Statistics and Its Interface Volume 8 (2015) 255–266

A simulation study of the p_1 model for directed random graphs

Ting Yan*,† and Chenlei Leng

The p_1 exponential-family distribution in Holland and Leinhardt (1981) is one of the earliest models for directed random graphs and has been widely used in practice. The conditions for the existence and uniqueness of the maximum likelihood estimate (MLE) have been derived. However, it is a daunting task to investigate the large-sample properties of the MLE theoretically as the number of graphical vertices goes to infinity. The uniform consistency and asymptotic normality of the MLE have been derived for some specialized models closely related to the p_1 model but general results are lacking. In this article, we explore the consistency and asymptotic normality of the MLE in the p_1 model as the network size increases through numerical simulations. The results indicate that the p_1 model also enjoys good asymptotic properties.

AMS 2000 subject classifications: Primary 62F12; secondary 62F10, 62E20, 05C80.
Keywords and phrases: Directed random graphs, Increasing number of parameters, Maximum likelihood estimation, p_1 exponential-family distribution.

1. INTRODUCTION

Complex social relationships can be conveniently represented by a directed graph, in which each vertex stands for a person and the directed edges indicate directed relationships between these individuals (Leinhardt, 1977). Such graph data are ubiquitous not only in social sciences but also in various other fields such as online communication networks, food webs and protein-protein interaction networks. We assume that there are no self-edges (one vertex connecting to itself) and at most one edge exists between any two distinct vertices in a given direction. To allow for the simultaneous estimation of parameters that measure both the strength of reciprocation of directed edges and the differential attractiveness exhibited by each vertex, Holland and Leinhardt (1981) proposed the earliest p_1 exponential-family distribution to model the directed network data.

*Corresponding author.
†Partially supported by the National Science Foundation of China (No. 11341001)

The p_1 model can be summarized as follows. Assume that there are n vertices $1,\dots,n$ in a directed graph. Let $X=(X_{ij})$ be its adjacent matrix, where

$$X_{ij}=\begin{cases}1, & \text{there is an edge from } i \text{ pointing to } j,\\ 0, & \text{otherwise.}\end{cases}$$

Denote $D_{ij}=(X_{ij},X_{ji})$, $i<j$, as the $n(n-1)/2$ dyads. By assuming that $\{D_{ij}\}_{1\le i<j\le n}$ are mutually independent, the p_1 probability distribution for each D_{ij} $(1\le i<j\le n)$ is specified as

$$\begin{aligned}p_{ij}(1,1) &:= P(D_{ij}=(1,1))\\ &= \frac{1}{k_{ij}}\exp(\rho+2\theta+\alpha_i+\alpha_j+\beta_i+\beta_j),\\ p_{ij}(1,0) &:= P(D_{ij}=(1,0)) = \frac{1}{k_{ij}}\exp(\theta+\alpha_i+\beta_j),\\ p_{ij}(0,1) &:= P(D_{ij}=(0,1)) = \frac{1}{k_{ij}}\exp(\theta+\alpha_j+\beta_i),\\ p_{ij}(0,0) &:= P(D_{ij}=(0,0)) = \frac{1}{k_{ij}},\end{aligned}$$

where

$$\begin{aligned}k_{ij} = 1+\exp(\theta+\alpha_i+\beta_j)+\exp(\theta+\alpha_j+\beta_i)\\ +\exp(\rho+2\theta+\alpha_i+\alpha_j+\beta_i+\beta_j)\end{aligned}$$

is the normalization constant. The parameter ρ measures the average tendency toward reciprocation for all pairs of vertices, θ is a density parameter of edges, and α_i quantifies the effect of an outgoing edge from vertex i. If α_i is large and positive, vertex i will tend to have a relatively large out-degree. On the other hand, β_j quantifies the effect of an incoming edge connecting to vertex j. If β_j is large and positive, vertex j will tend to have a large in-degree. The likelihood for the p_1 model is easily seen as

$$P(X=x)=\exp\Bigg(\rho m+\theta x_{++}+\sum_{i=1}^{n}\alpha_i x_{i+}+\sum_{j=1}^{n}\beta_j x_{+j}\Bigg)\times\Bigg(\prod_{1\le i<j\le n}k_{ij}\Bigg)^{-1}, \tag{1}$$

where m, x_{++}, x_{i+}, and x_{+j} are the observed values of $M=\sum_{1\le i<j\le n}X_{ij}X_{ji}$, $X_{++}=\sum_{i,j=1}^{n}X_{ij}$, $X_{i+}=\sum_{j=1}^{n}X_{ij}$,

and $X_{+j} = \sum_{i=1}^{n} X_{ij}$ respectively. For parameter identification, α_i and β_j are subject to the constraint that $\sum_{i=1}^{n} \alpha_i = 0$ and $\sum_{i=1}^{n} \beta_j = 0$. Therefore the dimension of the parameter space is $2n$ although the p_1 model is parametrized by a total of $2n + 2$ parameters.

Since Holland and Leinhardt (1981), various versions of exponential random graph models (ERGMs) have been proposed in which algorithms for obtaining the maximum likelihood estimate (MLE) are introduced as well [cf., Frank and Strauss (1986); Wang and Wong (1987); Wasserman and Pattison (1996); Wasserman and Robins (2005); Robins et al. 2007(a, b)]. However, there is a huge gap between the model development and the asymptotic theories of the MLE in ERGMs. Even for the simple p_1 model introduced by Holland and Leinhardt (1981), asymptotic results have not yet been provided. This is partly because networks are a non-standard type of data in which edges may be dependent and the number of parameters may be comparable to the size of network. As far as we know, the consistency and asymptotic normality of the MLE are derived only for selected models simpler than the p_1 model when the size of network goes to infinity. For example, the β-model coined by Chatterjee et al. (2011) with the degree sequence as its natural sufficient statistics, is a simple undirected version of p_1 model. It can also be seen as a heterozygous version of the Erdős-Rényi model (Erdős and Rényi, 1959) for undirected random graphs. The asymptotic and non-asymptotic properties of the β-model have been basically understood. In particular, Chatterjee et al. (2011) proved that the MLE in the β-model is uniformly consistent as the number of parameters goes to infinity, and developed a simple algorithm for calculating the MLE. Yan and Xu (2013) established asymptotic normality of the MLE by using a simple matrix to approximate the inverse of the Fisher information matrix, in which the asymptotic variance of the parameter for vertex i is the ith diagonal element of the inverse of the Fisher information matrix. Rinaldo et al. (2013) used the polytope of degree sequences to derive necessary and sufficient conditions for the existence and uniqueness of the MLE for the β-model as well as other network models including the p_1 model. Blitzstein and Diaconis (2011) developed a sequential importance sampling algorithm for generating a random graph with a given degree sequence.

Another exponential model closely related to the p_1 model is the Bradley-Terry model (Bradley and Terry, 1952) for ranking and rating individuals joined in paired comparisons. As pointed out by Frank (1981), the outcomes of paired comparisons can be represented in a directed random graph whose vertices represent the individuals and the weighted directed edges indicate the number of times that one individual is preferred to another individual. By assuming that each pair has the same number of comparisons, Simons and Yao (1999) proved that the MLE is uniformly consistent and asymptotically normal when the number of parameters goes to infinity. The proof of asymptotic normality is conducted through approximating the inverse of the Fisher information matrix. The asymptotic variances of the MLEs are the diagonal elements of the inverse of the Fisher information matrix. The Rasch model (Rasch, 1960) that are widely used for analyzing item response experiments, whose outcomes can be represented in a bipartite graph, is also a cousin to the p_1 model (Haberman, 1981). By assuming that the item parameters are bounded, Haberman (1977) proved that the MLE is uniformly consistent and asymptotically normal as the number of items goes to infinity.

For the p_1 model, Holland and Leinhardt (1981) provided a simple generalized iterative scaling algorithm for solving the MLE. This algorithm does not directly solve the MLE, but obtain the MLEs $\hat{p}_{ij}(a, b)$ of $p_{ij}(a, b), a, b = 0, 1$. Even if the MLE does not exist (i.e., the MLE is unbounded), corresponding to that some values of $\hat{p}_{ij}(a, b), a, b = 0, 1$ equal zero or one, the iterative scaling algorithm automatically adjusts for this and converges. Moreover, the computations for the iterative scaling algorithm are simple, only requiring row and column multiplications. For a given adjacent matrix $X = x$, in the online supplementary materials of Rinaldo et al. (2013) and Rinaldo et al. (2010), they derived the necessary and sufficient conditions for the existence and uniqueness of the MLE in the p_1 model. Feinberg et al. (2011) and Petrović (2010) carried out an algebraic statistics analysis of the p_1 model. As pointed out by these authors, investigating the asymptotic behaviors of the MLE is a daunting task since the model complexity is twice as many as the network size and statistical inferences are based on only one realization of the network. Although theoretical investigation is challenging, numerical studies are always possible. The aim of this article is to carry out simulation studies to investigate the asymptotic properties of the MLE. This is done in the next section. Further discussion is given in Section 3.

2. SIMULATIONS

Let $\boldsymbol{\varphi} = (\theta, \rho, \alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n)'$ and $\widehat{\boldsymbol{\varphi}}$ be its MLE. We evaluate the accuracy of the MLE through simulation studies by recording the average values of $\hat{\theta} - \theta$, $\hat{\rho} - \rho$, $\max_i |\hat{\alpha}_i - \alpha_i|$ and $\max_j |\hat{\beta}_j - \beta_j|$ as well as the probabilities that the MLE does not exist. Let V be the Fisher information matrix of $\boldsymbol{\varphi}$. Due to the restrictions $\sum_i \alpha_i = 0$ and $\sum_j \beta_j = 0$, V is not invertible. We use $\boldsymbol{\phi} = (\theta, \rho, \alpha_1, \ldots, \alpha_{n-1}, \beta_1, \ldots, \beta_{n-1})'$ as the independent parameters and V^* as its Fisher information matrix, whose expression is given in Appendix. Let U be the inverse matrix of V^*, and $\widehat{U}$ be its estimated value by replacing $\boldsymbol{\varphi}$ in U with $\widehat{\boldsymbol{\varphi}}$. Similar to Yan and Xu (2013), we use U as the covariance matrix of $\widehat{\boldsymbol{\phi}}$. In particular, the values of $(\hat{\theta} - \theta)/(\widehat{U}_{11})^{1/2}$, $(\hat{\rho} - \rho)/(\widehat{U}_{22})^{1/2}$,

$$\xi_i := \frac{(\hat{\alpha}_i - \alpha_i)}{(\widehat{U}_{2+i,2+i})^{1/2}}, \qquad \eta_j := \frac{(\hat{\beta}_j - \beta_j)}{(\widehat{U}_{n+1+j,n+1+j})^{1/2}},$$

were recorded. The quantile-quantile (QQ) plots were drawn to assess the asymptotic normality of these values. We also

Table 1. The accuracy of estimation for $\theta = 0$ and $\rho = 0$. The values in parentheses and square brackets are the estimated standard errors (the corresponding entries of the inverse of Fisher information matrix) and the frequencies that the MLE does not exist. If the square brackets are not shown, then the frequencies are zeros

L_n	0	$[\log(\log(n))]^{1/2}$	$\log(\log(n))$	$\log(n)$
		$n = 100$		
$\lvert\hat{\theta} - \theta\rvert$	0.029(0.035)	0.033(0.043)	0.036(0.048)	0.056(0.053)[0.2]
$\lvert\hat{\rho} - \rho\rvert$	0.048(0.058)	0.054(0.072)	0.061(0.083)	0.095(0.081)[0.2]
$\max_i \lvert\hat{\alpha}_i - \alpha_i\rvert$	0.566	0.648	0.683	1.104[0.2]
$\max_j \lvert\hat{\beta}_j - \beta_j\rvert$	0.568	0.646	0.687	1.106[0.2]
$\lvert\hat{\alpha}_1 - \alpha_1\rvert$	0.162(0.203)	0.169(0.214)	0.182(0.221)	0.254(0.318)[0.2]
$\lvert\hat{\alpha}_{n/2} - \alpha_{n/2}\rvert$	0.161(0.203)	0.199(0.248)	0.212(0.270)	0.388(0.446)[0.2]
$\lvert\hat{\alpha}_{n-1} - \alpha_{n-1}\rvert$	0.161(0.203)	0.193(0.243)	0.208(0.262)	0.371(0.429)[0.2]
$\lvert\hat{\beta}_1 - \beta_1\rvert$	0.163(0.203)	0.175(0.216)	0.175(0.222)	0.259(0.319)[0.2]
$\lvert\hat{\beta}_{n/2} - \beta_{n/2}\rvert$	0.159(0.203)	0.197(0.248)	0.213(0.269)	0.377(0.438)[0.2]
$\lvert\hat{\beta}_{n-1} - \beta_{n-1}\rvert$	0.162(0.203)	0.199(0.245)	0.217(0.263)	0.370(0.429)[0.2]
		$n = 500$		
$\lvert\hat{\theta} - \theta\rvert$	0.006(0.007)	0.007(0.009)	0.008(0.011)	0.145(0.012)
$\lvert\hat{\rho} - \rho\rvert$	0.009(0.011)	0.011(0.015)	0.014(0.019)	0.261(0.018)
$\max_i \lvert\hat{\alpha}_i - \alpha_i\rvert$	0.292	0.336	0.360	0.946
$\max_j \lvert\hat{\beta}_j - \beta_j\rvert$	0.292	0.335	0.362	0.947
$\lvert\hat{\alpha}_1 - \alpha_1\rvert$	0.073(0.090)	0.075(0.096)	0.083(0.101)	0.117(0.155)
$\lvert\hat{\alpha}_{n/2} - \alpha_{n/2}\rvert$	0.069(0.090)	0.088(0.112)	0.102(0.126)	0.529(0.218)
$\lvert\hat{\alpha}_{n-1} - \alpha_{n-1}\rvert$	0.070(0.090)	0.090(0.111)	0.098(0.125)	0.470(0.221)
$\lvert\hat{\beta}_1 - \beta_1\rvert$	0.074(0.090)	0.075(0.096)	0.079(0.101)	0.120(0.155)
$\lvert\hat{\beta}_{n/2} - \beta_{n/2}\rvert$	0.072(0.090)	0.091(0.112)	0.098(0.125)	0.554(0.216)
$\lvert\hat{\beta}_{n-1} - \beta_{n-1}\rvert$	0.070(0.090)	0.091(0.111)	0.101(0.125)	0.472(0.221)
		$n = 1{,}000$		
$\hat{\theta} - \theta$	0.003(0.003)	0.003(0.004)	0.004(0.006)	0.009(0.007)
$\hat{\rho} - \rho$	0.004(0.006)	0.006(0.007)	0.006(0.011)	0.015(0.011)
$\max_i \lvert\hat{\alpha}_i - \alpha_i\rvert$	0.218	0.251	0.277	0.468
$\max_j \lvert\hat{\beta}_j - \beta_j\rvert$	0.219	0.252	0.276	0.469
$\lvert\hat{\alpha}_1 - \alpha_1\rvert$	0.049(0.063)	0.056(0.068)	0.059(0.072)	0.095(0.118)
$\lvert\hat{\alpha}_{n/2} - \alpha_{n/2}\rvert$	0.052(0.063)	0.064(0.079)	0.070(0.091)	0.136(0.168)
$\lvert\hat{\alpha}_{n-1} - \alpha_{n-1}\rvert$	0.053(0.063)	0.065(0.080)	0.074(0.090)	0.137(0.167)
$\lvert\hat{\beta}_1 - \beta_1\rvert$	0.049(0.063)	0.054(0.068)	0.056(0.072)	0.094(0.118)
$\lvert\hat{\beta}_{n/2} - \beta_{n/2}\rvert$	0.049(0.063)	0.063(0.080)	0.068(0.091)	0.137(0.169)
$\lvert\hat{\beta}_{n-1} - \beta_{n-1}\rvert$	0.050(0.063)	0.065(0.079)	0.067(0.090)	0.137(0.167)

considered the combinations $\sum_{i=1}^r c_i\hat{\alpha}_i$ and $\sum_{i=1}^r c_i\hat{\beta}_i$ for some fixed r. For the sake of simplicity, we set $r = 2$. For some pairs (i, j), the QQ-plots of

$$\xi_{ij} := \frac{\hat{\alpha}_i - \hat{\alpha}_j - (\alpha_i - \alpha_j)}{\sqrt{\widehat{U}_{2+i,2+i} + \widehat{U}_{2+j,2+j} - 2\widehat{U}_{2+i,2+j}}},$$

$$\eta_{ij} := \frac{\hat{\beta}_i - \hat{\beta}_j - (\beta_i - \beta_j)}{\sqrt{\widehat{U}_{n+1+i,n+1+i} + \widehat{U}_{n+1+j,n+1+j} - 2\widehat{U}_{n+1+i,n+1+j}}}$$

were also depicted.

Following Yan and Xu (2013), the parameters α_i and β_i are specified via a linear type as

$$\beta_i = \alpha_i = \begin{cases} iL_n/(n/2), & i = 1, \ldots, n/2 \\ -(i - n/2)L_n/(n/2), & i = n/2 + 1, \ldots, n, \end{cases}$$

such that $\sum_i \alpha_i = 0$, $\sum_i \beta_i = 0$ and $\max_i |\alpha_i| = \max_i |\beta_i| = L_n$. Here we set $(\alpha_1, \ldots, \alpha_n) = (\beta_1, \ldots, \beta_n)$ considering that these two vectors have similar asymptotic behaviors since they are row and column exchangeable. Four different combinations of $(\theta, \rho) = (0, 0)$, $(-2, 0.5)$, $(-\log(\log n), \log(\log n))$ and $(-\log n, \log n)$ were considered. Since most of the real network data are sparse, we do not consider any positive density parameter for θ here. In fact, in the friendship data example of Holland and Leinhardt (1981), $\hat{\theta} = -2.5$. Due to the consideration of computation time, we only simulated three values for n as $n = 100$, 500, or 1,000, and four different values of L_n as 0, $[\log(\log n)]^{1/2}$, $\log(\log n)$ and $\log n$. Each simulation was based on 1,000 repetitions.

When $(\theta, \rho) = (-\log n, \log n)$, the frequencies that the MLE does not exist are close to 100% because under this

Table 2. The accuracy of estimation for $\theta = -2$ and $\rho = 0.5$. The values in parentheses and square brackets are the estimated standard errors (the corresponding entries of the inverse of Fisher information matrix) and the frequencies that the MLE does not exist. If the square brackets are not shown, then the frequencies are zeros

L_n	0	$[\log(\log(n))]^{1/2}$	$\log(\log(n))$	$\log(n)$
		$n = 100$		
$\lvert\hat{\theta} - \theta\rvert$	0.075(0.037)	0.082(0.041)[0.076]	0.085(0.043)[0.172]	0.091(0.059)[0.908]
$\lvert\hat{\rho} - \rho\rvert$	0.095(0.116)	0.080(0.095)[0.076]	0.078(0.093)[0.172]	0.146(0.090)[0.908]
$\max_i \lvert\hat{\alpha}_i - \alpha_i\rvert$	1.010	1.252[0.076]	1.303[0.172]	1.321[0.908]
$\max_j \lvert\hat{\beta}_j - \beta_j\rvert$	1.003	1.256[0.076]	1.300[0.172]	1.317[0.908]
$\lvert\hat{\alpha}_1 - \alpha_1\rvert$	0.254(0.307)	0.236(0.293)[0.076]	0.229(0.288)[0.172]	0.257(0.331)[0.908]
$\lvert\hat{\alpha}_{n/2} - \alpha_{n/2}\rvert$	0.251(0.308)	0.189(0.227)[0.076]	0.199(0.229)[0.172]	0.296(0.335)[0.908]
$\lvert\hat{\alpha}_{n-1} - \alpha_{n-1}\rvert$	0.250(0.307)	0.405(0.497)[0.076]	0.447(0.543)[0.172]	0.499(0.806)[0.908]
$\lvert\hat{\beta}_1 - \beta_1\rvert$	0.249(0.308)	0.233(0.293)[0.076]	0.231(0.289)[0.172]	0.258(0.331)[0.908]
$\lvert\hat{\beta}_{n/2} - \beta_{n/2}\rvert$	0.250(0.308)	0.188(0.227)[0.076]	0.189(0.231)[0.172]	0.333(0.339)[0.908]
$\lvert\hat{\beta}_{n-1} - \beta_{n-1}\rvert$	0.250(0.307)	0.410(0.496)[0.076]	0.428(0.538)[0.172]	0.470(0.797)[0.908]
		$n = 500$		
$\lvert\hat{\theta} - \theta\rvert$	0.014(0.007)	0.016(0.008)	0.017(0.008)	0.028(0.013)[0.029]
$\lvert\hat{\rho} - \rho\rvert$	0.018(0.023)	0.015(0.018)	0.015(0.018)	0.240(0.020)[0.029]
$\max_i \lvert\hat{\alpha}_i - \alpha_i\rvert$	0.460	0.611	0.685	1.146[0.029]
$\max_j \lvert\hat{\beta}_j - \beta_j\rvert$	0.464	0.610	0.686	1.152[0.029]
$\lvert\hat{\alpha}_1 - \alpha_1\rvert$	0.110(0.135)	0.102(0.128)	0.100(0.126)	0.129(0.162)[0.029]
$\lvert\hat{\alpha}_{n/2} - \alpha_{n/2}\rvert$	0.108(0.135)	0.079(0.099)	0.085(0.103)	0.477(0.176)[0.029]
$\lvert\hat{\alpha}_{n-1} - \alpha_{n-1}\rvert$	0.105(0.135)	0.172(0.219)	0.199(0.251)	0.502(0.478)[0.029]
$\lvert\hat{\beta}_1 - \beta_1\rvert$	0.110(0.135)	0.098(0.129)	0.106(0.126)	0.125(0.162)[0.029]
$\lvert\hat{\beta}_{n/2} - \beta_{n/2}\rvert$	0.108(0.135)	0.081(0.099)	0.085(0.104)	0.491(0.176)[0.029]
$\lvert\hat{\beta}_{n-1} - \beta_{n-1}\rvert$	0.110(0.135)	0.178(0.219)	0.198(0.250)	0.506(0.470)[0.029]
		$n = 1{,}000$		
$\hat{\theta} - \theta$	0.007(0.004)	0.008(0.004)	0.008(0.004)	0.015(0.007)[0.004]
$\hat{\rho} - \rho$	0.009(0.011)	0.007(0.009)	0.007(0.009)	0.016(0.011)[0.004]
$\max_i \lvert\hat{\alpha}_i - \alpha_i\rvert$	0.339	0.454	0.510	0.927[0.004]
$\max_j \lvert\hat{\beta}_j - \beta_j\rvert$	0.340	0.452	0.512	0.897[0.004]
$\lvert\hat{\alpha}_1 - \alpha_1\rvert$	0.077(0.095)	0.072(0.091)	0.070(0.089)	0.102(0.123)[0.004]
$\lvert\hat{\alpha}_{n/2} - \alpha_{n/2}\rvert$	0.074(0.095)	0.057(0.070)	0.059(0.074)	0.112(0.133)[0.004]
$\lvert\hat{\alpha}_{n-1} - \alpha_{n-1}\rvert$	0.077(0.095)	0.130(0.156)	0.142(0.180)	0.278(0.347)[0.004]
$\lvert\hat{\beta}_1 - \beta_1\rvert$	0.074(0.095)	0.073(0.091)	0.073(0.089)	0.097(0.123)[0.004]
$\lvert\hat{\beta}_{n/2} - \beta_{n/2}\rvert$	0.076(0.095)	0.057(0.070)	0.059(0.074)	0.110(0.133)[0.004]
$\lvert\hat{\beta}_{n-1} - \beta_{n-1}\rvert$	0.077(0.095)	0.126(0.156)	0.150(0.180)	0.273(0.342)[0.004]

setting most of the probabilities $p_{ij}(a, b)$ are nearly degenerate, close to 0 or 1. As a result, some estimated $\hat{p}_{ij}$ equal to 0 or 1 (i.e., the MLE does not exist) with a very high probability. These phenomena indicate that controlling the increasing rate of $\max_{1\le i\le 2n+2} |\boldsymbol{\varphi}_i|$ is crucial to guarantee the good asymptotic properties of the MLE.

By comparing Tables 1, 2 and 3, we have the following conclusions:

(1) The values of $|\hat{\theta}-\theta|$, $|\hat{\rho}-\rho|$, $\max_i |\hat{\alpha}_i-\alpha_i|$ and $\max_j |\hat{\beta}_j-\beta_j|$ decrease as n increases, and become larger as L_n increases. For example, when $L_n = 0$, $\theta = -2$ and $\rho = 0.5$, $|\hat{\theta}-\theta|$, $|\hat{\rho}-\rho|$ decrease from 0.075 and 0.095 to 0.007 and 0.009 respectively as n increases from 100 to 1,000. This is not difficult to understand since there are more samples as n increases when the range of parameters is fixed. On the other hand, the MLE is more variable as L_n becomes larger for fixed n. It can be expected that as n goes to infinity, the errors will decrease to zero when $\max_{1\le i\le 2n+2} |\boldsymbol{\varphi}_i|$ is controlled at an appropriate increasing rate. Similar phenomena could be observed for the square roots of diagonal elements of $\widehat{U}$.

(2) The accuracy of estimation in Table 1 is better than that in Tables 2 and 3. It may be due to that the data become sparse when θ is negative and deviate far away from 0.

(3) The values of $\max_i |\hat{\alpha}_i-\alpha_i|$ and $\max_j |\hat{\beta}_j-\beta_j|$ are much larger than those of single $|\hat{\alpha}_i - \alpha_i|$ and $|\hat{\beta}_j - \beta_j|$ and have a slower decreasing rate as n increases.

(4) For $1 \le i \le n$, $\hat{\alpha}_i$ and $\hat{\beta}_i$ have very similar errors, and $\widehat{U}_{2+i}$ and $\widehat{U}_{n+1+i}$ are almost equal. It shows that if α_i and β_i are equal, they may have similar asymptotic behaviors.

Table 3. The accuracy of estimation for $\theta = -\log\log(n)$ and $\rho = \log\log(n)$. The values in parentheses and square brackets are the estimated standard errors (the corresponding entries of the inverse of Fisher information matrix) and the frequencies that the MLE does not exist. If the square brackets are not shown, then the frequencies are zeros

L_n	0	$[\log(\log(n))]^{1/2}$	$\log(\log(n))$	$\log(n)$
		$n = 100$		
$\lvert\hat\theta - \theta\rvert$	0.043(0.034)	0.051(0.039)	0.055(0.042)	0.102(0.061)[0.240]
$\lvert\hat\rho - \rho\rvert$	0.067(0.072)	0.072(0.083)	0.078(0.092)	0.149(0.104)[0.240]
$\max_i \lvert\hat\alpha_i - \alpha_i\rvert$	0.702	0.879	0.922	1.389
$\max_j \lvert\hat\beta_j - \beta_j\rvert$	0.707	0.878	0.937	1.407
$\lvert\hat\alpha_1 - \alpha_1\rvert$	0.195(0.231)	0.201(0.239)	0.209(0.246)	0.296(0.362)[0.240]
$\lvert\hat\alpha_{n/2} - \alpha_{n/2}\rvert$	0.199(0.231)	0.199(0.234)	0.214(0.255)	0.398(0.426)[0.240]
$\lvert\hat\alpha_{n-1} - \alpha_{n-1}\rvert$	0.196(0.231)	0.306(0.362)	0.324(0.393)	0.523(0.658)[0.240]
$\lvert\hat\beta_1 - \beta_1\rvert$	0.199(0.231)	0.207(0.240)	0.209(0.246)	0.295(0.362)[0.240]
$\lvert\hat\beta_{n/2} - \beta_{n/2}\rvert$	0.193(0.231)	0.204(0.234)	0.218(0.255)	0.403(0.423)[0.240]
$\lvert\hat\beta_{n-1} - \beta_{n-1}\rvert$	0.194(0.231)	0.296(0.359)	0.327(0.394)	0.541(0.657)[0.240]
		$n = 500$		
$\lvert\hat\theta - \theta\rvert$	0.010(0.007)	0.012(0.008)	0.014(0.009)	0.092(0.014)
$\lvert\hat\rho - \rho\rvert$	0.014(0.015)	0.015(0.017)	0.018(0.022)	0.321(0.045)
$\max_i \lvert\hat\alpha_i - \alpha_i\rvert$	0.385	0.505	0.553	1.242
$\max_j \lvert\hat\beta_j - \beta_j\rvert$	0.384	0.499	0.559	1.238
$\lvert\hat\alpha_1 - \alpha_1\rvert$	0.093(0.109)	0.095(0.111)	0.102(0.116)	0.145(0.182)
$\lvert\hat\alpha_{n/2} - \alpha_{n/2}\rvert$	0.093(0.109)	0.089(0.104)	0.102(0.117)	0.640(0.265)
$\lvert\hat\alpha_{n-1} - \alpha_{n-1}\rvert$	0.091(0.109)	0.154(0.182)	0.173(0.207)	0.621(0.396)
$\lvert\hat\beta_1 - \beta_1\rvert$	0.094(0.109)	0.093(0.111)	0.101(0.116)	0.154(0.182)
$\lvert\hat\beta_{n/2} - \beta_{n/2}\rvert$	0.092(0.109)	0.091(0.104)	0.098(0.117)	0.711(0.263)
$\lvert\hat\beta_{n-1} - \beta_{n-1}\rvert$	0.094(0.109)	0.145(0.182)	0.166(0.207)	0.589(0.396)
		$n = 1{,}000$		
$\hat\theta - \theta$	0.006(0.003)	0.007(0.004)	0.008(0.005)	0.019(0.007)
$\hat\rho - \rho$	0.008(0.008)	0.008(0.009)	0.009(0.013)	0.031(0.012)
$\max_i \lvert\hat\alpha_i - \alpha_i\rvert$	0.296	0.390	0.445	0.752
$\max_j \lvert\hat\beta_j - \beta_j\rvert$	0.296	0.393	0.441	0.763
$\lvert\hat\alpha_1 - \alpha_1\rvert$	0.064(0.079)	0.069(0.080)	0.072(0.083)	0.121(0.137)
$\lvert\hat\alpha_{n/2} - \alpha_{n/2}\rvert$	0.066(0.079)	0.063(0.074)	0.073(0.084)	0.138(0.154)
$\lvert\hat\alpha_{n-1} - \alpha_{n-1}\rvert$	0.068(0.079)	0.116(0.135)	0.135(0.156)	0.261(0.298)
$\lvert\hat\beta_1 - \beta_1\rvert$	0.067(0.079)	0.068(0.080)	0.071(0.083)	0.116(0.137)
$\lvert\hat\beta_{n/2} - \beta_{n/2}\rvert$	0.068(0.079)	0.064(0.074)	0.073(0.084)	0.135(0.154)
$\lvert\hat\beta_{n-1} - \beta_{n-1}\rvert$	0.069(0.079)	0.112(0.135)	0.127(0.156)	0.249(0.298)

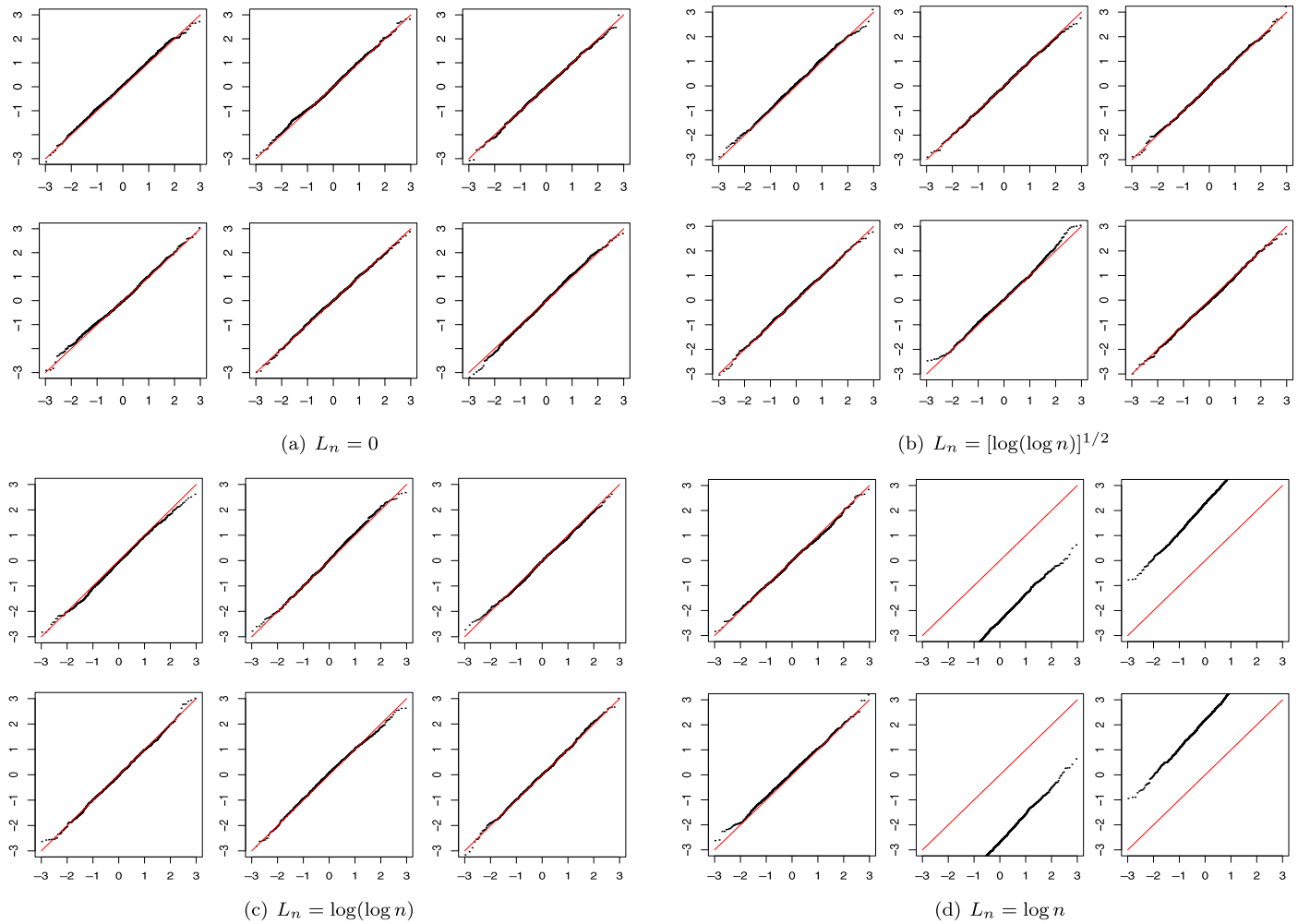
Due to limited space, the QQ-plots are shown here only for $n = 500$. By observing Figures 1–3, the quantiles of ξ_i and η_j coincide with the theoretical quantiles very well when $L_n \le \log(\log n)$, indicating that the MLE enjoys good asymptotic normality. There are similar phenomena for ξ_{ij} and η_{ij} when $L_n \le \log(\log n)$, whose figures are not shown. However, there are obvious deviations when $L_n = \log n$ in the middle and right subfigures of Figures 1–3 (d). For $i = j = n/2$, the QQ-plots of ξ_i and η_j deviate upward; for $i = j = n - 1$, they are in opposed directions. In Figure 4, only the middle subfigures have deviations corresponding to $\xi_{n/2,n/2+1}$ and $\eta_{n/2,n/2+1}$. It further confirms that asymptotic normality requires the control of the increasing rate of $\max_{1\le i\le 2n+2} \lvert\boldsymbol{\varphi}_i\rvert$.

Figure 5 shows that the sample quantiles of $(\hat\theta - \theta)/(\widehat{U}_{11})^{1/2}$ coincide with the theoretical quantiles only when $\max_i \lvert\alpha_i\rvert = \max_i \lvert\beta_i\rvert = 0$, $[\log(\log n)]^{1/2}$ and $\theta = 0$. For other cases, there are obvious deviations. From Figure 6, the QQ-plots of $(\hat\rho - \rho)/(\widehat{U}_{22})^{1/2}$ coincide with the diagonal lines only when $\max\{\theta, \rho, \max_i \lvert\alpha_i\rvert, \max_i \lvert\beta_i\rvert\} \le [\log(\log n)]^{1/2}$. For other subfigures, there are more or less deviations. Especially when $\max_i \lvert\alpha_i\rvert = \max_i \lvert\beta_i\rvert = \log n$, the deviation is very large.

The above phenomena were also observed in the figures of the QQ-plots when $n = 100$ or 1,000 (results not shown).

3. DISCUSSION

We have presented a simulated study about the large-sample properties of the p_1 model. Although we only presented the results for a linear sequence for the values of $\alpha_1, \ldots, \alpha_n$ and $\beta_1, \ldots, \beta_n$, similar phenomena were observed



(a) $L_n = 0$ (b) $L_n = [\log(\log n)]^{1/2}$

(c) $L_n = \log(\log n)$ (d) $L_n = \log n$

Figure 1. The QQ-plots for $\theta = 0$ and $\rho = 0$ ($n = 500$). The red color is the diagonal line ($y = x$). For each subfigure (a)–(d), the plots of the left, middle and right columns are ξ_1, $\xi_{n/2}$, ξ_{n-1} in the first row; the second row corresponds to η_1, $\eta_{n/2}$, η_{n-1}, respectively. The horizontal and vertical axises are the theoretical and sample quantiles, respectively.

for other values of parameters. The aim of the simulation is not to cover all possible combinations of these parameters, but to obtain some intuitive understandings on the asymptotic behaviors of the MLE by considering selected values for the parameters. The results present strong evidences for consistency and asymptotic normality of the maximum likelihood estimate in the p_1 model as the number of graphical vertices goes to infinity, when the increasing rate of $\max_i |\boldsymbol{\phi}_i|$ is controlled at some level. We make the following conjectures.

Conjecture 1 (Consistency)**.** *Assume that the density parameter θ and the reciprocity parameter ρ are constants. Let $L_n = \max\{\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n\}$, and $\hat{\theta}$, $\hat{\rho}$, $\hat{\alpha}_1, \ldots, \hat{\alpha}_n$, $\hat{\beta}_1, \ldots, \hat{\beta}_n$ be the MLEs of $\theta, \rho, \alpha_1, \ldots, \alpha_t, \beta_1, \ldots, \beta_n$. If $L_n = o(\log(\log n))$, then as n goes to infinity, the MLE exists with probability approaching one and satisfies*

$$|\hat{\theta} - \theta| = o_p(1), \quad |\hat{\rho} - \rho| = o_p(1),$$
$$\max_{i=1,\ldots,n} |\hat{\alpha}_i - \alpha_i| = o_p(1), \quad \max_{i=1,\ldots,n} |\hat{\beta}_j - \beta_j| = o_p(1)$$

Conjecture 2 (Central limit theorem)**.** *Let $\boldsymbol{\phi} = (\theta, \rho, \alpha_1, \ldots, \alpha_{n-1}, \beta_1, \ldots, \beta_{n-1})'$ and V be the Fisher information matrix of $\boldsymbol{\phi}$ whose entries are given in the Appendix. Denote $U = V^{-1}$. If $L_n = o(\log(\log n))$, then as n goes to infinity, for any fixed r,*

$$\frac{\hat{\rho} - \rho}{\sqrt{U_{22}}} \xrightarrow{D} N(0,1),$$
$$(\hat{\alpha}_1 - \alpha_1, \ldots, \hat{\alpha}_r - \alpha_r)^T \xrightarrow{D} N(0, U_{i,j=3,\ldots,2+r}),$$
$$(\hat{\beta}_1 - \beta_1, \ldots, \hat{\beta}_r - \beta_r)^T \xrightarrow{D} N(0, U_{i,j=2+n,\ldots,2+n+r}),$$

where $\xrightarrow{D}$ denotes convergence in distribution.

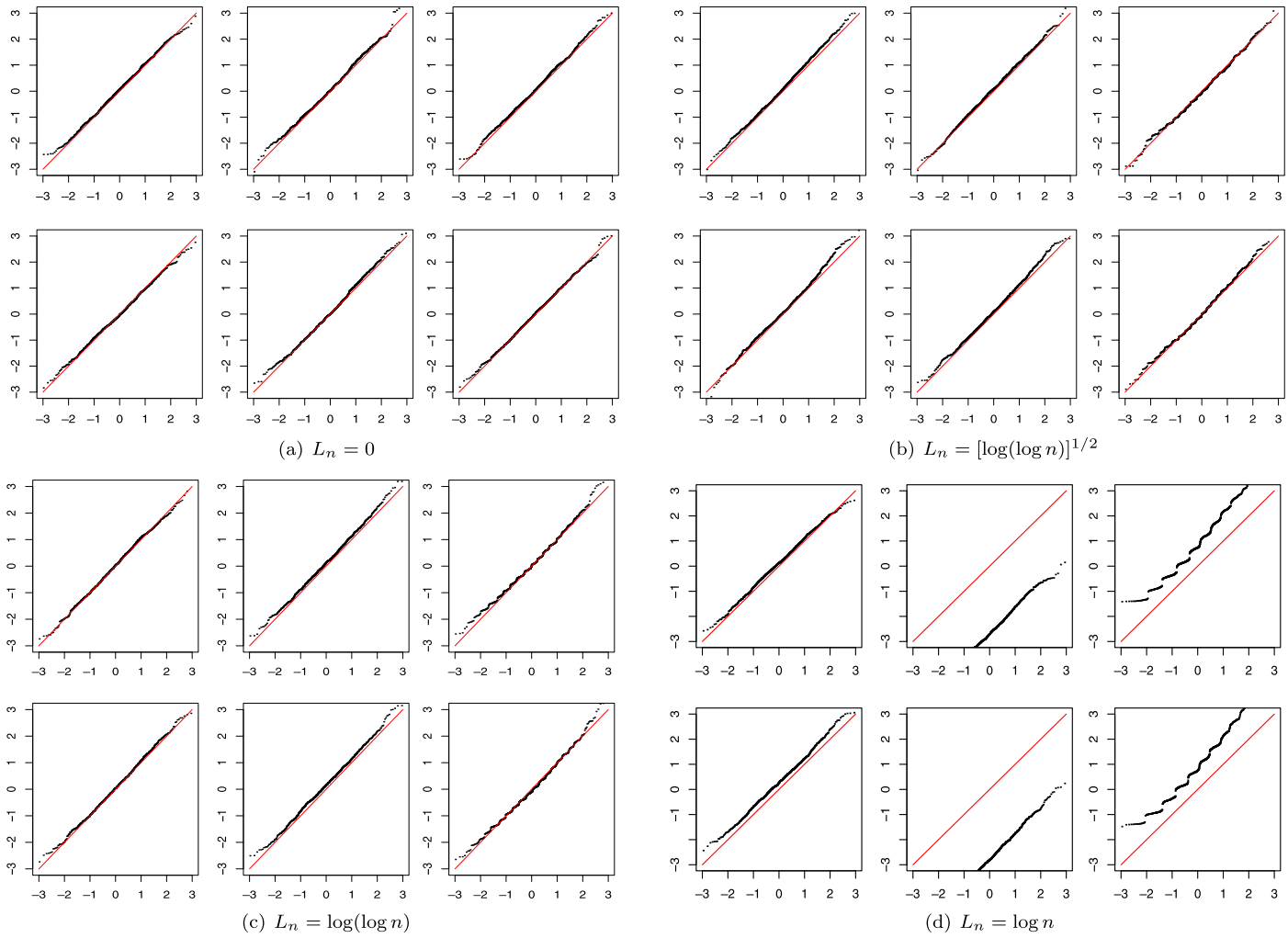


Figure 2. The QQ-plots for $\theta = -2$ and $\rho = 0.5$ ($n = 500$). The red color is the diagonal line ($y = x$). For each subfigure (a)–(d), the plots of the left, middle and right columns are ξ_1, $\xi_{n/2}$, ξ_{n-1} in the first row; the second row corresponds to η_1, $\eta_{n/2}$, η_{n-1}, respectively. The horizontal and vertical axises are the theoretical and sample quantiles, respectively.

The condition $L_n = o(\log(\log n))$ is motivated by a similar one in Yan and Xu (2013). For the above conjectures, another motivation is that the large-sample theories for some simpler network models (e.g., the β-model, the Bradley-Terry model, the Rasch model) have been established. We point out, however, that a different phenomenon is revealed. More specifically, $(\hat{\theta} - \theta)/\sqrt{U_{11}}$ does not follow asymptotically a standard normal distribution according to the simulations. The sufficient statistics of θ, α_i and β_j are $\sum_{i,j} X_{ij}$,$\sum_k X_{ik}$ and $\sum_k X_{kj}$, respectively. Clearly, the sufficient statistic of θ is a linear sum of $\sum_k X_{ik}$ tied to α_i (also of $\sum_k X_{kj}$ tied to β_j). In this sense, the asymptotic distribution of $\hat{\theta}$ depends on all of the estimated parameters $\hat{\alpha}_i$, $\hat{\beta}_i$, $1 \le i \le n$. Therefore, the asymptotic behavior of $\hat{\theta}$ may be different from other estimated parameters.

The independence assumption in the p_1 model implies that it cannot represent tendencies toward transitivity, cliquing, hierarchy, stars and so on. To deal with complicated dependent structures, other exponential random graph models have been proposed. Frank and Strauss (1986) introduced the Markov random graphs to tie sufficient statistics by counts of various triangles and stars. Wasserman and Pattison (1996) described a more general p^* model that can accommodate clustering and centralization characteristics in social networks. Hunter (2007) discussed curved exponential family models. For recent reviews, see Wasserman and Robins (2005) and Robins et al. (2007a, b). Like the p_1 model, little asymptotic properties are known for these general ERGMs. Recently, Shalizi and Rinaldo (2012) proved the consistency of maximum likelihood estimation under the projected ERGMs in which pa-

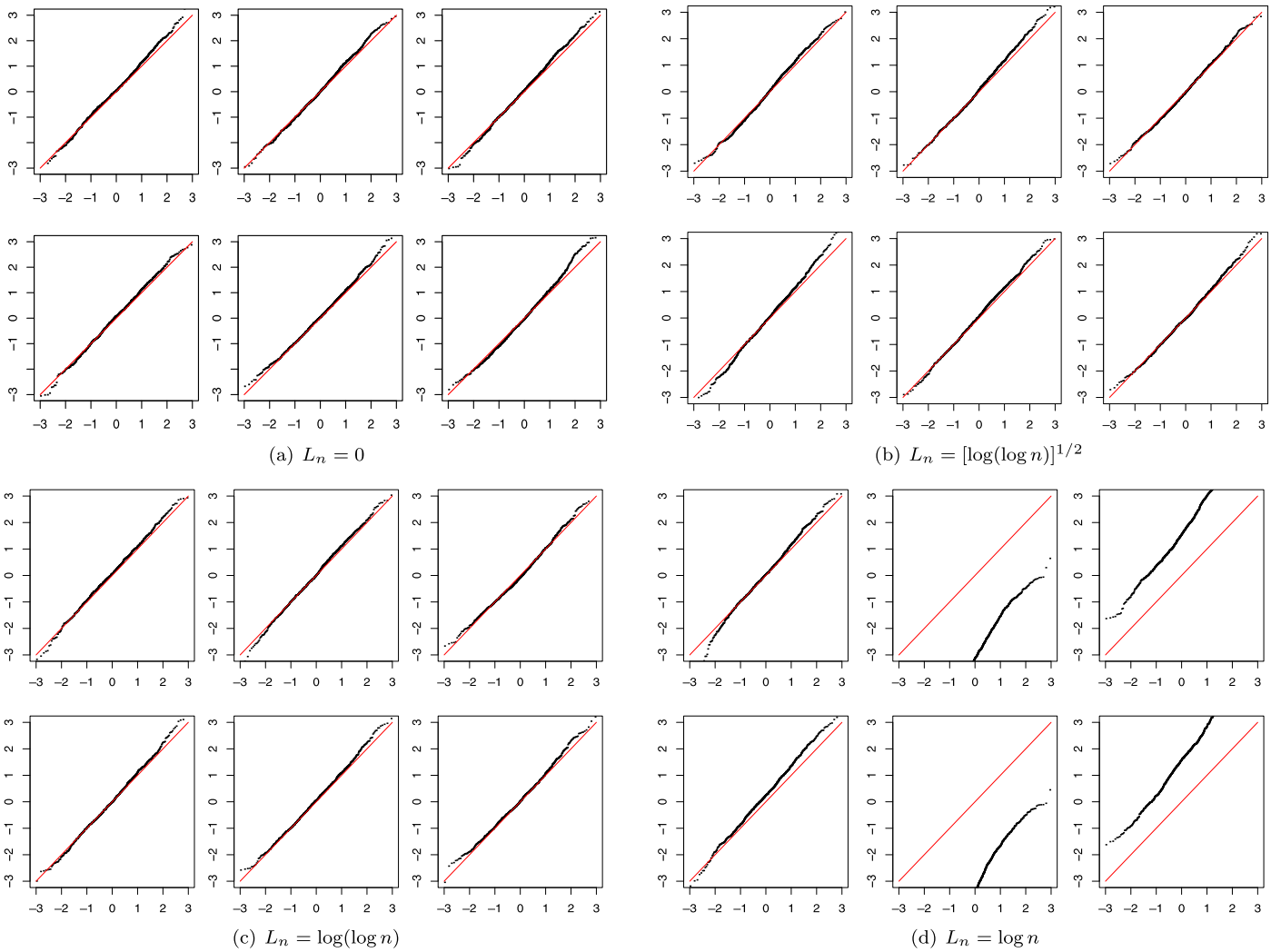


Figure 3. The QQ-plots for $\theta = -\log(\log n)$ and $\rho = \log(\log n)$ ($n = 500$). The red color is the diagonal line ($y = x$). For each subfigure (a)–(d), the plots of the left, middle and right columns are ξ_1, $\xi_{n/2}$, ξ_{n-1} in the first row; the second row corresponds to η_1, $\eta_{n/2}$, η_{n-1}, respectively. The horizontal and vertical axises are the theoretical and sample quantiles, respectively.

rameters derived from the observed sub-network could be consistently applied to the whole network. We hope that our simulations also shed light on the large-sample properties of maximum likelihood estimation for these general ERGMs.

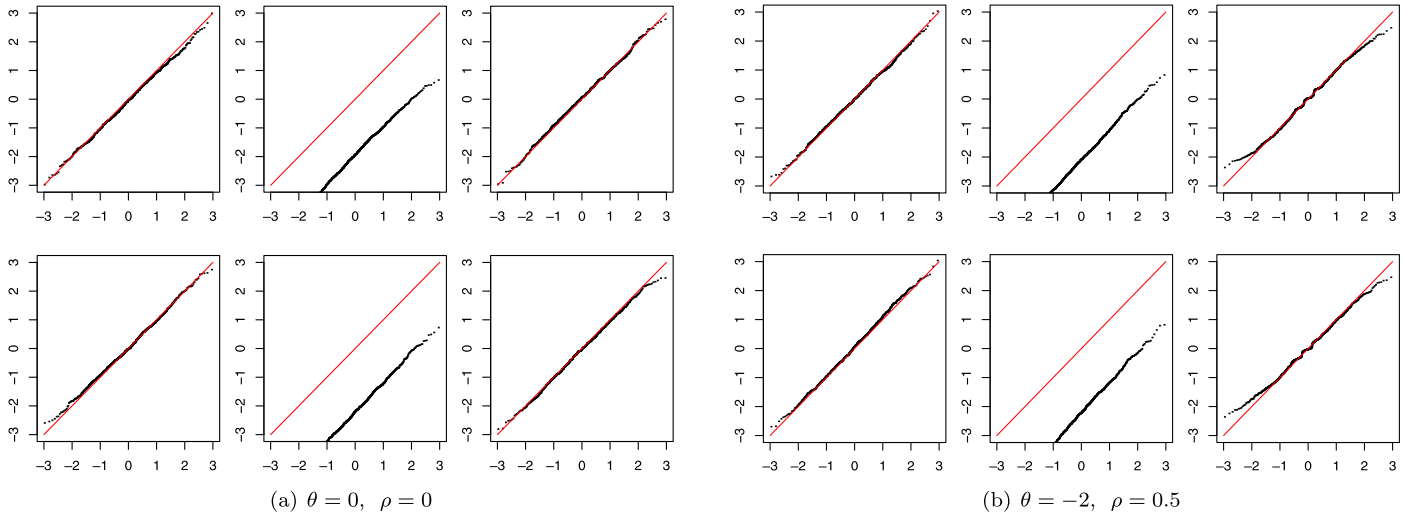
ACKNOWLEDGEMENTS

The authors would like to thank two anonymous referees for their valuable comments. R code for the simulations is available upon request by writing to Ting Yan (tingyanty@mail.ccnu.edu.cn). Leng is also affiliated with Department of Statistics and Applied Probability, National University of Singapore.

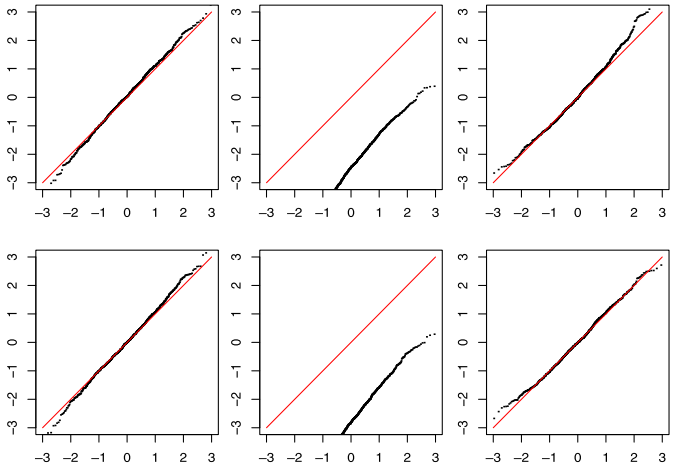
APPENDIX

The log-likelihood function of the p_1 model is easily seen as

$$\ell(\boldsymbol{\phi}) = \rho m + \theta x_{++} + \sum_{i=1}^{n} \alpha_i x_{i+} + \sum_{j=1}^{n} \beta_j x_{+j} - \frac{1}{2} \sum_{i,j=1; i \neq j}^{n} \log k_{ij} := \ell_1 - \frac{1}{2}\ell_2 - \ell_3,$$



(a) $\theta = 0,\ \rho = 0$ (b) $\theta = -2,\ \rho = 0.5$



(c) $\theta = -\log(\log n)$ and $\rho = \log(\log n)$

Figure 4. The QQ-plots for ξ_{ij} and η_{ij} when $L_n = \log n$. The red color is the diagonal line ($y = x$). For each subfigure (a)–(c), the plots of the left, middle and right columns are ξ_{12}, $\xi_{n/2,n/2+1}$, $\xi_{n-2,n-1}$ in the first row; the second row corresponds to η_{12}, $\eta_{n/2,n/2+1}$, $\eta_{n-2,n-1}$, respectively. The horizontal and vertical axises are the theoretical and sample quantiles, respectively.

where

$$\ell_1 = \rho m + \theta x_{++} + \sum_{i=1}^{n} \alpha_i x_{i+} + \sum_{j=1}^{n} \beta_j x_{+j},$$

$$\ell_2 = \sum_{h,g=1;h\neq g}^{n-1} \log k_{hg},$$

$$\ell_3 = \sum_{g=1}^{n-1} \log k_{ng}.$$

The Fisher information matrix V^* of the parameter vector $\boldsymbol{\phi}^*$ is

$$V = -E\left[\frac{\partial^2 \ell(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}^T \partial \boldsymbol{\phi}}\right] = \frac{1}{2}\frac{\partial^2 \ell_2(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}^T \partial \boldsymbol{\phi}} + \frac{\partial^2 \ell_3(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}^T \partial \boldsymbol{\phi}}.$$

For convenience, denote $a_{ij} = \exp(\theta+\alpha_i+\beta_j)$, $b_{ij} = \exp(\theta+\beta_i+\alpha_j)$ and $c_{ij} = \exp(\rho+2\theta+\alpha_i+\alpha_j+\beta_i+\beta_j)$ for all pairs $(i,j), i \neq j$ and define $a_{ii} = 0$, $b_{ii} = 0$ and $c_{ii} = 0$. Note that the transpose of A is B and that the matrix C is symmetric, where $A = (a_{ij})_{i,j=1,\ldots,n}$, $B = (b_{ij})_{i,j=1,\ldots,n}$ and $C = (c_{ij})_{i,j=1,\ldots,n}$. By direct calculations, the entries of $\frac{\partial^2 \ell_2(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}^T \partial \boldsymbol{\phi}}$ and $\frac{\partial^2 \ell_3(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}^T \partial \boldsymbol{\phi}}$ are given by

$$\frac{\partial^2 \ell_2}{\partial \rho^2} = \sum_{i,j=1;i\neq j}^{n-1} \left(\frac{c_{ij}}{k_{ij}} - \frac{c_{ij}^2}{k_{ij}^2}\right),$$

$$\frac{\partial^2 \ell_2}{\partial \theta \partial \rho} = \sum_{i,j=1;i\neq j}^{n-1} \left[\frac{2c_{ij}}{k_{ij}} - \frac{c_{ij}(a_{ij}+b_{ij}+2c_{ij})}{k_{ij}^2}\right],$$

$$\frac{\partial^2 \ell_2}{\partial \alpha_i \partial \rho} = 2\sum_{g=1;g\neq i}^{n-1} \left[\frac{c_{ig}}{k_{ig}} - \frac{c_{ig}(a_{ig}+c_{ig})}{k_{ig}^2}\right],$$

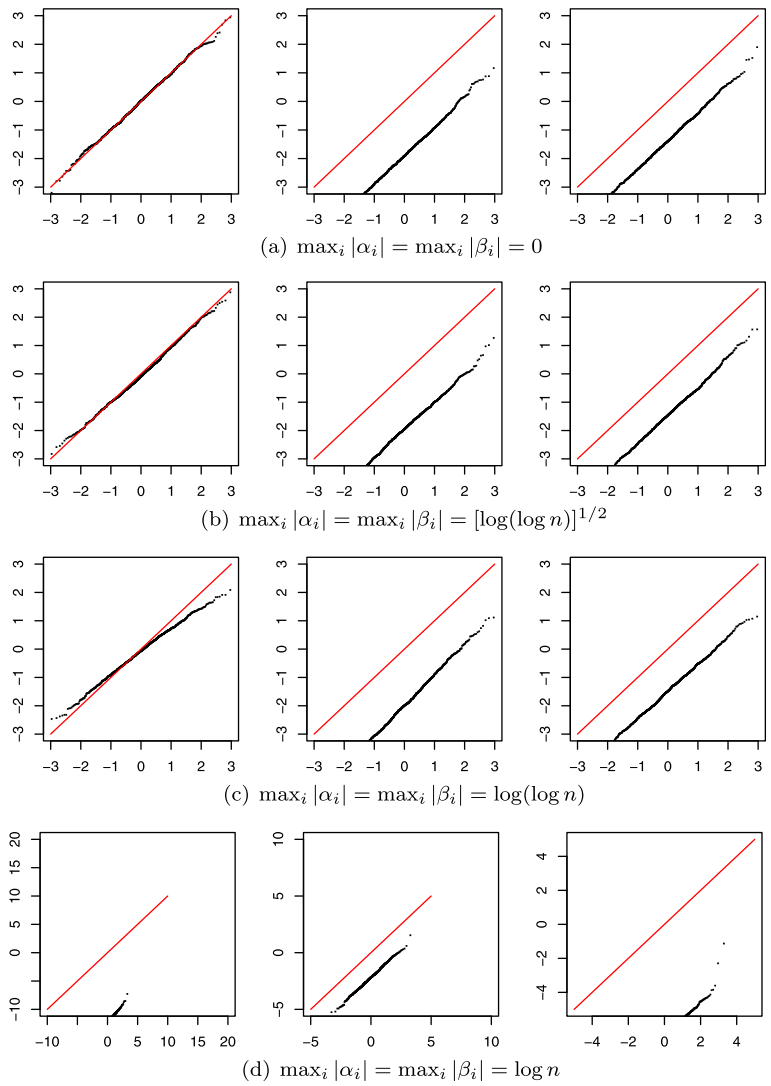


Figure 5. The QQ-plots of $(\hat{\theta} - \theta)/(\widehat{U}_{11})^{1/2}$ ($n = 500$). The red color is the diagonal line ($y = x$). For each subfigure (a)–(d), the plots of the left, middle and right columns correspond to $(\theta, \rho) = (0, 0)$, $(-2, 0.5)$ and $(-\log(\log n), \log(\log n))$, respectively. The horizontal and vertical axises are the theoretical and sample quantiles, respectively.

$$\frac{\partial^2 \ell_2}{\partial \beta_i \partial \rho} = 2 \sum_{g=1;g\neq i}^{n-1} \left[\frac{c_{ig}}{k_{ig}} - \frac{c_{ig}(b_{ig} + c_{ig})}{k_{ig}^2} \right],$$

$$\frac{\partial^2 \ell_2}{\partial \theta^2} = \sum_{i,j=1;i\neq j}^{n-1} \left[\frac{a_{ij} + b_{ij} + 4c_{ij}}{k_{ij}} - \frac{(a_{ij} + b_{ij} + 2c_{ij})^2}{k_{ij}^2} \right],$$

$$\frac{\partial^2 \ell_2}{\partial \alpha_i \partial \theta} = 2 \sum_{g=1,g\neq i}^{n-1} \frac{(a_{ig} + c_{ig})}{k_{ig}} \left[1 - \frac{(a_{ig} + b_{ig} + c_{ig})}{k_{ig}} \right],$$

$$\frac{\partial^2 \ell_2}{\partial \beta_i \partial \theta} = 2 \sum_{g=1,g\neq i}^{n-1} \frac{(b_{ig} + c_{ig})}{k_{ig}} \left[1 - \frac{(a_{ig} + b_{ig} + c_{ig})}{k_{ig}} \right],$$

$$\frac{\partial^2 \ell_2}{\partial \alpha_i^2} = 2 \sum_{g=1;g\neq i}^{n-1} \left[\frac{a_{ig} + c_{ig}}{k_{ig}} - \frac{(a_{ig} + c_{ig})^2}{k_{ig}^2} \right],$$

$$\frac{\partial^2 \ell_2}{\partial \alpha_i \partial \alpha_j} = 2 \left(\frac{c_{ij}}{k_{ij}} - \frac{c_{ij}^2}{k_{ij}^2} \right),$$

$$\frac{\partial^2 \ell_2}{\partial \alpha_i \partial \beta_j} = \frac{a_{ij} + c_{ij}}{k_{ij}} - \frac{(a_{ij} + c_{ij})^2}{k_{ij}^2} + \frac{b_{ji} + c_{ij}}{k_{ij}} - \frac{(b_{ji} + c_{ij})^2}{k_{ij}^2},$$

$$\frac{\partial^2 \ell_2}{\partial \beta_i^2} = 2 \sum_{g=1;g\neq i}^{n-1} \left[\frac{b_{ig} + c_{ig}}{k_{ig}} - \frac{(b_{ig} + c_{ig})^2}{k_{ig}^2} \right],$$

$$\frac{\partial^2 \ell_2}{\partial \beta_i \partial \beta_j} = 2 \left(\frac{c_{ij}}{k_{ij}} - \frac{c_{ij}^2}{k_{ij}^2} \right),$$

$$\frac{\partial^2 \ell_3}{\partial \rho^2} = \sum_{g=1}^{n-1} \left[\frac{c_{ng}}{k_{ng}} - \frac{c_{ng}^2}{k_{ng}^2} \right],$$

$$\frac{\partial^2 \ell_3}{\partial \theta \partial \rho} = \sum_{g=1}^{n-1} \left[\frac{2c_{ng}}{k_{ng}} - \frac{c_{ng}(a_{ng} + b_{ng} + 2c_{ng})}{k_{ng}^2} \right],$$

$$\frac{\partial^2 \ell_3}{\partial \alpha_i \partial \rho} = \sum_{g=1}^{n-1} \left[\frac{c_{ng}(a_{ng} + c_{ng})}{k_{ng}^2} - \frac{c_{ng}}{k_{ng}} \right] + \frac{c_{ni}}{k_{ni}} - \frac{c_{ni}(b_{ni} + c_{ni})}{k_{ni}^2},$$

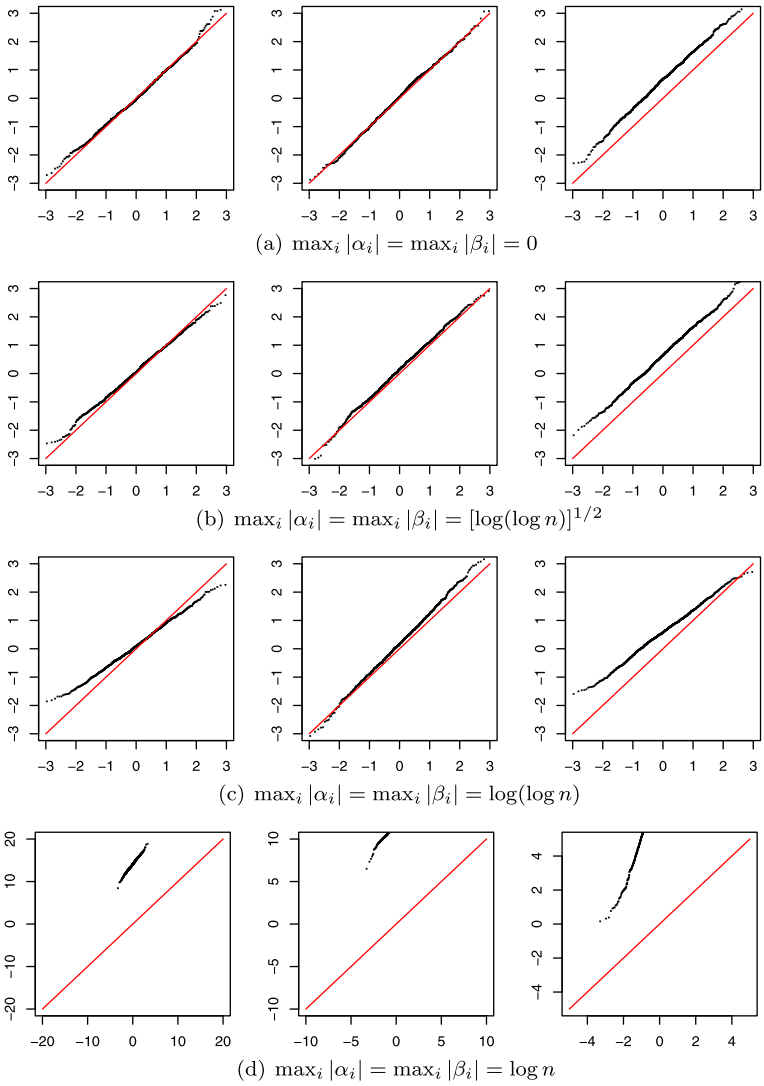


Figure 6. The QQ-plots of $(\hat{\rho} - \rho)/(\widehat{U}_{22})^{1/2}$ ($n = 500$). The red color is the diagonal line ($y = x$). For each subfigure (a)–(d), the plots of the left, middle and right columns correspond to $(\theta, \rho) = (0, 0)$, $(-2, 0.5)$ and $(-\log(\log n), \log(\log n))$, respectively. The horizontal and vertical axises are the theoretical and sample quantiles, respectively.

$$\frac{\partial^2 \ell_3}{\partial \beta_i \partial \rho} = \sum_{g=1}^{n-1} \left[\frac{c_{ng}(b_{ng} + c_{ng})}{k_{ng}^2} - \frac{c_{ng}}{k_{ng}} \right] + \frac{c_{ni}}{k_{ni}} - \frac{c_{ni}(a_{ni} + c_{ni})}{k_{ni}^2},$$

$$\frac{\partial^2 \ell_3}{\partial \theta^2} = \sum_{g=1}^{n-1} \left[\frac{a_{ng} + b_{ng} + 4c_{ng}}{k_{ng}} - \frac{(a_{ng} + b_{ng} + 2c_{ng})^2}{k_{ng}^2} \right],$$

$$\frac{\partial^2 \ell_3}{\partial \alpha_i \partial \theta} = \sum_{g=1}^{n-1} \left[\frac{(a_{ng} + b_{ng} + 2c_{ng})(a_{ng} + c_{ng})}{k_{ng}^2} - \frac{a_{ng} + 2c_{ng}}{k_{ng}} \right] + \frac{b_{ni} + 2c_{ni}}{k_{ni}} - \frac{(a_{ni} + b_{ni} + 2c_{ni})(b_{ni} + c_{ni})}{k_{ni}^2},$$

$$\frac{\partial^2 \ell_3}{\partial \beta_i \partial \theta} = \sum_{g=1}^{n-1} \left[\frac{(a_{ng} + b_{ng} + 2c_{ng})(b_{ng} + c_{ng})}{k_{ng}^2} - \frac{b_{ng} + 2c_{ng}}{k_{ng}} \right] + \frac{a_{ni} + 2c_{ni}}{k_{ni}} - \frac{(a_{ni} + b_{ni} + 2c_{ni})(a_{ni} + c_{ni})}{k_{ni}^2},$$

$$\frac{\partial^2 \ell_3}{\partial \alpha_i^2} = \sum_{g=1}^{n-1} \left[\frac{[a_{ng} + b_{ng} 1_{\{g=i\}} + c_{ng}(1 - 1_{\{g=i\}})]}{k_{ng}} - \frac{[a_{ng} - b_{ng} 1_{\{g=i\}} + c_{ng}(1 - 1_{\{g=i\}})]^2}{k_{ng}^2} \right],$$

$$\frac{\partial^2 \ell_3}{\partial \alpha_j \partial \alpha_i} = \sum_{g=1}^{n-1} \left[\frac{a_{ng} + c_{ng}}{k_{ng}} - \frac{(a_{ng} + c_{ng})^2}{k_{ng}^2} \right] + \frac{(b_{ni} + c_{ni})(a_{ni} + c_{ni})}{k_{ni}^2} + \frac{(b_{nj} + c_{nj})(a_{nj} + c_{nj})}{k_{nj}^2} - \frac{c_{ni}}{k_{ni}} - \frac{c_{nj}}{k_{nj}},$$

$$\frac{\partial^2 \ell_3}{\partial \beta_j \partial \alpha_i} = \sum_{g=1}^{n-1} \left[\frac{c_{ng}}{k_{ng}} - \frac{(a_{ng} + c_{ng})(b_{ng} + c_{ng})}{k_{ng}^2} \right]$$

$$+\frac{(b_{ni}+c_{ni})^2}{k_{ni}^2}+\frac{(a_{nj}+c_{nj})^2}{k_{nj}^2}-\frac{a_{nj}}{k_{nj}}-\frac{b_{ni}}{k_{ni}}-\frac{c_{ni}}{k_{ni}}-\frac{c_{nj}}{k_{nj}},$$

$$\frac{\partial^2\ell_3}{\partial\beta_i^2}=\sum_{g=1}^{n-1}\left[\frac{[b_{ng}+a_{ng}1_{\{g=i\}}+c_{ng}(1-1_{\{g=i\}})]}{k_{ng}}-\frac{[b_{ng}-a_{ng}1_{\{g=i\}}+c_{ng}(1-1_{\{g=i\}})]^2}{k_{ng}^2}\right],$$

$$\frac{\partial^2\ell_3}{\partial\beta_j\partial\beta_i}=\sum_{g=1}^{n-1}\left[\frac{b_{ng}+c_{ng}}{k_{ng}}-\frac{(b_{ng}+c_{ng})^2}{k_{ng}^2}\right]+\frac{(a_{ni}+c_{ni})(b_{ni}+c_{ni})}{k_{ni}^2}+\frac{(a_{nj}+c_{nj})(b_{nj}+c_{nj})}{k_{nj}^2}-\frac{c_{ni}}{k_{ni}}-\frac{c_{nj}}{k_{nj}}.$$

Received 30 April 2013

REFERENCES

Blitzstein, J. and Diaconis, P. (2010). A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Mathematics* **6** 489–522. MR2809836

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika* **39** 324–345. MR0070925

Chatterjee, S., Diaconis, P. and Sly, A. (2011). Random graphs with a given degree sequence. *Annals of Applied Probability* **21** 1400–1435. MR2857452

Erdős P. and Rényi A. (1959). On random graphs. I. *Publicationes Mathematicae* **6** 290–97. MR0120167

Frank O. (1981). An exponential family of probability distributions for directed graphs: Comment. *Journal of the American Statistical Association* **76** 58–59. MR0608176

Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association* **81** 832–842. MR0860518

Fienberg, S. E., Petrović, S. and Rinaldo, A. (2011). Algebraic statistics for random graph models: Markov bases and their uses. In: *Looking Back: Proceedings of a Conference in Honor of Paul W. Holland*, Dorans, N. J. and Sinharay, S., eds, Springer, New York, pp. 21–38. MR2856692

Haberman, S. J. (1977). Maximum likelihood estimates in exponential response models. *The Annals of Statistics* **5** 815–841. MR0501540

Haberman S. J. (1981). An exponential family of probability distributions for directed graphs: Comment. *Journal of the American Statistical Association* **76** 60–61. MR0608176

Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs (with discussion). *Journal of the American Statistical Association* **76** 33–65. MR0608176

Hunter D. R. (2007). Curved exponential family models for social networks. *Social Networks* **29** 216–230.

Leinhardt, S. (1977). *Social Networks: A Developing Paradigm*. Academic Press, New York.

Petrović, S., Rinaldo, A. and Fienberg, S. E. (2010). Algebraic statistics for a directed random graph model with reciprocation. In: *Algebraic Methods in Statistics and Probability II*, Viana, M. and Wynn, H., eds, Contemporary Mathematics, Vol. 516, American Mathematical Society, Providence, RI. MR2730754

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Paedagogiske Institut, Copenhagen.

Rinaldo, A., Petrović, S. and Fienberg, S. E. (2010). On the existence of the mle for a directed random graph network model with reciprocation. Available at http://arxiv.org/abs/1010.0745. MR2730754

Rinaldo, A., Petrović, S. and Fienberg, S. E. (2013). Maximum likelihood estimation in the beta model. *The Annals of Statistics* **41** 1085–1110. MR3113804

Robins, G. L., Pattison, P. E., Kalish, Y., and Lusher, D. (2007a). An introduction to exponential random graph (p^*) models for social networks. *Social Networks* **29** 173–191.

Robins, G. L., Snijders, T. A. B., Weng, P., Handcock, M. S., and Pattison, P. E. (2007b). Recent developments in exponential random graph (p^*) models for social networks. *Social Networks* **29** 192–215.

Shalizi, C. R. and Rinaldo, A. (2013). Consistency under sampling of exponential random graph models. *The Annals of Statistics* **41** 508–535. MR3099112

Simons, G. and Yao, Y.-C. (1999). Asymptotics when the number of parameters tends to infinity in the Bradley-Terry model for paired comparisons. *The Annals of Statistics* **27** 1041–1060. MR1724040

Wang, Y. J. and Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association* **82** 8–19. MR0883333

Wasserman, S. and Pattison, P. E. (1996). Logit models and logistic regressions for social networks. I. An introduction to Markov graphs and p^*. *Psychometrika* **61** 401–425. MR1424909

Wasserman, S. and Robins, G. L. (2005). An introduction to random graphs, dependence graphs, and p^*. In: *Models and Methods in Social Network Analysis*, Carrington, P., Scott, J., and Wasserman, S., eds, Cambridge University Press, New York, pp. 148–161.

Yan, T. and Xu, J. (2013). A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices. *Biometrika* **100** 519–524. MR3068452

Ting Yan
Department of Statistics
Central China Normal University
Wuhan 430079
China
E-mail address: tingyanty@mail.ccnu.edu.cn

Chenlei Leng
Department of Statistics
University of Warwick
CV4 7AL
United Kingdom
E-mail address: C.Leng@warwick.ac.uk