Research Report

Department of Statistics

Maximum Likelihood Estimation for Markov Graphs

by

Jukka Corander
Karin Dahmström
Per Dahmström

1998:8

Stockholm University, S-106 91 Stockholm, Sweden

Maximum Likelihood Estimation for Markov Graphs

by

Jukka Corander
Karin Dahmström
Per Dahmström

1998:8

Stockholm University
Department of Statistics

ISBN 91-7474-190-X
ISSN 0280-7564

October 1998

Maximum Likelihood Estimation for Markov Graphs

Jukka Corander, Karin Dahmström and Per Dahmström
Department of Statistics, Stockholm University, S-106 91 Stockholm

Abstract

Maximum likelihood estimation of Markov graph parameters is considered. An iterative technique using an expansion of the expected values of sufficient statistics in a Markov graph in terms of cumulants is introduced. Efficient starting values for parameter estimates are shown to be obtainable from the cumulants of graph statistics in uniform random graphs. A Markov chain Monte Carlo method is used to generate samples of Markov graphs with fixed parameter values at successive iteration steps. Complete enumeration is used in a small graph to show that the iterative estimation technique performs satisfactorily compared to the exact maximum likelihood estimates. Properties of the maximum likelihood estimators and pseudolikelihood estimators, suggested earlier in the statistical literature, are investigated in small and large graphs. Our results for an undirected graph with a univariate sufficient statistic suggest that in graphs with up to approximately 40 vertices the maximum likelihood estimator is uniformly better over such parts of the parameter space where the model behavior is not degenerate. In larger graphs with 40-100 nodes, the two estimators seem to be nearly equivalent. However, the pseudolikelihood estimators are shown to vary with different graphs with the same value of the sufficient statistic.
Key words: Cumulants, Markov Chain Monte Carlo, Markov Graph, Maximum Likelihood Estimation, Pseudolikelihood Estimation.

1. Introduction

In studies of social networks elaborate models allowing dependence between dyads have gained popularity during the recent years. An important class of such models, called Markov graphs, was introduced by Frank and Strauss (1986). However, maximum likelihood estimation of Markov graph parameters has been considered computationally too burdensome for routine data analysis. The only likelihood based method suggested in the literature is restricted to Markov graphs with a univariate sufficient statistic, see Frank and Strauss (1986), and Strauss (1986).

To enable estimation of Markov graph parameters in general, an approximative method based on pseudolikelihood has been developed by Frank and Strauss (1986). Applications are also given in Strauss and Ikeda (1990), Frank (1991), Frank and Nowicki (1993), and Wasserman and Pattison (1996). With this method parameter estimates are obtainable via logistic regression models, for which estimation procedures are available in standard statistical packages. No large-scale simulation studies of the performance of pseudolikelihood estimators seem to have been considered in the statistical literature. However, some limited comparisons are given in Frank and Strauss (1986) and in Strauss and Ikeda (1990) for Markov graphs with a univariate sufficient statistic. These results suggest that the pseudolikelihood estimators are satisfactory substitutes for the maximum likelihood estimators.

In the present paper, which is a revised and extended version of Dahmström and Dahmström (1993), we develop an iterative technique for maximum likelihood estimation of Markov graph parameters. Our method is based on an expansion of the expected values of the sufficient statistics in a Markov graph in terms of cumulants. For general Markov graphs, we use a linear approximation to the cumulant generating function of the sufficient statistics. Efficient starting values for the parameter estimates are shown to be obtainable from low-order cumulants of the graph statistics in uniform graphs, and algebraic expressions for such cumulants are given for some special cases. A Markov chain Monte Carlo technique is used to obtain samples of Markov graphs at successive iteration steps, and a highly efficient algorithm for generating such samples is described.

To investigate the performance of the iterative estimation we use complete enumeration to obtain the exact maximum likelihood estimates in a small graph, and for comparison, the pseudolikelihood estimates are also given. Complete enumeration and simulation are used to investigate the statistical properties of the maximum likelihood and pseudolikelihood estimators in small and large graphs, respectively. The variability of pseudolikelihood estimators with different graphs with the same value of the sufficient statistic is illustrated, and some problems associated with the interpretation of model parameters are discussed.

Necessary graph terminology and the definition of a Markov graph are reviewed in the next section. In Section 3 an estimation technique for Markov graphs with a univariate sufficient statistic is considered, and Section 4 gives a generalization to graphs with multivariate sufficient statistics. Comparison of the exact, iterative and pseudolikelihood estimates using complete enumeration is presented for some models in the section thereafter. Specific properties of the pseudolikelihood estimation are discussed in Section 6, and performance of the maximum likelihood and pseudolikelihood estimators is considered in Section 7. Some final remarks are given in Section 8.

2. Preliminaries

Let N denote a finite set of n elements. The set of all ordered pairs of distinct elements from N is $N^{(2)}$, and the set of all unordered pairs is $\binom{N}{2}$. For the cardinality of $\binom{N}{2}$ we write m. A graph G on N is a subset of pairs of elements in N. Directed and undirected graphs can be given as $G \subseteq N^{(2)}$ and $G \subseteq \binom{N}{2}$, respectively. The elements of $N, N^{(2)}$ and $\binom{N}{2}$ are called vertices, arcs and edges, respectively. An alternative representation of a graph is an adjacency matrix $\mathbf{Y}$, where

$$Y_{uv} = \begin{cases} 1, & \text{if } \{u,v\} \in G \\ 0, & \text{otherwise} \end{cases} \tag{2.1}$$

for $u, v \in N$. We shall solely consider cases where the *size* r of G is fixed, that is all models in the sequel are conditioned on r. There are $\binom{n(n-1)}{r}$ and $\binom{m}{r}$ distinct graphs of size r in the directed and undirected cases, respectively. It is further assumed here that the probability of a graph G is invariant under isomorphism. Under this condition, see Frank and Strauss (1986), any Markov graph G has probability

$$P(G) = c(\boldsymbol{\theta})^{-1} \exp(\boldsymbol{\theta}'\mathbf{x}) \tag{2.2}$$

where $\boldsymbol{\theta}$ is a k-dimensional vector of model parameters, $\mathbf{x}$ is a k-dimensional vector of the observed values of sufficient statistics of G, and $c(\boldsymbol{\theta})$ is a normalizing function of $\boldsymbol{\theta}$, equal to $\sum_G \exp(\boldsymbol{\theta}'\mathbf{x})$. Here $'$ denotes transposition and all vectors are understood to be column vectors. For definitions of the sufficient statistics in various models, see Frank and Strauss (1986). As a special case, when $\boldsymbol{\theta} = \mathbf{0}$, the probability (2.2) reduces to $\binom{n(n-1)}{r}^{-1}$ or to $\binom{m}{r}^{-1}$, and hence, G is a uniform graph with a fixed number of arcs or edges, and mutual independence between the elements of $\mathbf{Y}$. For a review of various properties of uniform graphs, see, for instance, Frank (1988).

3. Maximum likelihood estimation for Markov graphs with a univariate sufficient statistic

Consider a special case of the general model (2.2), where $\boldsymbol{\theta}$ and $\mathbf{x}$ are one-dimensional, and denoted by θ and x, respectively. A standard result for the exponential family is that the true maximum likelihood estimate $\hat{\theta}_{\text{ml}}$ is given as the solution to the equation

$$E_\theta(X) = \frac{\sum_G X \exp(\theta X)}{\sum_G \exp(\theta X)} = x \tag{3.1}$$

where x is the observed value of the sufficient statistic X. The complicated form of the normalizing constant in the model makes it difficult to obtain the solution directly from (3.1). Therefore, we shall expand the expected value in terms of cumulants of X (cf. with Frank and Strauss 1986), and use an iterative method to find the estimate $\hat{\theta}_{\text{ml}}$. However, for certain values of x, the maximum likelihood estimate is directly obtainable. First, it follows from (3.1), that for no finite value of θ, the expectation $E_\theta(X)$ is equal to $\min(X)$ or $\max(X)$. Hence, if the observed value x is equal to the minimum (maximum), then the estimate $\hat{\theta}_{\text{ml}}$ is equal to $-\infty$ (∞). Second, as was noted in Section 2, when the parameter θ is equal to 0, the Markov graph reduces to a uniform graph. Therefore, for x equal to the expected value of X in a uniform graph, the estimate $\hat{\theta}_{\text{ml}}$ equals 0.

For an arbitrary value of x apart from the minimum and maximum, the cumulant generating function of X can be used to determine the value of $\hat{\theta}_{\text{ml}}$. For a general discussion on cumulants, see Stuart and Ord (1994). It can be shown that the cumulant generating function of X, $C_X(\Delta) = \log E \exp(\Delta X)$, equals

$$\log c(\theta + \Delta) - \log c(\theta) \tag{3.2}$$

where $c(\cdot)$ is the normalizing function defined in the previous section. Since the cumulants are defined as the coefficients $K_i(\theta)$, $i = 1, 2, ...$, in the identity

$$C_X(\Delta) = \sum_{i=0}^{\infty} K_i(\theta) \frac{\Delta^i}{i!} \tag{3.3}$$

we can write the expectation as

$$E_{\theta+\Delta}(X) = \frac{\partial}{\partial\Delta}\log c(\theta+\Delta) = \sum_{i=1}^{\infty} K_i(\theta)\frac{\Delta^{i-1}}{(i-1)!} \tag{3.4}$$

By truncating the series expansion after a suitable number of terms, we can equal (3.4) with x iteratively to find the estimate $\hat{\theta}_{\mathrm{ml}}$. For an arbitrary value θ_0, the new, updated value of the estimate is $\theta_0 + \tilde{\Delta}$, where $\tilde{\Delta}$ is a root of

$$x - \sum_{i=1}^{j} K_i(\theta_0)\frac{\Delta^{i-1}}{(i-1)!} = 0 \tag{3.5}$$

The equation (3.5) can be solved explicitly for j up to order 5. A natural initial value for the iterative estimation procedure is $\theta_0 = 0$, for which explicit expressions for cumulants of low order are obtainable. Let $\hat{\theta}^{(j)}_{\mathrm{cum}}$ denote the estimate of θ obtained by using the first j cumulants $K_1(\theta_0), ..., K_j(\theta_0)$ given $\theta_0 = 0$. The estimate

$$\hat{\theta}^{(2)}_{\mathrm{cum}} = \frac{x - K_1(0)}{K_2(0)} \tag{3.6}$$

was suggested in Frank and Strauss (1986). A third order estimate $\hat{\theta}^{(3)}_{\mathrm{cum}}$ is obtained by solving the equation

$$x - K_1(0) + K_2(0)\Delta + K_3(0)\frac{\Delta^2}{2!} = 0 \tag{3.7}$$

We have derived algebraic formulas for the cumulants $K_1(0), K_2(0)$ and $K_3(0)$ for two sufficient statistics of an undirected graph (see the end of this section). These are used in Section 5 to illustrate the performance of the estimate $\hat{\theta}^{(3)}_{\mathrm{cum}}$, which is used as a starting value in the estimation procedure.

Generally, when $\theta_0 \neq 0$, the cumulants have to be obtained by simulating the distribution of X for θ_0. Here, we describe a general procedure for simulating Markov graphs with the *Metropolis-Hastings* algorithm, proposed in Strauss (1986). Given a fixed value θ_0, the algorithm produces a Markov chain $G_t; t = 0, 1, ...,$ of random graphs, such that the stationary distribution of the chain is in accordance with (2.2). At the initial step $t = 0$, G_0 is set to a randomly chosen Bernoulli graph. Let $\mathbf{Y}_t$ denote the adjacency matrix of G_t. Further, at step t, let G^* be a *proposal graph*, obtained by setting two randomly chosen elements

Y_{tuv} and Y_{twz}, such that $Y_{tuv} = 1$ and $Y_{twz} = 0$, equal to 0 and 1, respectively. The next graph G_{t+1} is set equal to G^* with probability α, where α is given by

$$\min\left(1, \frac{P(G^*)}{P(G_t)}\right) = \min(1, \exp(\theta_0(x^* - x_t))) \tag{3.8}$$

and with probability $1 - \alpha$, G_{t+1} is set equal to G_t. In (3.8) x^* and x_t denote the observed value of X in G^* and G_t, respectively. From the simulated Markov chain, it is possible to obtain estimates of cumulants of X up to an arbitrary order j. A highly efficient way of coding the Metropolis-Hastings algorithm is to divide all possible nonredundant elements $\{u, v\}$ in the graph G into two vectors, one with $Y_{uv} = 1$, and the other with $Y_{uv} = 0$. A proposal graph may then be generated by randomly switching elements between the two vectors, and the corresponding change in the value of the sufficient statistic is easily calculated. We have programmed such an algorithm in Fortran, and the resulting code enables generation of 10^6 Markov graphs within approximately 3 seconds on a PC with a 150MHz Pentium processor (for n up to 40).

To ensure that the simulated values are a sample from the stationary distribution, a burn-in period of suitable length, say 1000, should be excluded from the beginning of the sequence. For general details on the Metropolis-Hastings algorithm, and for a discussion on the convergence of Markov chains, see, for instance, Gilks et al. (1996) or Gamerman (1997).

We have derived analytic expressions for the first three cumulants of two sufficient statistics in an undirected Markov graph with $\theta = 0$. In a clustering model, the sufficient statistic $X = S$ is the number of two-stars in G and $\theta = \sigma$ is the clustering parameter. In a transitivity model, $X = T$ is the number of triangles and $\theta = \tau$ is the transitivity parameter. For more information on these two models, see Frank and Strauss (1986). The cumulants up to order 3 are related to the moments according to

$$\begin{aligned} K_1 &= E(X) \\ K_2 &= \mathrm{Var}(X) = E(X^2) - (E(X))^2 \\ K_3 &= E(X^3) - 3E(X^2)E(X) + 2(E(X))^3 \end{aligned} \tag{3.9}$$

Let $v^{(l)}$ denote $v!/(v-l)!$, and let $I(.)$ be an indicator function which is equal to 1 when the argument is true, and 0 otherwise. Further, let $g(y) = \frac{\binom{m-y}{r-y}}{\binom{m}{r}}$. For the number of two-stars we have

$$E(S) = \frac{2r(r-1)}{n+1} \tag{3.10}$$

$$\begin{aligned}
\mathrm{Var}(S) &= (2n-5)\frac{n^{(3)}r^{(3)}}{m^{(3)}} + \frac{n^2-6}{4}\frac{n^{(4)}r^{(4)}}{m^{(4)}} - E(S)[E(S)-1] \\
E(S^3) &= E(S) + 6Z_1 + 6Z_2
\end{aligned}$$

where Z_1 and Z_2 are given by

$$\begin{aligned}
Z_1 &= \frac{2n-5}{2}\frac{n^{(3)}r^{(3)}}{m^{(3)}} + \frac{n^2-6}{8}\frac{n^{(4)}r^{(4)}}{m^{(4)}} \qquad (3.11) \\
Z_2 &= \frac{n-2}{6}\frac{n^{(3)}r^{(3)}}{m^{(3)}} + \frac{16n-37}{6}\frac{n^{(4)}r^{(4)}}{m^{(4)}} + 4\frac{n^{(4)}r^{(5)}}{m^{(5)}} + \frac{2n^2+3n-8}{4}\frac{n^{(5)}r^{(5)}}{m^{(5)}} + \\
&+ \frac{n^3+6n^2+14n+60}{48}\frac{n^{(6)}r^{(6)}}{m^{(6)}} + \frac{47}{6}\frac{n^{(5)}r^{(6)}}{m^{(6)}} + \frac{1}{3}\frac{n^{(4)}r^{(6)}}{m^{(6)}}
\end{aligned}$$

Expressions for $E(S)$ and $\mathrm{Var}(S)$ in this special case were also given in Frank and Strauss (1986). Correspondingly, for the triangles we have

$$\begin{aligned}
E(T) &= \binom{n}{3} g(3) \qquad (3.12) \\
E(T^2) &= \binom{n}{3} \sum_{j=0}^{3} \binom{3}{j}\binom{n-3}{3-j} g\left(6 - \binom{j}{2}\right) \\
E(T^3) &= \binom{n}{3} \sum_{j=0}^{3} \binom{3}{j}\binom{n-3}{3-j} z(j)
\end{aligned}$$

where

$$\begin{aligned}
z(j) &= g\left(9-\binom{j}{2}\right)\left[\binom{n-6+j}{3} + (6-j)\binom{n-6+j}{2} + (3-j)^2(n-6+j)\right] + \qquad (3.13) \\
&\quad g\left(8-\binom{j}{2}\right)\left[n\left(6-\binom{j}{2}\right) - 18I(j=0) - 26I(j=1) - 20I(j=2) - 9I(j=3)\right] + \\
&\quad g\left(7-\binom{j}{2}\right)[4I(j=1) + 2I(j=2)] + \\
&\quad g\left(6-\binom{j}{2}\right)[2 - I(j=3)]
\end{aligned}$$

The above expressions for $E(T)$ and $E(T^2)$ were earlier obtained by Frank (1971). For a general discussion on the moment properties of subgraph counts in random graphs, see Frank (1979).

4. Maximum likelihood estimation for general Markov graphs

We now consider a generalization of the estimation technique introduced in the previous section. In a general model we have a k-dimensional vector $\mathbf{X}$ of sufficient statistics and a vector $\boldsymbol{\theta}$ of the corresponding parameters. For single components of $\mathbf{X}$ and $\boldsymbol{\theta}$ we write X_i and θ_i, respectively. As in the univariate case, the maximum likelihood estimate $\hat{\boldsymbol{\theta}}_{\mathrm{ml}}$ is obtained as the solution to

$$E_{\boldsymbol{\theta}}(\mathbf{X}) = \frac{\sum_G \mathbf{X} \exp(\boldsymbol{\theta}'\mathbf{X})}{\sum_G \exp(\boldsymbol{\theta}'\mathbf{X})} = \mathbf{x} \tag{4.1}$$

and analogously, the estimate $\hat{\boldsymbol{\theta}}_{\mathrm{ml}}$ is equal to $\mathbf{0}$ when $\mathbf{x}$ equals the expected value of $\mathbf{X}$ in a uniform graph. However, the situation is more complicated in the general model when any x_i equals the minimum or maximum value of X_i. Then we know that the estimate of θ_i has no finite value. The simplest solution to this problem is to exclude the parameters from the model for which no maximum likelihood estimate exists, and then proceed with the estimation of the remaining parameters. Another possibility is to condition the model on all sufficient statistics X_i for which the observed value x_i equals the minimum or maximum.

As in Section 3 we use the cumulant generating function to find a solution to the likelihood equation (4.1). The multivariate cumulant generating function, see Stuart and Ord (1994), can be written as

$$C_{\boldsymbol{\theta}+\boldsymbol{\Delta}}(\mathbf{X}) = \sum_{w_1,\ldots,w_k=0}^{\infty} K^{12\ldots k}_{w_1 w_2 \ldots w_k}(\boldsymbol{\theta}) \frac{\Delta_1^{w_1}\Delta_2^{w_2}\cdots\Delta_k^{w_k}}{w_1!w_2!\cdots w_k!} \tag{4.2}$$

where the coefficients $K^{12\ldots k}_{w_1 w_2 \ldots w_k}(\boldsymbol{\theta})$ are the multivariate cumulants and $\boldsymbol{\Delta}$ denotes the vector $(\Delta_1, \Delta_2, \ldots, \Delta_k)'$. Define the order of a cumulant $K^{12\ldots k}_{w_1 w_2 \ldots w_k}(\boldsymbol{\theta})$ as $\sum_{i=1}^k w_i$. In order to obtain a useful expression for the expectation (4.1) we use a linear approximation to (4.2) and set all cumulants of order 3 or higher equal to 0. To simplify the notation we exclude all null indices in $K^{12\ldots k}_{w_1 w_2 \ldots w_k}(\boldsymbol{\theta})$ and write $K^{i}_{w_i}(\boldsymbol{\theta})$ and $K^{ij}_{w_i w_j}(\boldsymbol{\theta})$, for $K^{12\ldots k}_{0\ldots w_i \ldots 0}(\boldsymbol{\theta})$ and $K^{12\ldots k}_{0\ldots w_i w_j \ldots 0}(\boldsymbol{\theta})$, respectively. In particular, $K^{ij}_{11}(\boldsymbol{\theta})$ is the covariance of X_i and X_j. Given the linear approximation we may write the cumulant generating function of $\mathbf{X}$ as

$$\begin{bmatrix} K^1_1(\boldsymbol{\theta})\Delta_1 + \frac{1}{2}K^1_2(\boldsymbol{\theta})\Delta_1^2 + K^{12}_{11}(\boldsymbol{\theta})\Delta_1\Delta_2 + \ldots + K^{1k}_{11}(\boldsymbol{\theta})\Delta_1\Delta_k \\ \vdots \\ K^k_1(\boldsymbol{\theta})\Delta_k + \frac{1}{2}K^k_2(\boldsymbol{\theta})\Delta_k^2 + K^{k1}_{11}(\boldsymbol{\theta})\Delta_k\Delta_1 + \ldots + K^{k(k-1)}_{11}(\boldsymbol{\theta})\Delta_k\Delta_{k-1} \end{bmatrix} \tag{4.3}$$

By differentiating (4.3) we obtain, with some algebra, an expression for the expectation (4.1) as

$$E_{\boldsymbol{\theta}+\boldsymbol{\Delta}}(\mathbf{X}) = \mathbf{K}_1(\boldsymbol{\theta}) + \mathbf{K}_{11}(\boldsymbol{\theta})\boldsymbol{\Delta} \tag{4.4}$$

where $\mathbf{K}_1(\boldsymbol{\theta})$ is the mean vector and $\mathbf{K}_{11}(\boldsymbol{\theta})$ the covariance matrix of $\mathbf{X}$. At $\boldsymbol{\theta}_0$, solving the system of linear equations

$$\mathbf{x} - \mathbf{K}_1(\boldsymbol{\theta}_0) - \mathbf{K}_{11}(\boldsymbol{\theta}_0)\boldsymbol{\Delta} = \mathbf{0} \tag{4.5}$$

for $\boldsymbol{\Delta}$ gives the updated estimate $\boldsymbol{\theta}_0+\tilde{\boldsymbol{\Delta}}$. Although it is possible to obtain a closed form solution for a general k, the expression becomes quite lengthy already for $k = 4$, and is suppressed here. For an illustration consider the bivariate case, where the solution is

$$\tilde{\Delta}_1 = -\frac{K_{11}^{12}(\boldsymbol{\theta}_0)x_2 - K_{11}^{12}(\boldsymbol{\theta}_0)K_1^2(\boldsymbol{\theta}_0) - K_2^2(\boldsymbol{\theta}_0)x_1 + K_2^2(\boldsymbol{\theta}_0)K_1^1(\boldsymbol{\theta}_0)}{-K_{11}^{12}(\boldsymbol{\theta}_0) + K_2^2(\boldsymbol{\theta}_0)K_2^1(\boldsymbol{\theta}_0)} \tag{4.6}$$

$$\tilde{\Delta}_2 = \frac{-K_{11}^{12}(\boldsymbol{\theta}_0)x_1 + K_{11}^{12}(\boldsymbol{\theta}_0)K_1^1(\boldsymbol{\theta}_0) + K_2^1(\boldsymbol{\theta}_0)x_2 - K_2^1(\boldsymbol{\theta}_0)K_1^2(\boldsymbol{\theta}_0)}{-K_{11}^{12}(\boldsymbol{\theta}_0) + K_2^2(\boldsymbol{\theta}_0)K_2^1(\boldsymbol{\theta}_0)}$$

For the iterative estimation procedure we set the initial value $\boldsymbol{\theta}_0$ equal to $\mathbf{0}$. For this particular value it is possible to derive algebraic expressions for the covariances $K_{11}^{ij}(\boldsymbol{\theta}_0)$, and we have done this in a bivariate case (see below). In general, when $\boldsymbol{\theta}_0 \neq \mathbf{0}$, the cumulants have to be obtained by simulating the distribution of $\mathbf{X}$ for $\boldsymbol{\theta}_0$. Since this can be done by a straightforward modification of the univariate simulation, we omit the details here.

As an illustration of general models we consider a Markov graph model with S and T as sufficient statistics. The covariance of S and T in a uniform graph can be written as

$$\begin{aligned} K_{11}^{ST}(\mathbf{0}) \quad = \quad & \binom{n}{3}(3n-9)g(5) + \\ & \binom{n}{3}\sum_{j=0}^{3}\left[\binom{3}{j}\binom{n-3}{3-j}\left[\left(3-\binom{j}{2}\right)+3\binom{j}{3}\right]g\left(5-\binom{j}{2}+\binom{j}{3}\right)\right] \\ & -E(S)E(T) \end{aligned} \tag{4.7}$$

where $g(.), E(S)$ and $E(T)$ are defined as in Section 3.

Due to the linear approximation used in (4.3) it may happen that the stepwise estimation procedure fails to converge when the initial estimates based on (4.6) are far from $\hat{\boldsymbol{\theta}}_{\mathrm{ml}}$. Alternative starting values are then needed in the estimation procedure. One possibility is to simulate the expectation $E_{\boldsymbol{\theta}}(\mathbf{X})$ at a grid of values of $\boldsymbol{\theta}$, and seek for values closer to $\hat{\boldsymbol{\theta}}_{\mathrm{ml}}$. This is computationally feasible even for a large value of k due to the efficiency of the Metropolis-Hasting algorithm. It is also possible to replace the relatively simple updating procedure for the parameter estimates at the successive iteration steps by a more complex algorithm including control parameters which ensure that the updating does not lead the estimation procedure towards infinite values.

We shall describe briefly the pseudolikelihood method introduced in Frank and Strauss (1986), which has been used for the estimation of Markov graph parameters by Strauss and Ikeda (1990), Frank (1991), Frank and Nowicki (1993), and Wasserman and Pattison (1996). In the pseudolikelihood method, the log-odds of the probability of $Y_{uv} = 1$, conditional on the rest of the graph, is written as

$$\text{logit } P(Y_{uv} = 1|\text{rest}) = \sum_{i=1}^{k} \theta_i \delta_{uv}^i \tag{4.8}$$

where δ_{uv}^i is equal to the *difference* between the value of X_i when we set $Y_{uv} = 1$ and $Y_{uv} = 0$, respectively. The representation (4.8) enables the estimation of θ_i's to be done for $k \geq 1$ with the available standard logistic regression packages. One advantage of the use of (4.8) is that the normalizing function $c(\boldsymbol{\theta})$ is not involved in the estimation procedure. However, it should be noted that the pseudolikelihood method is not based on the sufficient statistics X_i, but on the "change variables" δ_{uv}^i.

5. Enumeration results in small graphs

To investigate the performance of the iterative estimates and to compare them with the pseudolikelihood estimates, we have chosen a graph with $n = 7, r = 12$, and used complete enumeration to find the solutions to the equation (3.1). There are 293930 distinct graphs for $n = 7, r = 12$ and the distribution of the sufficient statistics is given for the clustering and transitivity models (Section 3) in Table 1.

Value of S	Frequency	Value of T	Frequency
30	19355	0	35
31	45360	2	1260
32	71190	3	5880
33	48055	4	39375
34	51030	5	78330
35	27090	6	87360
36	18585	7	52080
37	8190	8	25795
38	4200	9	1260
39	875	10	2135
		11	420
Total	293930		293930

Table 1. Distributions of S and T, respectively, for an undirected graph with $n = 7, r = 12$.

By using the distributions in Table 1, it is possible to obtain the maximum likelihood estimates of σ and τ as solutions to (3.1) for each possible value of the sufficient statistics. In Table 2, estimates of σ using different methods are given. Here, $\hat{\sigma}_{\text{ml}}$ is the true maximum likelihood estimate, and $\hat{\sigma}_{\text{iter}}$ is the estimate obtained by using several steps of iteration with $\hat{\sigma}^{(3)}_{\text{cum}}$ as a starting value. For comparison, the pseudolikelihood estimates $\hat{\sigma}_{\text{pl}}$ are also given. For some values of the sufficient statistic, the pseudolikelihood estimate failed to converge, these cases are denoted by $*$. The corresponding estimates of τ are given in Table 3.

In the iterative estimation procedure we have initially used Markov chains of length 50000, and then increased the length with 50000 at each iteration step to obtain better precision of the estimates of cumulants in the neighbourhood of $\hat{\theta}_{\text{ml}}$. Estimates of the four first cumulants were used at each iteration step to obtain a new value for $\hat{\theta}_{\text{iter}}$. For S and T, the minimum and maximum values can be obtained numerically for any n and r using certain combinatorial functions.

From Tables 2 and 3 it can be seen that the iterative estimates are in a very close agreement with the true maximum likelihood estimates for all possible values of the sufficient statistics. It is reasonable to assume that the iterative estimates perform similarly in larger graphs. Moreover, the estimates based on the three first cumulants at $\theta = 0$ are surprisingly good compared to the pseudolikelihood estimates. For most values of S and T, the estimate $\hat{\theta}^{(3)}_{\text{cum}}$ is closer to the maximum likelihood estimate than the pseudolikelihood estimate, which is considerably biased. Due to variability of the pseudolikelihood estimates among graphs with the same value of the sufficient statistic, the pseudolikelihood estimates given in Tables 2 and 3 are arithmetic means of the pseudolikelihood estimates for all possible graphs with the specified value of the sufficient statistic. The variability of the pseudolikelihood estimates is considered further in the next section.

Value of S	$\hat{\sigma}_{\text{ml}}$	$\hat{\sigma}^{(3)}_{\text{cum}}$	$\hat{\sigma}_{\text{iter}}$	$\hat{\sigma}_{\text{pl}}$
30	$-\infty$	$-\infty$	$-\infty$	*
31	$-.978$	$-.970$	$-.979$	-1.859
32	$-.350$	$-.352$	$-.355$	$-.851$
33	0	0	0	$-.391$
34	.257	.255	.258	$-.077$
35	.486	.465	.487	.183
36	.727	.648	.730	.434
37	1.038	.812	1.034	.731
38	1.597	.962	1.612	1.173
39	∞	∞	∞	2.065

Table 2. Estimates of σ for an undirected graph with $n = 7, r = 12$.

Value of T	$\hat{\tau}_{\text{ml}}$	$\hat{\tau}^{(3)}_{\text{cum}}$	$\hat{\tau}_{\text{iter}}$	$\hat{\tau}_{\text{pl}}$
0	$-\infty$	$-\infty$	$-\infty$	*
2	-1.936	-3.34	-1.935	-2.841
3	-1.571	-2.69	-1.570	-2.282
4	-1.103	-1.27	-1.102	-2.144
5	$-.485$	$-.490$	$-.485$	-1.108
6	.119	.119	.119	$-.290$
7	.610	.611	.609	.427
8	.989	1.09	.989	.861
9	1.339	1.51	1.34	.739
10	1.811	1.89	1.810	1.425
11	∞	∞	∞	1.721

Table 3. Estimates of τ for an undirected graph with $n = 7, r = 12$.

We now consider iterative estimation for the Markov graph model with S and T as sufficient statistics. The bivariate distribution of the statistics is given in Table 4.

$S\backslash T$	0	2	3	4	5	6	7	8	9	10	11
30	35	1260	3780	9660	4620	–	–	–	–	–	–
31	–	–	1260	18900	20790	4410	–	–	–	–	–
32	–	–	840	9450	39060	18900	2940	–	–	–	–
33	–	–	–	735	11340	30030	5460	490	–	–	–
34	–	–	–	630	2520	27720	17640	2520	–	–	–
35	–	–	–	–	–	4620	16590	5670	–	210	–
36	–	–	–	–	–	1680	6720	10185	–	–	–
37	–	–	–	–	–	–	2520	4410	1260	–	–
38	–	–	–	–	–	–	–	2520	–	1680	–
39	–	–	–	–	–	–	210	–	–	245	420

Table 4. The bivariate distribution of S and T for an undirected graph with $n = 7, r = 12$.

Using the values in Table 4 we have calculated the true maximum likelihood estimates of (σ, τ) for all possible values of S and T, such that neither S nor T is equal to the minimum or maximum value. These estimates are given in Table 5. At some boundary values of (S, T) the maximum likelihood estimates are numerically unstable, that is they tend towards infinity. Such values are $(31, 6), (32, 7), (33, 8)$ and $(35, 10)$, for which the estimates given in Table 5 are the smallest (in absolute value) with the deviation $|E_{\boldsymbol{\theta}}(S) - s| + |E_{\boldsymbol{\theta}}(T) - t| \leq 0.005$. The reason why the estimates are nearly divergent at these values of the sufficient statistics is not quite clear. However, our simulation results for models with a univariate sufficient statistic (Section 7) show that in larger graphs, values of sufficient statistics close to their maximum are observed only when the model behavior is degenerate. It might be that such degeneracy problems appear already for small n in general Markov models. The iterative estimates of (σ, τ) based on the starting values (4.6) at $\boldsymbol{\theta}_0 = \mathbf{0}$ are given in Table 6.

S\T	3	4	5	6	7	8	9	10
31	.397,-1.758	-.396,-.851	-1.466,.781	-6.150,6.140				
32	1.396,-2.426	.751,-1.661	-.136,-.349	-1.261,1.498	-5.010,5.640			
33		1.459,-2.293	.790,-1.321	-.234,.398	-1.245,2.025	-5.400,6.390		
34		2.082,-2.961	1.425,-2.030	.640,-.672	-.404,1.097	-1.239,2.394		
35				1.291,-1.515	.443,.073	-.547,1.635		-3.350,6.700
36				1.811,-2.216	1.124,-.743	.310,.630		
37					1.631,-1.339	1.140,-.165	.240,1.080	
38						1.895,-.550		.215,1.590

Table 5. The true maximum likelihood estimates of (σ, τ) for an undirected graph with $n = 7, r = 12$.

S\T	3	4	5	6	7	8	9	10
31	.396, -1.758	-.396, -.851	-1.466, .781	-8.304, 8.303				
32	*	.751, -1.661	-.136, -.349	-1.261, 1.497	-8.347, 8.971			
33		1.459, -2.293	.790, -1.321	-.234, .398	-1.245, 2.025	-7.516, 8.504		
34		*	1.424, -2.030	.640, -.673	-.404, 1.097	-1.240, 2.394		
35				1.291, -1.515	.443, .072	-.547, 1.635		-4.108, 8.223
36				1.810, -2.215	1.124, -.743	.317, .630		
37					1.631, -1.339	1.142, -.168	.245, 1.075	
38						1.895, -.549		.207, 1.594

Table 6. The iterative estimates of (σ, τ).

For the two values $(32, 3)$ and $(34, 4)$ of (S, T) the estimation procedure failed to converge. We experimented with various starting values of $\hat{\boldsymbol{\theta}}$ for these two particular cases, and found that convergence was achieved for the starting values at $(2.50, -3.25)$ or closer to the maximum likelihood estimate for $(32, 3)$. The iterative estimate with $(2.50, -3.25)$ as the starting value attained the value $(1.396, -2.425)$. Correspondingly, for $(34, 4)$ convergence was achieved for the starting values at $(3.20, -4.00)$ or closer to the maximum likelihood estimate. Given this particular starting value, the iterative estimate attained the value $(2.081, -2.960)$.

The results in Table 5 show that it can be difficult to interpret the obtained parameter estimates in general models. Due to the complex dependence between the two sufficient statistics considered here, the estimates cannot be interpreted as in the models with a single sufficient statistic, where a small or large parameter value corresponds to an observed graph which is extreme with respect to the property represented by the particular sufficient statistic. For instance, for

$(S,T) = (31,6)$ the maximum likelihood estimate equals $(-6.150, 6.140)$. The extremely large value of $\hat{\tau}$ thus indicates a high degree of transitivity in the observed graph, whereas the value $T = 6$ is closest to the expected value of T in a uniform graph. Hence, we would expect to observe the same degree of transitivity in a simple random graph. The value $T = 6$ is extreme in the conditional distribution of T given $S = 31$, which leads to the large value of $\hat{\tau}$. In practise the obtained estimate of θ_i should therefore be interpreted in terms of the conditional distribution of X_i given the observed values of the remaining sufficient statistics $X_1, ..., X_{i-1}, X_{i+1}, ..., X_k$.

As for the univariate models, the iterative estimates are well in accordance with the maximum likelihood estimates in the bivariate case, and the pseudolikelihood estimates given in Table 7 are considerably biased. For approximately 30% of the possible values of (S,T), pseudolikelihood estimation procedure failed to converge, these cases are denoted by $*$. Deviance between the values of the iterative and true maximum likelihood estimators is negligible, except for the values $(31,6), (32,7), (33,8)$ and $(35,10)$, for which the true maximum likelihood estimates are nearly divergent. For these values the iterative estimates are larger in absolute value, but they preserve quite accurately the relation between the values of $\hat{\sigma}$ and $\hat{\tau}$.

$S\backslash T$	0	2	3	4	5	6
30	*	*	*	*	*	
31			-.338,-1.302	-1.044,-1.163	-4.410,2.458	*
32			.902,-2.539	.405,-2.958	-.531,-.467	-2.152,2.366
33				1.014,-1.971	.629,-1.788	-.338,-1.302
34				*	1.499,-2.788	.829,-1.267
35						1.171,-1.442
36						*

$S\backslash T$	7	8	9	10	11
32	*				
33	-2.209,3.320	*			
34	-1.006,1.414	-2.338,3.115			
35	.003,.323	-2.171,3.629		*	
36	1.217,-1.452	-.479,1.437			
37	1.651,-2.231	.967,-.396	-.412,1.140		
38		1.620,-.758		-.537,2.097	
39	2.208,-2.391			*	-.459,2.161

Table 7. The pseudolikelihood estimates of (σ, τ).

6. Specific properties of the pseudolikelihood estimation

In the papers concerning the pseudolikelihood estimation for Markov graphs there has not so far appeared any discussion about the variability of pseudolikelihood estimators for graphs with fixed values of the sufficient statistics. To investigate how the use of δ^i_{uv} affects the estimation, we have obtained the pseudolikelihood estimates of σ and τ for all possible graphs with $n = 7, r = 12$ and with $n = 8, r = 14$, respectively. All possible graphs in the former case were summarized in Table 1. In the latter case there are 40116600 distinct graphs, and it does not seem feasible to study enumeratively still larger graphs. Estimation results for σ and τ are given in Tables 8 and 9, respectively, for the graphs with $n = 7, r = 12$. Corresponding results for the graphs with $n = 8, r = 14$ are given in Tables 10 and 11.

S	#graphs with convergence	#graphs without convergence	Mean	Std.dev.	Min	Max
30	0	19355	*	*	*	*
31	43470	1890	−1.859	.223	−2.167	−1.602
32	70770	420	−.851	.049	−.981	−.792
33	48055	0	−.391	.011	−.427	−.381
34	51030	0	−.077	.0003	−.078	−.076
35	27090	0	.183	.001	.181	.186
36	18585	0	.434	.009	.426	.456
37	8190	0	.731	.029	.677	.765
38	4200	0	1.173	.112	1.012	1.260
39	630	245	2.065	.475	1.723	2.736

Table 8. Properties of $\hat{\sigma}_{\text{pl}}$ for given values of S for graphs with $n = 7, r = 12$.

T	#graphs with convergence	#graphs without convergence	Mean	Std.dev.	Min	Max
0	0	35	*	*	*	*
2	1260	0	−2.841	.000	−2.841	−2.841
3	4620	1260	−2.282	.547	−2.758	−1.511
4	37590	1785	−2.144	.641	−3.379	−1.187
5	78330	0	−1.108	.274	−1.661	−.698
6	87360	0	−.290	.209	−.870	.127
7	52080	0	.427	.287	−.370	1.092
8	25725	70	.861	.243	.434	1.719
9	1260	0	.739	.000	.739	.739
10	2135	0	1.425	.126	1.047	1.562
11	420	0	1.721	.000	1.721	1.721

Table 9. Properties of $\hat{\tau}_{\text{pl}}$ for given values of T for graphs with $n = 7, r = 12$.

S	#graphs with convergence	#graphs without convergence	Mean	Std.dev.	Min	Max
36	0	1024380	*	*	*	*
37	2906400	57120	−2.032	.273	−2.442	−1.810
38	6074880	33600	−1.050	.069	−1.299	−.954
39	5404560	0	−.623	.026	−.777	−.595
40	7111440	0	−.350	.005	−.372	−.343
41	4841760	0	−.156	.0004	−.158	−.155
42	4844160	0	.000	.000	.000	.000
43	2772000	0	.135	.0003	.134	.136
44	2328900	0	.258	.002	.254	.282
45	1139040	0	.380	.007	.372	.447
46	910560	0	.503	.008	.496	.530
47	357840	0	.646	.009	.639	.725
48	208320	0	.831	.013	.796	.862
49	70560	0	1.090	.030	1.040	1.113
50	26880	3360	1.710	.395	1.482	2.395
52	840	0	2.295	.000	2.295	2.295

Table 10. Properties of $\hat{\sigma}_{\mathrm{pl}}$ for given values of S for graphs with $n = 8, r = 14$.

T	#graphs with convergence	#graphs without convergence	Mean	Std.dev.	Min	Max
0	0	5040	*	*	*	*
2	152880	40320	−2.193	.454	−3.283	−1.665
3	922320	30240	−2.128	.581	−3.577	−.932
4	3805620	7980	−1.667	.486	−3.200	−.746
5	8322720	0	−1.018	.270	−2.231	−.489
6	10445400	0	−.407	.162	−1.014	.000
7	8466240	0	.075	.188	−.647	.963
8	4825800	0	.490	.238	−.394	1.764
9	2204160	0	.802	.307	.329	3.506
10	561960	0	.878	.130	.563	1.136
11	246960	0	1.228	.255	.806	2.183
12	39900	0	1.281	.237	1.047	1.723
13	21840	0	1.409	.015	1.370	1.422
16	420	0	1.884	.000	1.884	1.884

Table 11. Properties of $\hat{\tau}_{\mathrm{pl}}$ for given values of T for graphs with $n = 8, r = 14$.

From the above tables it is evident that the specific graph with its specific combination of values in $\mathbf{Y}$ is decisive for the pseudolikelihood estimation, that is the "change variables" δ^i_{uv} do not give unique estimates. This is in contrast to the iterative method in which the sufficient statistics are used. Using simulation we have seen that the "between-graph" variation decreases relative to the total sampling variation when larger graphs are considered. However, the magnitude of difference between minimum and maximum values of the pseudolikelihood estimates is considerable for negative parameter values even for moderately large graphs. In order to save space, we do not present the results here, but they can be obtained from the authors.

Another problem associated with the pseudolikelihood estimates is the lack of convergence in special cases. Assume there exists a threshold value z, such that for all $\delta^i_{uv} < z$, the indicator Y_{uv} takes one specific value (say 1), while for all $\delta^i_{uv} > z$, the indicator Y_{uv} takes the opposite value (here 0). According to Haberman (1974), this is a sufficient condition for the nonexistence of the parameter estimate in a logistic regression model. An illustration of such a graph, one of the 70 graphs with $T = 8$ in Table 9, for which the pseudolikelihood estimate of τ does not exist, is given in Figure 1.

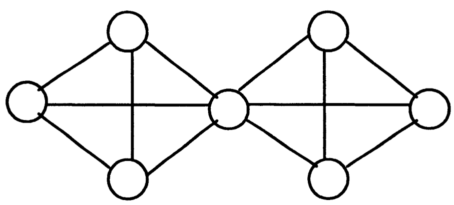

Figure 1. An undirected graph with $n = 7, r = 12$ for which the pseudolikelihood estimate of τ does not exist.

7. Comparison of the maximum likelihood and pseudolikelihood estimators

To investigate the statistical properties of the maximum likelihood and pseudolikelihood estimators, we have studied the clustering and transitivity parameters σ and τ in graphs with $n = 7, r = 12$, and the clustering parameter in graphs with (n, r) equal to $(25, 40), (40, 60), (50, 70), (60, 100)$ and $(100, 200)$, respectively.

Complete enumeration was used for the graphs with $n = 7, r = 12$, whereas simulation was used for the larger graphs. Properties of the maximum likelihood estimator were investigated using 10^6 replicates of the Markov model for each σ-value in the range -2,2(.5). The maximum likelihood estimates were here determined by a grid search using interpolation of σ, in order to determine the parameter values which correspond to the integer values of $E_\sigma(S)$.

Due to the extensive calculations involved, properties of the pseudolikelihood estimator were investigated using 10^5 replicates of the Markov model for the graphs with n equal to 25 and 40, whereas 10^4 replicates were used for the remaining graphs. Means, variances, mean squared errors (MSE) and the probability of convergence of the two estimators are given in Tables 12 to 18 (see Appendix). Plots of the bias and MSE of the estimators of σ are given in Figures 2 to 5 for n equal to 7 and 40.

To check precision of the calculations we used the simulation method also for the graph with $n = 7, r = 12$, and compared the results with those obtained by complete enumeration. Only some smaller deviations, usually in the third decimal, appeared.

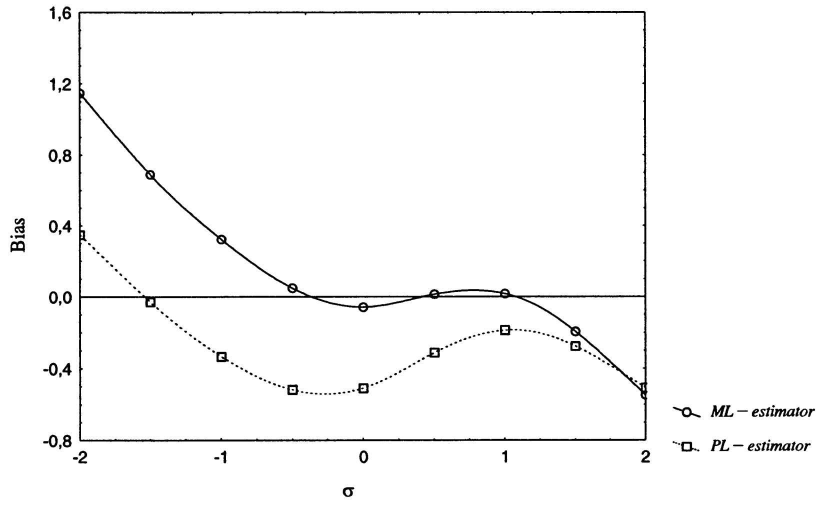

Figure 2. Bias of $\hat{\sigma}_{\mathrm{ml}}$ and $\hat{\sigma}_{\mathrm{pl}}$ for graphs with $n = 7, r = 12$.

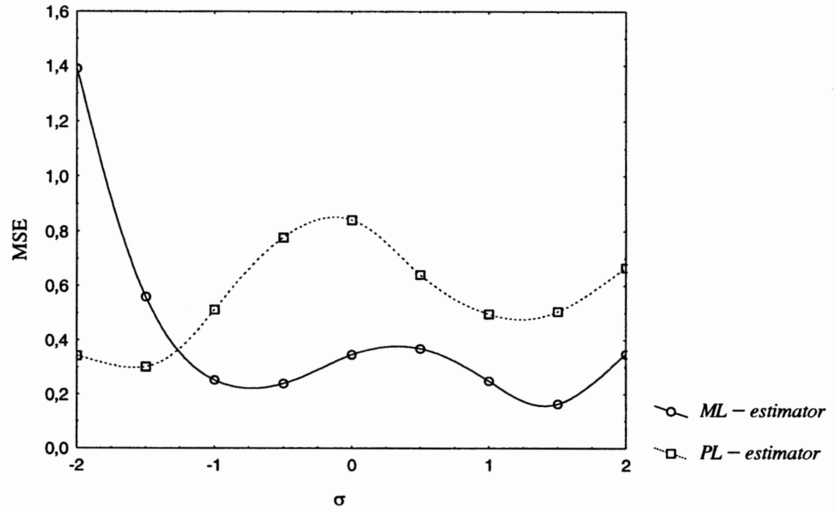

Figure 3. MSE of $\hat{\sigma}_{\mathrm{ml}}$ and $\hat{\sigma}_{\mathrm{pl}}$ for graphs with $n = 7, r = 12$.

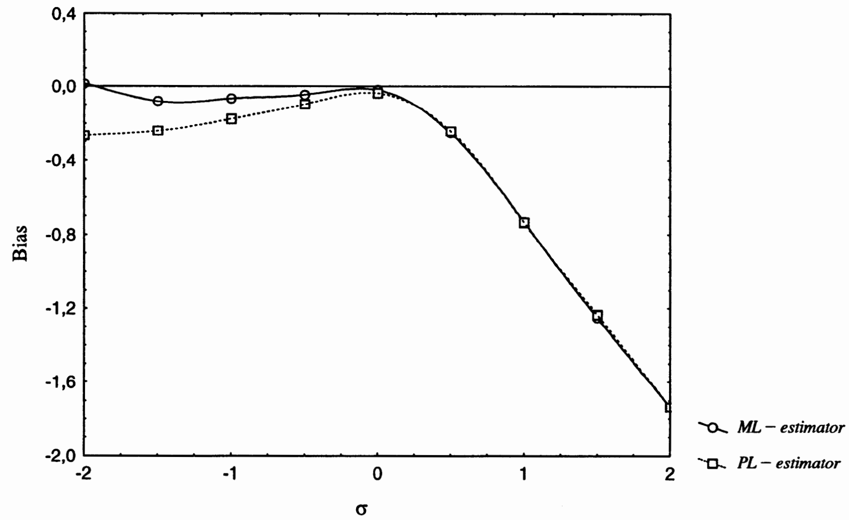

Figure 4. Bias of $\hat{\sigma}_{\mathrm{ml}}$ and $\hat{\sigma}_{\mathrm{pl}}$ for graphs with $n = 40, r = 60$.

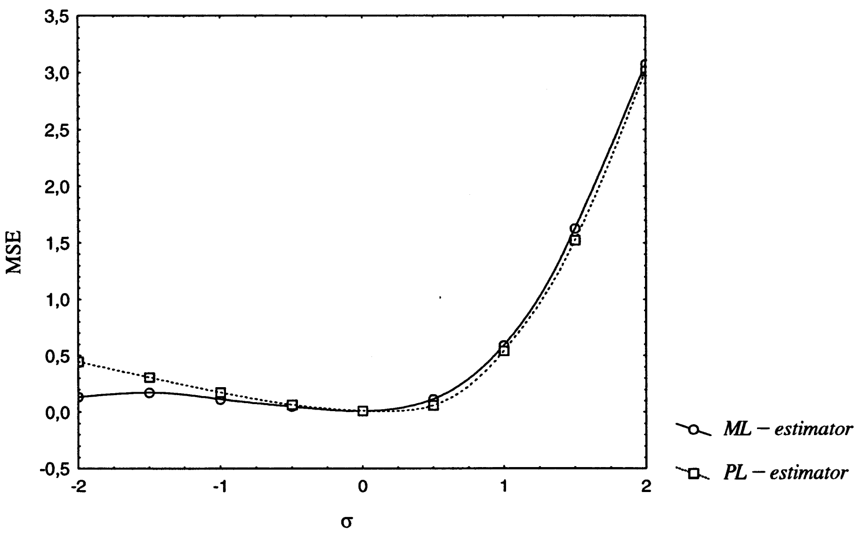

Figure 5. MSE of $\hat{\sigma}_{\text{ml}}$ and $\hat{\sigma}_{\text{pl}}$ for graphs with $n = 40, r = 60$.

It is evident from our simulation results that the maximum likelihood estimator generally has smaller bias and variance than the pseudolikelihood estimator. In the cases considered here, the pseudolikelihood estimator has smaller bias than the maximum likelihood estimator only in that part of the parameter space where the model behavior is degenerate. By model degeneracy we mean that the observed values of the sufficient statistics are concentrated on the extreme values, either at the minimum or maximum. Such extreme values of the sufficient statistics are not expected to be met in reality. The results for the larger graphs with 40-100 vertices suggest that the maximum likelihood and pseudolikelihood estimators are nearly equivalent with respect to the magnitude of the bias and MSE. Moreover, the bias of the estimators tends to 0 in the neighborhood of $\sigma = 0$ as the number of vertices increases. Thus, the estimators seem to be, in some sense, asymptotically consistent, for Markov models close to the uniform graph model.

It is evident that the behavior of the Markov model is problematic as n gets larger. The larger n is, the smaller positive value of a parameter is needed to produce graphs where the value of the sufficient statistic approaches its maximum. The asymptotic behavior of the Markov graph model and degeneracy problems were discussed in detail by Strauss (1986). The degeneracy problem is even more severe in models where the graph size r is allowed to be stochastic. An example of this is the transitivity model for a graph with $n = 30$, considered in Frank and Strauss (1986), where parameter values only in a narrow range around zero produce "realistic" graphs.

8. Remarks

The numerical results presented in the previous sections show that the maximum likelihood estimates seem to be accurate enough and computationally feasible to be preferred in routine data analysis. Moreover, we have shown that the pseudo-likelihood method does not yield unique estimates for fixed values of the sufficient statistics. Concerning the aspects of inference, likelihood ratio tests of the uniform model corresponding to $\boldsymbol{\theta} = \mathbf{0}$ could be performed, albeit approximately, by using the simulation method to obtain an approximation to the distribution of $P(\mathbf{x}|\boldsymbol{\theta} = \mathbf{0})/P(\mathbf{x}|\hat{\boldsymbol{\theta}})$ under the null hypothesis.

An important aspect of interpretation of parameters in Markov models is related to the size of n. Any particular parameter value distinct from 0 will have a different interpretation in graphs with different values of n. To avoid this complication the definition of the graph probability (2.2) should be altered in some suitable manner. It will be a task for future research to investigate this and other properties of Markov graphs.

References

[1] **Dahmström, K. and Dahmström, P.** (1993). ML-estimation of the clustering parameter in a Markov graph model. Research Report, Department of Statistics, Stockholm University.

[2] **Frank, O.** (1971). Statistical inference in graphs. Stockholm: Försvarets forskningsanstalt.

[3] **Frank, O.** (1979). Moment properties of subgraph counts in stochastic graphs. *Ann. N. Y. Academy Sciences*, **319**, 207-218.

[4] **Frank, O.** (1988). Random sampling and social networks. *Math. Inf. Sci. hum.*, **26**, 19-33.

[5] **Frank, O.** (1991). Statistical analysis of change in networks. *Statistica Neerlandica,* **45**, 283-293.

[6] **Frank, O. and Nowicki, K.** (1993). Exploratory statistical analysis of networks. *Ann. Disc. Math.*, **55**, 349-366.

[7] **Frank, O. and Strauss, D.** (1986). Markov graphs. *J. Amer. Stat. Assoc.,* **81**, 832-842.

[8] **Gamerman, D.** (1997). Markov Chain Monte Carlo: Stochastic simulation for Bayesian inference. London: Chapman&Hall.

[9] **Gilks, W., Richardson, S. and Spiegelhalter, D.** (1996). (Eds.) Markov chain Monte Carlo in practice. London: Chapman&Hall.

[10] **Haberman, S.** (1974). The analysis of frequency data. Chicago: University of Chicago Press.

[11] **Strauss, D.** (1986). On a general class of models for interaction. *SIAM Review*, **28**, 513-527.

[12] **Strauss, D. and Ikeda, M.** (1990). Pseudolikelihood estimation for social networks. *J. Amer. Stat. Assoc.,* **85**, 204-212.

[13] **Stuart, A. and Ord, K.** (1994). Kendall's advanced theory of statistics, Vol 1. 6th ed. London: Edward Arnold.

[14] **Wasserman, S. and Pattison, P.** (1996). Logit models and logistic regressions for social networks: An introduction to Markov graphs and p^*. *Psychometrika*, **61**, 401-425.

Appendix

$\underline{\sigma}$	$E(\hat{\sigma}_{ml})$	$E(\hat{\sigma}_{pl})$	$V(\hat{\sigma}_{ml})$	$V(\hat{\sigma}_{pl})$	$MSE(\hat{\sigma}_{ml})$	$MSE(\hat{\sigma}_{pl})$	P^{ml}_{conv}	P^{pl}_{conv}
−2.0	−.852	−1.652	.071	.221	1.389	.342	.281	.272
−1.5	−.810	−1.529	.084	.301	.560	.302	.426	.413
−1.0	−.677	−1.333	.148	.401	.252	.512	.607	.592
−.5	−.451	−1.018	.237	.509	.239	.778	.794	.781
0	−.058	−.509	.343	.582	.346	.842	.931	.926
.5	.514	.187	.369	.544	.369	.642	.955	.979
1.0	1.017	.812	.249	.461	.249	.496	.840	.955
1.5	1.305	1.223	.126	.428	.164	.505	.654	.903
2.0	1.453	1.492	.048	.409	.348	.667	.470	.852

Table 12. Properties of $\hat{\sigma}_{ml}$ and $\hat{\sigma}_{pl}$ for graphs with $n = 7, r = 12$.

$\underline{\sigma}$	$E(\hat{\sigma}_{ml})$	$E(\hat{\sigma}_{pl})$	$V(\hat{\sigma}_{ml})$	$V(\hat{\sigma}_{pl})$	$MSE(\hat{\sigma}_{ml})$	$MSE(\hat{\sigma}_{pl})$	P^{ml}_{conv}	P^{pl}_{conv}
−2.0	−1.126	−2.446	.752	.808	1.516	1.006	.983	.986
−1.5	−1.299	−1.905	.369	.557	.409	.721	.997	.998
−1.0	−1.070	−1.280	.176	.279	.181	.357	1.000	1.000
−.5	−.573	−.658	.085	.109	.090	.134	1.000	1.000
0	−.032	−.069	.018	.018	.019	.025	1.000	1.000
.5	.428	.332	.049	.001	.055	.029	.999	1.000
1.0	.356	.336	.123	.0006	.539	.403	.999	1.000
1.5	.373	.369	.130	.0006	1.400	1.280	.999	1.000
2.0	.355	.370	.133	.0006	2.839	2.657	.999	1.000

Table 13. Properties of $\hat{\sigma}_{ml}$ and $\hat{\sigma}_{pl}$ for graphs with $n = 25, r = 40$.

$\underline{\sigma}$	$E(\hat{\sigma}_{ml})$	$E(\hat{\sigma}_{pl})$	$V(\hat{\sigma}_{ml})$	$V(\hat{\sigma}_{pl})$	$MSE(\hat{\sigma}_{ml})$	$MSE(\hat{\sigma}_{pl})$	P^{ml}_{conv}	P^{pl}_{conv}
−2.0	−1.987	−2.267	.134	.3374	.134	.446	1.000	1.000
−1.5	−1.580	−1.740	.166	.281	.172	.308	1.000	1.000
−1.0	−1.066	−1.175	.110	.142	.114	.174	1.000	1.000
−.5	−.546	−.597	.047	.049	.049	.066	1.000	1.000
0	−.021	−.038	.009	.008	.009	.011	1.000	1.000
.5	.248	.259	.050	.003	.114	.059	.999	1.000
1.0	.270	.265	.056	.0004	.589	.541	.999	1.000
1.5	.249	.266	.061	.0004	1.627	1.523	.999	1.000
2.0	.266	.265	.064	.0004	3.069	3.009	.999	1.000

Table 14. Properties of $\hat{\sigma}_{ml}$ and $\hat{\sigma}_{pl}$ for graphs with $n = 40, r = 60$.

σ	$E(\hat\sigma_{\mathrm{ml}})$	$E(\hat\sigma_{\mathrm{pl}})$	$V(\hat\sigma_{\mathrm{ml}})$	$V(\hat\sigma_{\mathrm{pl}})$	$MSE(\hat\sigma_{\mathrm{ml}})$	$MSE(\hat\sigma_{\mathrm{pl}})$	$P^{\mathrm{ml}}_{\mathrm{conv}}$	$P^{\mathrm{pl}}_{\mathrm{conv}}$
−2.0	−1.987	−2.310	.113	.325	.113	.421	1.000	1.000
−1.5	−1.563	−1.701	.138	.218	.142	.259	1.000	1.000
−1.0	−1.059	−1.137	.088	.130	.091	.148	1.000	1.000
−.5	−.536	−.586	.037	.043	.038	.050	1.000	1.000
0	−.018	−.050	.008	.008	.008	.011	1.000	1.000
.5	.223	.211	.041	.003	.117	.086	1.000	1.000
1.0	.205	.227	.042	.0004	.674	.598	1.000	1.000
1.5	.215	.234	.047	.0004	1.698	1.603	1.000	1.000
2.0	.200	.234	.050	.0004	3.290	3.121	1.000	1.000

Table 15. Properties of $\hat\sigma_{\mathrm{ml}}$ and $\hat\sigma_{\mathrm{pl}}$ for graphs with $n = 50, r = 70$.

σ	$E(\hat\sigma_{\mathrm{ml}})$	$E(\hat\sigma_{\mathrm{pl}})$	$V(\hat\sigma_{\mathrm{ml}})$	$V(\hat\sigma_{\mathrm{pl}})$	$MSE(\hat\sigma_{\mathrm{ml}})$	$MSE(\hat\sigma_{\mathrm{pl}})$	$P^{\mathrm{ml}}_{\mathrm{conv}}$	$P^{\mathrm{pl}}_{\mathrm{conv}}$
−2.0	−1.453	−2.167	.694	.261	.994	.289	1.000	1.000
−1.5	−1.497	−1.664	.156	.140	.156	.167	1.000	1.000
−1.0	−1.041	−1.145	.065	.090	.067	.112	1.000	1.000
−.5	−.526	−.579	.026	.035	.026	.041	1.000	1.000
0	−.012	−.025	.004	.004	.004	.005	1.000	1.000
.5	.167	.159	.018	.0004	.129	.116	.999	1.000
1.0	.161	.171	.020	.0004	.725	.687	.999	1.000
1.5	.160	.175	.019	.0004	1.814	1.757	1.000	1.000
2.0	.178	.175	.017	.0004	3.339	3.332	1.000	1.000

Table 16. Properties of $\hat\sigma_{\mathrm{ml}}$ and $\hat\sigma_{\mathrm{pl}}$ σ for graphs with $n = 60, r = 100$.

σ	$E(\hat\sigma_{\mathrm{ml}})$	$E(\hat\sigma_{\mathrm{pl}})$	$V(\hat\sigma_{\mathrm{ml}})$	$V(\hat\sigma_{\mathrm{pl}})$	$MSE(\hat\sigma_{\mathrm{ml}})$	$MSE(\hat\sigma_{\mathrm{pl}})$	$P^{\mathrm{ml}}_{\mathrm{conv}}$	$P^{\mathrm{pl}}_{\mathrm{conv}}$
−2.0	−1.834	−2.105	.339	.137	.367	.148	1.000	1.000
−1.5	−1.531	−1.573	.066	.078	.067	.083	1.000	1.000
−1.0	−1.030	−1.052	.034	.034	.035	.037	1.000	1.000
−.5	−.517	−.543	.013	.017	.013	.019	1.000	1.000
0	−.006	−.009	.002	.002	.002	.002	1.000	1.000
.5	.510	.094	.485	.000	.485	.165	1.000	.997
1.0	.661	.095	.644	.000	.759	.819	1.000	.986
1.5	.815	.095	.769	.000	1.238	1.973	1.000	.955
2.0	1.010	.095	.858	.000	1.839	3.629	1.000	.860

Table 17. Properties of $\hat\sigma_{\mathrm{ml}}$ and $\hat\sigma_{\mathrm{pl}}$ for graphs with $n = 100, r = 200$.

τ	$E(\hat{\tau}_{\text{ml}})$	$E(\hat{\tau}_{\text{pl}})$	$V(\hat{\tau}_{\text{ml}})$	$V(\hat{\tau}_{\text{pl}})$	$MSE(\hat{\tau}_{\text{ml}})$	$MSE(\hat{\tau}_{\text{pl}})$	$P^{\text{ml}}_{\text{conv}}$	$P^{\text{pl}}_{\text{conv}}$
−2.0	−1.523	−2.395	.214	.450	.442	.606	.611	.570
−1.5	−1.246	−2.075	.319	.693	.383	1.023	.889	.831
−1.0	−.896	−1.643	.404	.912	.414	1.326	.983	.935
−.5	−.500	−1.124	.460	1.034	.460	1.424	.998	.972
0	−.068	−.543	.475	1.005	.480	1.300	.999	.989
.5	.387	.066	.449	.834	.462	1.023	.985	.997
1.0	.856	.681	.413	.605	.433	.707	.905	.999
1.5	1.291	1.202	.314	.328	.358	.417	.707	.999
2.0	1.579	1.482	.167	.132	.344	.400	.487	.999

Table 18. Properties of $\hat{\tau}_{\text{ml}}$ and $\hat{\tau}_{\text{pl}}$ for graphs with $n = 7, r = 12$.