3!} \left[ \sum_{i,j=1}^{n} (\phi_i + \phi_j - 2m)^3 - 8 \sum_{i=1}^{n} (\phi_i - m)^3 \right]
\]

\[
= \frac{a_3}{3!} (n - 4) \sum_{i=1}^{n} (\phi_i - m)^3 + 3n(\bar{\phi} - m) \sum_{i=1}^{n} (\phi_i - m)^2
\]

\[
R_{4, f_n} := \frac{a_4}{4!} \sum_{1 \leq i < j \leq n} (\phi_i + \phi_j - 2m)^4.
\]

Fixing \( b_4 > a_3^2/3a_1 \) arbitrary and introducing the u.n.d \( h_n(\phi) \) as

\[- \log h_n(\phi) := \frac{n(n - 1)}{2} p(m, m) + R_{1, h_n} + R_{3, h_n} + R_{4, h_n},
\]

with \( R_{1, h_n} := \frac{a_1 n}{2} \sum_{i=1}^{n} (\phi_i - m)^2, R_{3, h_n} := \frac{na_3}{3!} \sum_{i=1}^{n} (\phi_i - m)^3, R_{4, h_n} := \frac{b_4 n}{4!} \sum_{i=1}^{n} (\phi_i - m)^4, \)
To complete the proof, with $H_n$ denoting the implied law of the u.n.d. $h_n$ it suffices to show the following:

$$\log\left(\frac{f_n}{h_n}\right) = O_p(1) \text{ under } F_n,$$

$$\log\left(\frac{g_n}{h_n}\right) = O_p(1) \text{ under } G_n,$$

$$\log\left(\frac{f_n}{h_n}\right) = O_p(1) \text{ under } H_n,$$

$$\log\left(\frac{g_n}{h_n}\right) = O_p(1) \text{ under } H_n.$$

Indeed, these implications along with part (a) of Lemma 3.1 gives the desired conclusion. It thus remains to prove these implications.

- $\log\left(\frac{f_n}{h_n}\right) = O_p(1) \text{ under } F_n.$

Noting that

$$|R_{1,f_n} - R_{1,h_n}| \leq O(1) \sum_{i=1}^{n} (\phi_i - m)^2,$$

$$|R_{2,f_n}| \leq O(1) \left[ n^2(\bar{\phi} - m)^2 + \sum_{i=1}^{n} (\phi_i - m)^2 \right]$$

$$|R_{3,f_n} - R_{3,h_n}| \leq O(1) \left[ \sum_{i=1}^{n} (\phi_i - m)^3 + |n(\bar{\phi} - m)| \sum_{i=1}^{n} (\phi_i - m)^2 \right],$$

$$|R_{4,f_n} - R_{4,h_n}| \leq O(1)n \sum_{i=1}^{n} (\phi_i - m)^4,$$

$$|\sum_{i<j} R(\phi_i, \phi_j)| \leq O(1)n \sum_{i=1}^{n} |\phi_i - m|^5,$$

it suffices to show that under $F_n$,

$$n(\bar{\phi} - m) = O_p(1), \quad n^{l/2-1} \sum_{i=1}^{n} |\phi_i - m|^l = O_p(1), l = 2, 3, 4, 5.$$

The first claim follows from Lemma 4.3, and the second claim follows from Lemma 4.2.

- $\log\left(\frac{g_n}{h_n}\right) = O_p(1) \text{ under } G_n.$

As before it suffices to show that under $G_n$,

$$n \sum_{i=1}^{n} (\phi_i - m)^4 = O_p(1), \quad n \sum_{i=1}^{n} (\phi_i - m)^3 = O_p(1), \quad n(\bar{\phi} - m) = O_p(1). \quad (4.3)$$

To this effect, note that there is an orthogonal matrix $A_n$ such that

$$\psi := A_n(\phi - m)1, \quad \psi_1 = \sqrt{n}(\bar{\phi} - m), \quad \sum_{i=2}^{n} \psi_i^2 = \sum_{i=1}^{n} (\phi_i - \bar{\phi})^2.$$
Thus in terms of $\psi$ the u.n.d. becomes
\[
\frac{n(n-1)}{2} p(m,m) + \frac{n(a_1-a_2)}{2} \psi_1^2 + \frac{na_1}{2} \sum_{i=2}^{n} \psi_i^2,
\]
and so $\{\psi_i\}_{i=1}^{n}$ are mutually independent under $G_n$, with
\[
\psi_1 \sim N\left(0, \frac{1}{n(a_1-a_2)}\right), \quad \psi_i \sim N\left(0, \frac{1}{na_1}\right), \quad i \geq 2.
\]
From this representation it is easy to check all the conditions of (4.3).

- $\log(f_n/h_n), \log(g_n/h_n) = O_p(1)$ under $H_n$.

To show the above, it suffices to show the following proposition:

**Proposition 4.1.** Under $H_n$,
\[
n \sum_{i=1}^{n} (\phi_i - m)^3 = O_p(1), \quad n \sum_{i=1}^{n} (\phi_i - m)^4 = O_p(1).
\]

Note that under $\phi_i$'s are i.i.d. under $H_n$. The proof of Proposition 4.1 has been moved to the appendix.

This completes the proof of Lemma 4.4. □

Armed with Lemma 4.4, the proof of Lemma 3.2 for $\theta \in \Theta_1$ is carried out next.

**Proof of Lemma 3.2 for $\theta \in \Theta_1$.** To begin, first note that
\[
\eta_1 = \frac{1}{a_1 - a_2}, \quad \eta_2 = \frac{1}{a_1}.
\]
Now under $G_n$,
\[
n(\bar{\phi} - m) = \sqrt{n} \psi_1 \sim N\left(0, \frac{1}{a_1 - a_2}\right) = N(0, \eta_1).
\]
Also $(\log(f_n/g_n), n(\bar{\phi} - m))$ converges to a joint gaussian distribution under $G_n$. To compute the covariance of this joint limiting distribution, first note that
\[
\left[ n(\bar{\phi} - m)^2, \sum_{i=1}^{n} (\phi_i - \bar{\phi})^2, R_{4,f_n}, \sum_{1 \leq i < j \leq n} R(\phi_i, \phi_j) \right] \xrightarrow{p} \left[ 0, \frac{1}{a_1}, \frac{a_1}{4a_1^2}, 0 \right],
\]
and so the limiting covariance is given by
\[
- \lim_{n \to \infty} E_{G_n} n(\bar{\phi} - m) R_{3,f_n}.
\]
Since
\[
n \sum_{i=1}^{n} (\phi_i - m)^3 = n \sum_{i=1}^{n} (\phi_i - \bar{\phi})^3 + 3n(\bar{\phi} - m) \sum_{i=1}^{n} (\phi_i - \bar{\phi})^2 + n^2(\bar{\phi} - m)^3
\]
with $n^2(\bar{\phi} - m)^3 \xrightarrow{p} 0$, and $(\phi_i - \bar{\phi})^3$ is uncorrelated with $n(\bar{\phi} - m)$, the limiting covariance is same as
\[
- \frac{a_3}{2} \lim_{n \to \infty} E_{G_n} n^2(\bar{\phi} - m)^2 \sum_{i=1}^{n} [(\phi_i - m)^2 + (\phi_i - \bar{\phi})^2] = - \frac{a_3}{a_1(a_1 - a_2)} = -\mu.
\]
Thus by part (a) of Lemma 3.1 the limiting distribution of $n(\bar{\phi} - m)$ under $F_n$ is $N(-\mu, \eta_2)$ which proves (3.3).

(The fact that distribution convergence implies convergence in moments has been used above repeatedly, and this holds here because of uniform integrability implied given by Lemma 4.2.)

(3.6) follows trivially from part (a) of Lemma 4.3.

Finally to prove (3.7), note that under $G_n$

$$\sum_{i=1}^{n}(\phi_i - \bar{\phi})^2 = \sum_{i=2}^{n}\psi_i^2 \overset{P}{\to} \eta_2,$$

and $G_n$ and $F_n$ are mutually contiguous.

4.2. **Non-uniqueness domain $\theta \in \Theta_2$.**

*Proof of Lemma 3.2 for $\theta \in \Theta_3$. By [M, Lemma 4.3]*

$P_{n,\beta}(\phi \in (0, \infty)^n) \overset{n \to \infty}{\to} \frac{1}{2}.$

Upon conditioning on $\phi_i > 0$ for all $i$ and using Lemma 4.2 with $U = (0, \infty)$ and $\phi_0 = m$, a similar analysis as in Domain 1 gives:

$$\mathbb{E}_{P_{n,\beta}}(\sum_{i=1}^{n}(\phi_i - \bar{\phi})^2 | \phi \in (0, \infty)^n) = O(1),$$

$$[n(\bar{\phi} - m) | \phi \in (0, \infty)^n] \overset{d}{\to} N(-\mu, \eta),$$

$$\sum_{i=1}^{n}(\phi_i - \bar{\phi})^2 \overset{P}{\to} \eta_2.$$

Similarly calculations hold on the set $\phi \in (-\infty, 0)^n$, giving

$$\mathbb{E}_{P_{n,\beta}}(\sum_{i=1}^{n}(\phi_i - \bar{\phi})^2 | \phi \in (-\infty, 0)^n) = O(1),$$

$$[n(\bar{\phi} + m) | \phi \in (-\infty, 0)^n] \overset{d}{\to} N(\mu, \eta),$$

$$\sum_{i=1}^{n}(\phi_i - \bar{\phi})^2 \overset{P}{\to} \eta_2.$$

This readily gives all the conclusions (3.4), (3.5), (3.6) and (3.7), thus completing the proof. □
5. Simulations

In both the simulations below the number of vertices \( n \) has been taken to be \( n = 100 \), and the burn in period has been taken to be 200. The plotted diagrams are the histograms of \( S_1 \), which is a linear translate of the number of edges. The number of independent samples drawn for each histogram is 5000.

5.1. Domain 1. The first figure shows the histogram and qq-plot of \( S_1 \) for \((\theta_1 = 0, \theta_2 = .25) \in \Theta_{11}\). The number of bins for the histogram is 50.

![Histogram and qq-plot of number of edges in the first domain](image)

**Figure 1.** Histogram and qq-plot of number of edges in the first domain

The histogram has an approximate uni-modular bell shape, and the qq-plot confirms that the data is close to normal. This confirms the predictions of Theorem 2.1 that \( S_1 \) is asymptotically normal.
5.2. **Domain 2.** The second figure shows the histogram for simulations from \((\theta_1 = 0, \theta_2 = .55) \in \Theta_2\) with 80 bins. The figure shows two well separated histograms.
In the third figure each of the histograms above are zoomed in. The number of bins chosen for each of the histograms is 50.

![Histograms zoomed in](image)

**Figure 3.** Separate histograms for positive and negative values

Each histogram is again approximately bell shaped uni-modal, thus confirming the predictions that both the conditional distributions are normal.
The fourth figure shows the qq-plot of the positive and negative values separately. This again confirms that each of them is roughly normal.

**Figure 4.** Separate qq-plots for positive and negative values
6. Acknowledgement

I am grateful to my advisor Dr. Persi Diaconis for introducing me to this problem, and for his continued help and support during my Ph.D.

References

[AWC] C. J. Anderson, S. Wasserman and B. Crouch, A $p^*$ primer: logit models for social networks. Social Networks 21: 37-66, 1999.
[B1] J. Besag, Spatial Interaction and the Statistical Analysis of Lattice Systems. Journal of the Royal Statistical Society. Series B (Methodological). 36 (2): 192-236, 1974.
[B2] J. Besag, Statistical Analysis of Non-Lattice Data. Journal of the Royal Statistical Society. Series D (The Statistician). 24 (3): 179-195, 1975.
[BSB] S. Bhamidi, S. A.Sly and G. Bresler, Mixing time of exponential random graphs, Annals of Applied Probability 21, 2146-2170, 2011.
[BD] J. Blitzstein and P. Diaconis, A Sequential Importance Sampling Algorithm for Generating Random Graphs with Prescribed Degrees. Internet Mathematics. 6 (4): 489-522, 2011.
[C] S. Chatterjee, Estimation in spin glasses: A first step. Ann. Statist., 35 no. 5, 1931-1946, 2007.
[CD] S. Chatterjee and P. Diaconis. Estimating and Understanding Exponential Random Graph Models. To appear in Ann. Statist.
[CDS] S. Chatterjee, P. Diaconis and A. Sly, Random graphs with a given degree sequence. Annals of Applied Probability, 21, 4, 1400-1435, 2011.
[CS] S. Chatterjee and Q.M. Shao, Nonnormal approximation by Steins method of exchangeable pairs with application to the Curie-Weiss model. The Annals of Applied Probability. 21 (2): 464-483, 2011.
[DM] A. Dembo and A. Montanari, Gibbs measures and phase transitions on sparse random graphs. Brazilian J. of Probab. and Stat. 24, pp. 137-211, 2010.
[FS] O. Frank and D. Strauss, Markov Graphs. J. Amer. Statist. Assoc., 81, 832842, 1986.
[Gr] G. Grimmett, The random-cluster model. Berlin : New York : Springer, c2006.
[GT] C. Geyer and E. Thompson, Constrained Monte Carlo Maximum Likelihood for Dependent Data. Journal of the Royal Statistical Society. Series B (methodological), 54, 3, 657-699, 1992.
[H] M. Handcock, Assessing Degeneracy in Statistical Models of Social Networks. Working Paper no. 39, Center for Statistics and the Social Sciences University of Washington, 2003.
[HI] P. Holland and S. Leinhardt, An Exponential Family of Probability Distributions for Directed Graphs. Journal of the American Statistical Association. 76 (373): 33-50, 1981.
[MHH] M. Morris, D. Hunter, D and M. Handcock, Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects. Journal of Statistical Software 42(4), 2008.
[Newman] M.E.J. Newman, The Structure and Function of Complex Networks. SIAM Rev., 45(2), 167256, 2003.
[PN] J. Park and M.E.J. Newman, Solution of the 2-star model of a network, Phys. Rev. E 70, 066146, 2004.
[PN2] J. Park and M.E.J. Newman, Solution for the properties of a clustered network. Phys. Rev. E (3), 72 026136, 5, 2005.
[PW] S. Wasserman and P. Pattison, Logit models and logistic regressions for social networks. I. An introduction to Markov graphs and p. Psychometrika, 61, 401425, 1996.
[M] S. Mukherjee, Phase transition in the two star Exponential Random Graph Model. http://arxiv.org/abs/1310.4164.
[R] G. Roussas, Contiguity of probability measures: some applications in statistics. Cambridge : Cambridge University Press, 1972.
[RPKL] G. Robins, P. Pattison, Y. Kalish and D. Lusher, An introduction to exponential random graph $p^*$ models for social networks. Social Networks, 29, 2, 173-191, 2007.
[S] T. A. B. Snijders, Markov chain Monte Carlo estimation of exponential random graph models. J. Social Structure, 2, 2002.
[S] Strauss, D. Strauss, On a General Class of Models for Interaction. SIAM Review. 28 (4): 513-527, 1986.
[S] T.A.B. Snijders, P. Pattison, G.L. Robins and M. Handcock, New Specifications for Exponential Random Graph Models. Sociological Methodology;36:99153, 2006.
[WF] S. Wasserman and K. Faust, Social network analysis: Methods and applications. Cambridge: Cambridge University Press, 1994.
7. Appendix

Proof of Lemma 3.1. Let $A_n$ be a sequence of sets such that $\mathbb{G}_n(A_n) \to 0$. It suffices to show that $\mathbb{H}_n(A_n) \to 0$, as then the other implication also follows by symmetry.

Fix $0 < \epsilon < 1$, arbitrary, and let $M = M(\epsilon)$ be such that

$$\mathbb{H}_n(B_n(\epsilon)) \geq 1 - \epsilon, \mathbb{G}_n(B_n(\epsilon)) \geq 1 - \epsilon, \quad B_n(\epsilon) := \{e^{-M} \leq \frac{h_n}{g_n} \leq e^M\}$$

Thus setting $\epsilon = 1/2$,

$$b_n = \int g_n d\mu_n = \int_{B_n(1/2)} g_n d\mu_n + \int_{B_n(1/2)^c} g_n d\mu_n \leq e^{M(1/2)} \int_{B_n(1/2)} h_n d\mu_n + \frac{b_n}{2} \leq e^{M(1/2)} a_n + \frac{b_n}{2}.$$ 

This gives $b_n \leq 2e^{M(1/2)}a_n$, i.e. $b_n/a_n$ is bounded above. Now for any $\epsilon > 0$,

$$\mathbb{H}_n(A_n) \leq \mathbb{H}_n(A_n \cap B_n(\epsilon)) + \epsilon = \frac{\int_{A_n \cap B_n(\epsilon)} h_n d\mu_n}{a_n} + \epsilon \leq e^{M(\epsilon)} \mathbb{G}_n(A_n) \frac{b_n}{a_n} + \epsilon \leq 2e^{M(\epsilon) + M(1/2)} \mathbb{G}_n(A_n) + \epsilon$$

Taking lim sup on both sides as $n \to \infty$ gives lim sup $\mathbb{H}_n(A_n) \leq \epsilon$. Since $\epsilon < 1$ is arbitrary, contiguity follows.

(a) Note that the above proof implies that $r_n := \log a_n - \log b_n$ is a bounded sequence of reals, and so there is a subsequence $\{n_k\}$ along which $r_{n_k} \to r \in \mathbb{R}$. Then with

$$h_n := \frac{h_n}{a_n}, g_n := \frac{g_n}{b_n}, L_n := \log h_n - \log g_n = L_n - r_n$$

$$(X_{n_k}, (L_{n_k})) \overset{d, G_{n_k}}{\to} N(\mu_1, \mu_2 - r, \sigma_1^2, \sigma_2^2, \sigma_{12})$$

But then [R, Corollary 7.1, Chapter 1] gives $r = \mu_2 + \frac{\sigma_2^2}{2}$, and so any convergent subsequence of $r_n$ has to converge to $\mu_2 + \frac{\sigma_2^2}{2}$. Thus $r_n$ converges, and so

$$(X_n, (L_n)) \overset{d, G_n}{\to} N(\mu_1, \mu_2 - r, \sigma_1^2, \sigma_2^2, \sigma_{12}),$$

from which [R, Theorem 7.1, Chapter 1] gives

$$X_n \overset{d, \mathbb{H}_n}{\to} N(\mu_1 + \sigma_{12}, \sigma_1^2).$$

(b) Since $\mathbb{H}_n$ and $\mathbb{G}_n$ are mutually contiguous, it follows that $L_n \overset{d, \mathbb{H}_n}{\to} c$. Fixing $\epsilon, \delta > 0$ arbitrary, for all large $n$,

$$\mathbb{H}_n(B_n) > 1 - \epsilon, \mathbb{G}_n(B_n) > 1 - \epsilon, \quad B_n := \{e^{c-\delta} \leq \frac{h_n}{g_n} < e^{c+\delta}\}.$$ 

Thus

$$b_n = \int_{B_n} g_n d\mu_n + \int_{\overline{B_n}} g_n d\mu_n \leq e^{-c-\delta} \int_{B_n} h_n d\mu_n + \epsilon b_n \leq e^{-c-\delta} a_n + \epsilon b_n$$

which gives $\frac{b_n}{a_n} \leq \frac{e^{-c-\delta}}{1-\epsilon}$.
Similar calculations give \( \frac{a_n}{b_n} \leq \frac{e^r}{1-r} \), and so \( \frac{a_n}{b_n} \to e^c \). Consequently \( \frac{H_n}{G_n} \sim d \) under both \( H_n \) and \( G_n \). Letting \( \epsilon > 0 \) be fixed and \( C_n := \{ |H_n/G_n| - 1| > \epsilon \} \),

\[
2\|H_n - G_n\|_{TV} \leq \int_{C_n} |H_n + G_n| d\mu_n + \int_{C_n} |H_n + G_n - 1| d\mu_n \leq H_n(C_n) + G_n(C_n) + \epsilon,
\]

which gives

\[
\limsup_{n \to \infty} \|H_n - G_n\|_{TV} \leq \epsilon/2.
\]

This completes the proof of part (b).

\[\square\]

**Proof of Proposition 4.1.** Note that \( \phi_i - m \) are i.i.d. with density proportional to

\[
e^{-\frac{a_1}{2} x^2 - \frac{a_3}{3!} x^3 + \frac{b_4}{4!} x^4} = e^{-nx^2\eta(x)}, \quad \eta(x) := \frac{a_1}{2} x + \frac{a_3}{3!} x + \frac{b_4}{4!} x^2
\]

Thus it suffices to show that for any \( l \in \mathbb{Z}, l \geq 0 \),

\[
b_{l,n} := \int_{-\infty}^{\infty} x^l e^{-nx^2\eta(x)} dx
\]

satisfies

\[
b_{l,n} = -\frac{C_l}{\sqrt{n}} (1 + O(n^{-1/2})) \text{ if } n \text{ is odd }, \tag{7.1}\\
= \frac{D_l}{\sqrt{n}} (1 + O(n^{-1/2})) \text{ if } n \text{ is even}, \tag{7.2}
\]

where

\[
C_l := \frac{a_3}{3!} \sqrt{\left( \frac{2\pi}{a_1^{1+4}} \right)} E Z^{l+3}, \quad D_l := \sqrt{\left( \frac{2\pi}{a_1^{2+4}} \right)} E Z^l,
\]

with \( Z \sim N(0, 1) \). Indeed, given (7.2), setting \( l = 4 \) it follows that

\[
n \mathbb{E}_{\mathbb{G}_n} \sum_{i=1}^{n} (\phi_i - \phi_0)^4 = nb_{4,n}/b_{0,n} = O(1),
\]

proving \( \sum_{i=1}^{n} (\phi_i - m)^4 = O_p(1) \). Also (7.2) along with Cauchy-Schwarz inequality gives

\[
\mathbb{E}_{\mathbb{G}_1,n} \left[ n \sum_{i=1}^{n} (\phi_i - m)^3 \right]^2 \leq n^3 \frac{b_{6,n}}{b_{0,n}} = O(1),
\]

proving \( \sum_{i=1}^{n} (\phi_i - m)^3 = O_p(1) \), which completes the proof of the Proposition.

Turning to prove (7.1) and (7.2), first note that \( 3b_4a_1 > a_3^2 \), and so the discriminant of \( \eta(x) \) is negative. Thus \( \eta \) has no real roots on \( \mathbb{R} \), and consequently

\[
\inf_{x \in \mathbb{R}} \eta(x) =: d/2 > 0.
\]
Fixing $\gamma > 0$ write

\[ b_{l,n} = \int_{|x| > \sqrt{\gamma \log n}} x^l e^{-nx^2 \eta(x)} \, dx + \int_{|x| \leq \sqrt{\gamma \log n}} x^l e^{-nx^2 \eta(x)} \, dx \]

The first integral can be bounded by

\[ \int_{|x| > \sqrt{\gamma \log n}} |x|^l e^{-\frac{\eta d x^2}{2}} \, dx \leq \left( \frac{1}{\sqrt{nd}} \right)^{l+1} \sqrt{2\pi} \mathbb{E}(|Z| \mathbb{I}_{|Z| \geq \gamma d \log n}) \leq \left( \frac{1}{\sqrt{nd}} \right)^{l+1} \sqrt{2\pi} \sqrt{n} \mathbb{P}(|Z| > \gamma d \log n), \]

where $Z \sim N(0,1)$. Proceeding to estimate the second integral, write it as

\[ \left( \frac{1}{\sqrt{n}} \right)^{l+1} \int_{|x| \leq \sqrt{\gamma \log n}} x^l e^{-\frac{a_1 x^2}{2}} e^{-\frac{a_3}{3!} \frac{x^3}{\sqrt{n}}} - \frac{b_4}{4!} \frac{x^4}{n} \, dx \]

Using the fact that $\exp\{-x\} = 1 - x + O(x^2)$ on $[-1,1]$, the integrand can be written as

\[ x^l e^{-\frac{a_1 x^2}{2}} \left[ 1 - a_3 \frac{x^3}{3!} \sqrt{n} + O\left( \frac{x^4}{n} \right) \right] \]

and integrate term by term to get, for $l$ odd, the following estimate for the integral:

\[ -\frac{a_3}{3!} \sqrt{n} \int_{-\sqrt{\gamma \log n}}^{\sqrt{\gamma \log n}} x^{l+3} e^{-\frac{a_1 x^2}{2}} \, dx + O\left( \frac{1}{n} \right). \]

Finally noting that

\[ \int_{-\sqrt{\gamma \log n}}^{\sqrt{\gamma \log n}} x^{l+3} e^{-\frac{a_1 x^2}{2}} \, dx = \sqrt{\frac{2\pi}{a_1^{l+4}}} \mathbb{E}[Z^{l+3}, |Z| \leq \sqrt{a_1 \gamma \log n}] \]

\[ = \sqrt{\frac{2\pi}{a_1^{l+4}}} \left[ \mathbb{E}Z^{l+3} - \mathbb{E}Z^{l+6} \sqrt{n} \mathbb{P}(|Z| \geq \sqrt{a_1 \gamma \log n}) \right], \]

choosing $\gamma$ fixed but large it follows that

\[ b_{l,n} = -\frac{C_l}{(\sqrt{n})^{l+2}} \left( 1 + O(n^{-1/2}) \right). \]

The case for $l$ even follows by a similar computation.

\[ \Box \]

† Department of Statistics, Stanford University
Sequoia Hall, 390 Serra Mall, Stanford, California 94305