ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Computational Statistics & Data Analysis ▮▮▮ (▮▮▮▮) ▮▮▮–▮▮▮

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Markov models for digraph panel data: Monte Carlo-based derivative estimation

Michael Schweinberger*, [1], Tom A.B. Snijders

ICS, University of Groningen, Grote Rozenstraat 31, 9712 TG Groningen, The Netherlands

Received 10 June 2005; received in revised form 7 July 2006; accepted 10 July 2006

Abstract

A parametric, continuous-time Markov model for digraph panel data is considered. The parameter is estimated by the method of moments. A convenient method for estimating the variance–covariance matrix of the moment estimator relies on the delta method, requiring the Jacobian matrix—that is, the matrix of partial derivatives—of the estimating function. The Jacobian matrix was estimated hitherto by Monte Carlo methods based on finite differences. Three new Monte Carlo estimators of the Jacobian matrix are proposed, which are related to the likelihood ratio/score function method of derivative estimation and have theoretical and practical advantages compared to the finite differences method. Some light is shed on the practical performance of the methods by applying them in a situation where the true Jacobian matrix is known and in a situation where the true Jacobian matrix is unknown.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Digraphs; Continuous-time Markov process; Gradient estimation; Likelihood ratio/score function method; Variance reduction; Control variates

1. Introduction

The present paper considers digraph panel data, that is, data that can be represented by a directed relation (or digraph) on a set of nodes observed at two or more discrete time points. Social scientists use digraphs to study, among other things, informal relations (e.g., friendship) among individuals embedded in formal organizations (business firms, schools, etc.). Such data tend to display second- and third-order dependencies among observed arcs, so that modeling digraph panel data requires to take into account such dependencies.

A flexible approach to model such dependent data is based on the assumption that the observed digraphs are outcomes at some discrete time points of a Markov process evolving in continuous time, indexed by a parameter θ. The basic, underlying idea dates back to Holland and Leinhardt (1977) and Wasserman (1979, 1980), and was expanded by Snijders (2001) to model third- and higher-order dependencies.

Since the continuous-time Markov process is not observed in continuous time but at discrete time points, the likelihood function cannot be written in closed form and thus likelihood-based inference is hard. The method of moments was proposed by Snijders (2001) to estimate the parameter θ. A convenient method to estimate the variance–covariance

* Corresponding author. Tel.: +31 50 3636197; fax: +31 50 3636226.
E-mail address: M.Schweinberger@rug.nl (M. Schweinberger).

[1] Supported by grant 401-01-550 from the Netherlands Organisation for Scientific Research (NWO).

0167-9473/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2006.07.014

matrix of the moment estimator is based on the delta method, which requires the Jacobian matrix of the estimating function, that is, the matrix of first-order partial derivatives. Snijders (2001) estimated the Jacobian matrix by Monte Carlo methods based on finite differences with common random numbers.

The present paper proposes three new estimators of the Jacobian matrix, all of them related to the likelihood ratio/score function method of derivative estimation (Aleksandrov et al., 1968), and two of them utilizing variance reduction methods based on control variates. The three estimators have theoretical advantages compared to the finite differences method, but an important practical motivation for the estimators is that they roughly cut down the computational burden by a factor $L+1$, where L is the dimension of θ. The achieved reduction in computation time is of great practical value, since in practice computation time is an issue, and it is not unusual for L to be between 10 and 30.

The paper is structured as follows. The probabilistic framework is outlined in Section 2. The central argument is presented in Section 3. Section 4 compares the estimators of the Jacobian matrix in a situation where the true Jacobian matrix is known and in a situation where the true Jacobian matrix is unknown.

2. Probabilistic framework

It is assumed that a binary, directed relation $\longrightarrow$ on a finite set of nodes $\mathscr{N}=\{1,2,\ldots,n\}$ has been observed at discrete, ordered time points $t_0<t_1<\cdots<t_G$. These observations may be represented by digraphs and stored as binary matrices $x\left(t_0\right), x\left(t_1\right), \ldots, x\left(t_G\right)$, where element $x_{ij}\left(t_g\right)$ of $n\times n$ matrix $x\left(t_g\right)$ is defined as

$$x_{ij}\left(t_g\right)=\begin{cases}1 & \text{if } i\longrightarrow j \text{ at time point } t_g,\\ 0 & \text{otherwise,}\end{cases} \tag{1}$$

where $i\longrightarrow j$ means that node i is related to node j; the fact that the relation is directed means that $x_{ij}\left(t_g\right)$ may be different from $x_{ji}\left(t_g\right)$; the diagonal elements $x_{ii}\left(t_g\right)$ are regarded as structural zeros.

It is postulated that the observed digraphs are generated by an unobserved, continuous-time stochastic process. The discrete time points $t_0<\cdots<t_G$ are embedded in the time interval $[t_0,t_G]$. The digraph $x(t_0)$ observed at time point t_0 is not modeled, that is, the statistical modeling is done conditional on $x\left(t_0\right)$. Consider the case $G=1$.

A simple process can be constructed by assuming that the process is a Markov process. Then the model is specified by the generator of the Markov process, which corresponds to a $W\times W$ matrix Q_θ indexed by a parameter θ, where $W=2^{n(n-1)}$ is the number of digraphs on $\mathscr{N}$. The elements $q_\theta\left(x^\star,x\right)$ of generator Q_θ are the rates of moving from digraph $x^\star$ to digraph x.

Let $x^\star$ and x be two arbitrary digraphs on $\mathscr{N}$. If x deviates from $x^\star$ in more than one arc variable $x_{ij}^\star$, then Snijders (2001) assumes that $q_\theta\left(x^\star,x\right)=0$; in other words, the process moves forward by changing not more than one arc variable $x_{ij}^\star$ at the time. Let $x^\star$ be an arbitrary digraph on $\mathscr{N}$, and let x be the digraph that is obtained from $x^\star$ by changing one and only one specified arc variable, say $x_{ij}^\star$. Since the transition from $x^\star$ to x involves only the ordered pair of nodes (i,j), one can rewrite $q_\theta\left(x^\star,x\right)$ as $q_\theta\left(x^\star,i,j\right)$ and decompose $q_\theta\left(x^\star,i,j\right)$ as follows:

$$q_\theta\left(x^\star,i,j\right)=\lambda_i\left(\theta,x^\star\right)r_i(\theta,x,j), \tag{2}$$

where

$$\lambda_i\left(\theta,x^\star\right)=\sum_{h\neq i}^{n}q_\theta\left(x^\star,i,h\right) \tag{3}$$

and

$$r_i(\theta,x,j)=\frac{q_\theta\left(x^\star,i,j\right)}{\lambda_i\left(\theta,x^\star\right)}. \tag{4}$$

The interpretation is that λ_i—called the rate function—is the rate at which the set $\left\{x_{ij}^\star : j\neq i\in\mathscr{N}\right\}$ of arcs emanating from node i is changed, while r_i gives the conditional probabilities of such changes.

A simple specification of the rate function is

$$\lambda_i\left(\theta,x^\star\right)=\rho, \tag{5}$$

Please cite this article as: Michael Schweinberger, Tom A.B. Snijders, Markov models for digraph panel data: Monte Carlo-based derivative estimation, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.014

that is, as constant across nodes i and digraphs $x^\star$, equal to some rate parameter ρ. A more general specification of the rate function is given by

$$\lambda_i\left(\theta, x^\star\right) = \rho \exp\left[\alpha' a_i\left(x^\star, c_i\right)\right], \tag{6}$$

where $\alpha = (\alpha_k)$ is a vector valued parameter and $a_i = (a_{ik})$ is a vector valued function of covariates c_i depending on node i and graph-dependent statistics involving the arcs of node i. When there is more than one time interval, that is, $G > 1$, then parameter ρ can be made dependent on time interval $\left[t_{g-1}, t_g\right]$.

A convenient, multinomial logit parametrization of $r_i(\theta, x, j)$ is given by

$$r_i(\theta, x, j) = \frac{\exp\left[f_i(\beta, x, j)\right]}{\sum_{h \neq i}^{n} \exp\left[f_i(\beta, x, h)\right]}, \tag{7}$$

where

$$f_i(\beta, x, j) = \beta' s_i(x, j), \tag{8}$$

while $\beta = \left(\beta_k\right)$ is a vector valued parameter and $s_i = (s_{ik})$ is a vector valued statistic. The function f_i is called the objective function. Examples of statistics s_{ik} are the number of arcs $\sum_{h=1}^{n} x_{ih}$, the number of transitive triplets $\sum_{h,l=1}^{n} x_{ih} x_{hl} x_{il}$, or other statistics involving the arcs of node i and covariates; such statistics can be used to define third- and higher-order dependencies.

Remark 2.1 (*Simulation of the Markov process*). The methods proposed in Section 3 rely on Monte Carlo simulation of the Markov process. Therefore, it is worthwhile to consider how the Markov process can be simulated. Let $\text{Exp}(\psi)$ be the negative exponential distribution with parameter ψ. The Markov process can be simulated in time interval $[t_0, t_1]$ conditional on digraph $x\left(t_0\right)$ observed at time point t_0 by starting at time point $t = t_0$ at digraph $x^{(0)} = x\left(t_0\right)$ and iterating the following steps (with initial value $M = 0$):

```
Increment $M$.
(1) Increment $t$:
    — Sample $h_M \sim \text{Exp}\left(\sum_{k=1}^{n} \lambda_k\left(\theta, x^{(M-1)}\right)\right)$.
    — Set $t = t + h_M$.
(2) If $t < t_1$ then:
    — Sample node $i$ with probability $\lambda_i\left(\theta, x^{(M-1)}\right) / \sum_{k=1}^{n} \lambda_k\left(\theta, x^{(M-1)}\right)$.
    — Given $i$, sample node $j$ with probability $r_i\left(\theta, x^{(M)}, j\right)$ and
      set $x_{ij}^{(M)} = 1 - x_{ij}^{(M-1)}$ and $x_{kl}^{(M)} = x_{kl}^{(M-1)}$ for all $(k, l) \neq (i, j)$.
Else: terminate.
```

Remark 2.2 (*Extensions*). Extensions to $G > 1$ time intervals are straightforward due to the Markov property. It is furthermore possible to deal with the co-evolution of digraphs and other outcome variables (Snijders et al., 2006).

3. Derivative estimation

The present section begins by briefly discussing the estimation of the parameter vector θ and its variance–covariance matrix, which requires the knowledge of the Jacobian matrix of the estimating function. The problem is that the Jacobian matrix of the estimating function cannot be written in closed form and hence must be estimated. Section 3.1 outlines a conventional Monte Carlo estimator of the Jacobian matrix based on finite differences, while in Section 3.2 three new Monte Carlo estimators are proposed, based on the likelihood ratio/score function method. Attention is restricted to the case of $G = 1$ time intervals; the extension to the case $G > 1$ is immediate.

Snijders (2001) proposed to estimate the parameter $\theta \in \Theta \subset \mathscr{R}^L$ by the method of moments (Pearson, 1902a,b). Let $U\left(t_1\right)$ be a suitable $L \times 1$ vector function (statistic) of the digraph and covariates at time point t_1 with observed value $u\left(t_1\right)$; the function $U\left(t_1\right)$ may in addition depend on the digraph $x\left(t_0\right)$ and covariates observed at time point t_0. Let $\hat{\theta}$ be the solution of the moment equation

$$E_\theta\left[U\left(t_1\right) \mid X\left(t_0\right) = x\left(t_0\right)\right] = u\left(t_1\right). \tag{9}$$

Please cite this article as: Michael Schweinberger, Tom A.B. Snijders, Markov models for digraph panel data: Monte Carlo-based derivative estimation, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.014

In the sequel, the left-hand side of (9) is referred to as $E_\theta U$. The estimation problem amounts to finding a root of $E_\theta U - u$ as a function of θ, where $u = u(t_1)$. Experience suggests that, for suitable statistics U, Eq. (9) has a unique root in almost all cases. The moment estimator $\hat{\theta}$ is not available in closed form. However, it is possible to simulate the Markov process so that root-finding algorithms based on stochastic approximation (Robbins and Monro, 1951) can be used to find $\hat{\theta}$ (see Snijders, 2001).

To estimate the variance–covariance matrix of $\hat{\theta}$, it is inconvenient to use bootstrap (or resampling) methods, because each of the multiple estimation runs required by resampling methods is time-consuming; an alternative is to utilize the delta method (Lehmann, 1999, p. 315) and the implicit function theorem, giving the approximation

$$\mathrm{Cov}_\theta\, \hat{\theta} \approx \Delta^{-1}(\theta)\Sigma(\theta)\left[\Delta^{-1}(\theta)\right]', \tag{10}$$

where

$$\Sigma(\theta) = \mathrm{Cov}_\theta\, U \tag{11}$$

is the $L \times L$ variance–covariance matrix of U, and

$$\Delta(\theta) = \frac{\partial}{\partial \theta'} E_\theta U \tag{12}$$

is the Jacobian matrix of $E_\theta U$ at θ, that is, the $L \times L$ matrix of first-order partial derivatives of $E_\theta U$ evaluated at θ. Here again, no closed form expressions are available.

The variance–covariance matrix (10) can be estimated by plugging in the moment estimator $\hat{\theta}$ in (11) and (12). Monte Carlo estimation of $\Sigma\left(\hat{\theta}\right)$ is straightforward; the issue is how to construct a Monte Carlo estimator of $\Delta\left(\hat{\theta}\right)$. In Section 3.1, a conventional Monte Carlo estimator of $\Delta\left(\hat{\theta}\right)$ based on finite differences is described, whereas in Section 3.2 three new Monte Carlo estimators are proposed, based on the likelihood ratio/score function method.

3.1. *Finite differences method*

By definition,

$$\Delta_l(\theta) = \lim_{\varepsilon \longrightarrow 0} \frac{E_{\theta + e_l \varepsilon} U - E_\theta U}{\varepsilon}, \quad l = 1, \ldots, L, \tag{13}$$

where $\Delta_l(\theta)$ refers to the lth column of $\Delta(\theta)$. The finite differences method to estimate $\Delta_l(\theta)$ is based on

$$\Delta_{l,\varepsilon}(\theta) = \frac{E_{\theta + e_l \varepsilon} U - E_\theta U}{\varepsilon}, \quad l = 1, \ldots, L. \tag{14}$$

The expectations $E_{\theta + e_l \varepsilon} U$, $l = 1, \ldots, L$, and $E_\theta U$ are not available in closed form, but can be estimated by the corresponding Monte Carlo sample averages: given $\theta + e_l \varepsilon$ and a pseudo-random number generator (see, e.g., Marsaglia and Zaman, 1991), one can simulate the Markov process as described in Remark 2.1 of Section 2; having simulated the Markov process multiple times, the Monte Carlo sample average of U can be used as an estimate of $E_{\theta + e_l \varepsilon} U$; the expectation $E_\theta U$ can be estimated accordingly by simulating the Markov process multiple times given θ and using the Monte Carlo sample average of U as an estimate of $E_\theta U$. Thus, to estimate the $L + 1$ expectations, $L + 1$ Monte Carlo samples are required, because the parameters are different.

It is well-known (see, e.g., L'Ecuyer, 1991) that the resulting estimator is biased. Under regularity conditions, the bias is of order ε, which suggests to make ε as small as possible. On the other hand, it is clear from (14) that, when using independent draws of U under the distributions corresponding to $\theta + e_l \varepsilon$ and to θ, the variance of the resulting estimator is of order ε^{-2}, which implies that small values of ε are undesirable. A way out of this dilemma is provided by using common random numbers (Hammersley and Handscomb, 1964, pp. 48–49) for simulating the random variable U under the distributions corresponding to $\theta + e_l \varepsilon$ and to θ. Denoting by W the random number stream and by $U_\theta(W)$ the result of the simulation procedure as a function of W and θ, this means that the same W is

Please cite this article as: Michael Schweinberger, Tom A.B. Snijders, Markov models for digraph panel data: Monte Carlo-based derivative estimation, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.014

used for generating U under $\theta + e_l\varepsilon$ and under θ, so that the random variable used for generating a value of $\Delta_{l,\varepsilon}(\theta)$ is given by

$$\frac{U_{\theta+e_l\varepsilon}(W) - U_\theta(W)}{\varepsilon}. \tag{15}$$

If $U_\theta(w)$ would be a continuously differentiable function of θ_l for any given w, then under regularity conditions the random variable (15) would tend to the derivative $\partial U_\theta(w)/\partial\theta_l$, its variance would be bounded for $\varepsilon \longrightarrow 0$, and ε could be taken very small to get a finite differences estimator (15) which is practically unbiased and N-consistent, where N is the size of the Monte Carlo sample. However, the discrete nature of the outcome variable U in the considered model implies that $U_\theta(w)$ is a discontinuous function of θ. The following lemma can be used to determine the order of magnitude of the variance of (15).

Lemma 1. *Let $D(\varepsilon)$ be a random variable with a finite outcome space that does not depend on ε, let $d_\varepsilon = E\,D(\varepsilon)/\varepsilon$, and suppose that $d = \lim_{\varepsilon\longrightarrow 0} d_\varepsilon$ is finite and non-zero. Then*

$$\liminf_{\varepsilon\longrightarrow 0}\ |\varepsilon|\ \mathrm{Var}\left(\frac{D(\varepsilon)}{\varepsilon}\right) > 0. \tag{16}$$

Proof. Let d_0 be the smallest non-zero outcome of $|D(\varepsilon)|$. Then $D^2(\varepsilon) \geqslant d_0|D(\varepsilon)|$ and

$$E\,\frac{D^2(\varepsilon)}{\varepsilon^2} \geqslant \frac{d_0}{|\varepsilon|}E\,\frac{|D(\varepsilon)|}{|\varepsilon|}. \tag{17}$$

From $\mathrm{Var}(D(\varepsilon)/\varepsilon) = E\left[D^2(\varepsilon)/\varepsilon^2\right] - d_\varepsilon^2$ it follows that

$$\liminf_{\varepsilon\longrightarrow 0}\left\{|\varepsilon|\ \mathrm{Var}\left(\frac{D(\varepsilon)}{\varepsilon}\right)\right\} \geqslant \liminf_{\varepsilon\longrightarrow 0}\left\{d_0 E\,\frac{|D(\varepsilon)|}{|\varepsilon|} - |\varepsilon| d_\varepsilon^2\right\} \geqslant d_0|d| > 0. \quad \square \tag{18}$$

Applying Lemma 1 to $D(\varepsilon) = U_{\theta+e_l\varepsilon}(W) - U_\theta(W)$ shows that if the derivative to be estimated is non-zero, the variance of (15) is at least of order ε^{-1}. Note that if the derivative is zero, then $E_\theta U$ is not sensitive to changes in θ, and therefore U is not a sensible choice for estimating θ in a method of moments framework (see Snijders, 2001).

In the remainder of the paper, $D_0(\varepsilon)$ refers to the Monte Carlo finite differences estimator with common random numbers which estimates the columns $\Delta_l(\theta)$ of $\Delta(\theta)$ by $\hat{\Delta}_{l,\varepsilon}(\theta)$, where $\hat{\Delta}_{l,\varepsilon}(\theta)$ is the Monte Carlo estimator of $\Delta_{l,\varepsilon}(\theta)$ obtained by replacing the expectations $E_{\theta+e_l\varepsilon}U$ and $E_\theta U$ by the corresponding Monte Carlo sample averages. An important practical implication of using $D_0(\varepsilon)$ is that $L+1$ Monte Carlo samples are required. Section 3.2 considers an alternative method that produces unbiased and N-consistent estimators, in contrast to $D_0(\varepsilon)$, and requires only one Monte Carlo sample.

3.2. Likelihood ratio/score function method

The alternative method is related to the likelihood ratio/score function method of derivative estimation, which can be traced back to Aleksandrov et al. (1968). Some related papers are Rubinstein (1986, 1989) and Glynn and L'Ecuyer (1995).

Denote the complete data—that is, the holding times of the Markov process and the sequence of arc changes in time interval $[t_0, t_1]$—by Z. Let P_θ be the probability law governing Z, admitting a probability density $p_\theta = \mathrm{d}P_\theta/\mathrm{d}\mu$ with respect to some dominating measure μ, and let $Z_1, Z_2, \ldots$ be Monte Carlo generated random variables with distribution P_θ.

Three likelihood ratio/score function (LR) estimators of the Jacobian matrix $\Delta(\theta)$ are derived below, called D_{I}, D_{II}, and D_{III}; the dependence of the D-estimators on θ is left implicit. The LR estimator D_{I} (Section 3.2.1) is the basic LR estimator, while LR estimators D_{II} and D_{III} (Section 3.2.2) use variance reduction methods and have less Monte Carlo variance than D_{I}.

3.2.1. Estimator D_{I}: the basic LR estimator

Let

$$D_{\mathrm{I}} = \frac{1}{N}\sum_{i=1}^{N} U_i \frac{\partial \ln p_\theta(Z_i)}{\partial \theta'}, \tag{19}$$

where N is the size of the Monte Carlo sample.

Lemma 2. *Let U be any function of $x(t_0)$, $x(t_1)$, and covariates that does not depend on θ. Then D_{I} is an unbiased and N-consistent estimator of the Jacobian matrix $\Delta(\theta)$ defined in* (12).

Proof. By definition,

$$\Delta_l(\theta) = \lim_{\varepsilon \longrightarrow 0} \frac{E_{\theta+e_l\varepsilon}U - E_\theta U}{\varepsilon} = \lim_{\varepsilon \longrightarrow 0} \int_{\mathscr{Z}} U \frac{p_{\theta+e_l\varepsilon}(z) - p_\theta(z)}{\varepsilon}\,\mathrm{d}\mu(z). \tag{20}$$

The model definition of Section 2 implies that the model for Z is a family of negative exponential-multinomial distributions, which is an exponential family of distributions. The outcome space of U is finite, so that $E_\theta U$ exists for all $\theta \in \Theta$ and is finite. Hence, by Theorem 2.7.1 of Lehmann and Romano (2005, p. 49), it is admissible to interchange the order of differentiation and integration:

$$\lim_{\varepsilon \longrightarrow 0} \int_{\mathscr{Z}} U \frac{p_{\theta+e_l\varepsilon}(z) - p_\theta(z)}{\varepsilon}\,\mathrm{d}\mu(z) = \int_{\mathscr{Z}} U \lim_{\varepsilon \longrightarrow 0} \frac{p_{\theta+e_l\varepsilon}(z) - p_\theta(z)}{\varepsilon}\,\mathrm{d}\mu(z). \tag{21}$$

Thus, the Jacobian matrix can be written as

$$\Delta(\theta) = \int_{\mathscr{Z}} U \frac{\partial p_\theta(z)}{\partial \theta'}\,\mathrm{d}\mu(z) = E_\theta\left[U \frac{\partial \ln p_\theta(Z)}{\partial \theta'}\right]. \tag{22}$$

Eq. (22) proves the unbiasedness of D_{I} for estimating $\Delta(\theta)$. N-consistency follows from the strong law of large numbers (see, e.g., Ferguson, 1996, p. 21). □

Since in practice the aim is to evaluate the Jacobian matrix $\Delta(\theta)$ at the moment estimate $\hat{\theta}$ of θ, $\hat{\theta}$ is plugged in for θ. Given $\hat{\theta}$, the Monte Carlo generation of random variables Z_i $(i=1, 2, \ldots, N)$ with probability law $P_{\hat{\theta}}$ is straightforward; see Remark 2.1 of Section 2. The complete-data efficient score $\partial \ln p_{\hat{\theta}}(Z_i)/\partial\hat{\theta}'$ is derived in Appendix A.

The Monte Carlo variance of the basic LR estimator D_{I} may be too large for practical purposes. It is therefore sensible to reduce the variance of D_{I} by using variance reduction methods, which is explored in Section 3.2.2.

3.2.2. Estimators D_{II} and D_{III}: LR estimators exploiting variance reduction methods based on control variates

In the present section, two LR estimators are proposed which have less variance than D_{I}. Both estimators are based on the idea of reducing the variance of LR estimators by using the complete-data efficient score as a control variate (see Fieller and Hartley, 1954; Rubinstein, 1986, 1989).

It will be convenient to let

$$S'_\theta = S'_\theta(Z) = \frac{\partial \ln p_\theta(Z)}{\partial \theta'}, \tag{23}$$

and to rewrite the Jacobian matrix $\Delta(\theta)$ given by (22) using the vec operator as

$$\operatorname{vec} \Delta(\theta) = E_\theta\left[\operatorname{vec}\left(U S'_\theta\right)\right]. \tag{24}$$

Let A be any non-random matrix of order $L^2 \times L$. Observe that $E_\theta S_\theta = 0$ and that

$$\begin{aligned}\operatorname{vec} \Delta(\theta) &= E_\theta\left[\operatorname{vec}\left(U S'_\theta\right)\right] = E_\theta\left[\operatorname{vec}\left(U S'_\theta\right) - A\left(S_\theta - E_\theta S_\theta\right)\right] \\ &= E_\theta\left[\operatorname{vec}\left(U S'_\theta\right) - A S_\theta\right].\end{aligned} \tag{25}$$

Please cite this article as: Michael Schweinberger, Tom A.B. Snijders, Markov models for digraph panel data: Monte Carlo-based derivative estimation, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.014

Eq. (25) suggests that (24) can be estimated by

$$D_C(A) = \frac{1}{N}\sum_{i=1}^{N}\left[\mathrm{vec}\left(U_i S'_{\theta,i}\right) - A S_{\theta,i}\right] = \mathrm{vec}\left(D_{\mathrm{I}}\right) - A\,\frac{1}{N}\sum_{i=1}^{N} S_{\theta,i}. \tag{26}$$

If A is constant, then $D_C(A)$ as an estimator of (24) is both unbiased and N-consistent.

The idea of control variates is to exploit the fact that $E_\theta S_\theta$ and hence the Monte Carlo integration error $(1/N)\sum_{i=1}^{N} S_{\theta,i} - E_\theta S_\theta$ is known; let $L=1$ so that $U S_\theta$ and S_θ are scalars; if $U S_\theta$ and S_θ are correlated, the knowledge of the integration error can be used to (linearly) transform D_{I} such that D_{I} gets closer to its expectation, resulting in variance reduction. Two choices of A, and hence two control variate estimators, are elaborated below.

3.2.2.1. Estimator D_{II}: heuristic LR control variate estimator A simple, heuristic control variate estimator is obtained as follows. Let

$$U^\star = U - E_\theta U, \tag{27}$$

and observe that at the moment estimate $\hat\theta$ of θ, $E_\theta U$ is known and given by $E_{\hat\theta} U = u$, where u is the observed value of statistic U. The centering (27) is equivalent to using S_θ as a control variate with $A = I_L \otimes u$, where I_L is the $L \times L$ identity matrix, which is evident from the identity

$$\mathrm{vec}\left(U^\star S'_\theta\right) = \mathrm{vec}\left(U S'_\theta\right) - \mathrm{vec}\left(u S'_\theta\right) \tag{28}$$

and the fact that $\mathrm{vec}\left(u S'_\theta\right)$ can be written as

$$\mathrm{vec}\left(u S'_\theta\right) = \mathrm{vec}\left(u S'_\theta I_L\right) = \left(I_L \otimes u\right)\mathrm{vec}\left(S'_\theta\right) = A S_\theta. \tag{29}$$

The resulting estimator, which is (26) with $A = I_L \otimes u$, is denoted by D_{II}:

$$D_{\mathrm{II}} = D_C\left(I_L \otimes u\right). \tag{30}$$

Since $A = I_L \otimes u$ is constant, D_{II} as an estimator of (24) is both unbiased and N-consistent.

To give some insight into the behavior of estimator D_{II}, let $L=1$ so that U, S_θ, D_{I}, and D_{II} are scalars. Using $E_\theta\left[U^\star S_\theta\right] = E_\theta\left[U S_\theta\right]$,

$$\begin{aligned}\mathrm{Var}_\theta\left(D_{\mathrm{I}}\right) - \mathrm{Var}_\theta\left(D_{\mathrm{II}}\right) &= \frac{E_\theta\left[\left(U S_\theta\right)^2\right] - E_\theta\left[\left(U^\star S_\theta\right)^2\right]}{N}\\ &= E_\theta\left[\left(\frac{2U}{u} - 1\right) u^2 S_\theta^2\right] \Big/ N.\end{aligned} \tag{31}$$

Thus, under the (unrealistic) assumption $P_\theta(U/u \geqslant \frac{1}{2}) = 1$, the right-hand side of (31) is non-negative and $\mathrm{Var}_\theta\left(D_{\mathrm{I}}\right) \geqslant \mathrm{Var}_\theta\left(D_{\mathrm{II}}\right)$. In practice, (31) suggests that if the ratio of the standard deviation of U to the absolute value of the expectation $E_{\hat\theta} U = u$ is less than, say, $\frac{1}{4}$, then the variance of D_{II} is probably much smaller than the variance of D_{I}; by Chebyshev's inequality, a more conservative guess is that variance reduction is very likely if the ratio is less than $\frac{1}{10}$. Furthermore, when this ratio is close to 0 and thus U is "almost constant", the right-hand side of (31) is roughly $u^2\,\mathrm{Var}_\theta\left(S_\theta\right)/N$.

However, while the variance of D_{II} may be considerably smaller than the variance of D_{I}, its variance may not be the minimum variance that can be achieved by using S_θ as a control variate. The estimator proposed below attains the minimum variance by construction.

3.2.2.2. Estimator D_{III}: minimum variance LR control variate estimator Let $\mathrm{Cov}_\theta\left[D_C(A)\right]$ be the variance–covariance matrix of the vector valued LR control variate estimator $D_C(A)$ as given by (26), and let $\left|\mathrm{Cov}_\theta\left[D_C(A)\right]\right|$ be the determinant of $\mathrm{Cov}_\theta\left[D_C(A)\right]$, called the generalized variance of $D_C(A)$.

Lemma 3 (Rubinstein and Marcus, 1985). *The value of A that minimizes the generalized variance of $D_C(A)$ is given by*

$$B = \Gamma_{21}(\theta)\Gamma_{11}^{-1}(\theta), \tag{32}$$

where

$$\Gamma_{11}(\theta) = E_\theta\left[S_\theta S_\theta'\right] \tag{33}$$

is the variance-covariance matrix of S_θ, and

$$\Gamma_{21}(\theta) = E_\theta\left[\left\{\mathrm{vec}\left(U S_\theta'\right) - E_\theta\left[\mathrm{vec}\left(U S_\theta'\right)\right]\right\} S_\theta'\right] \tag{34}$$

is the covariance matrix of $\mathrm{vec}\left(U S_\theta'\right)$ and S_θ.

Proof. See Rubinstein and Marcus (1985). □

Let $L=1$ so that S_θ, US_θ, D_{I}, and $D_C(B)$ are scalars, where $D_C(B)$ is given by (26) with $A=B$. Then the variance of $D_C(B)$ is equal to $(1-\varrho^2)\,\mathrm{Var}_\theta(D_{\mathrm{I}})$, where ϱ is the correlation between US_θ and S_θ: thus, the higher the absolute correlation between US_θ and S_θ, the greater is the reduction in variance. A similar argument for $L>1$ follows from Rubinstein and Marcus (1985).

The matrix B can be estimated by

$$\hat{B} = V_{21}V_{11}^{-1}, \tag{35}$$

where V_{11} and V_{21} are Monte Carlo estimators of $\Gamma_{11}(\theta)$ and $\Gamma_{21}(\theta)$, respectively, estimated from the Monte Carlo sample $Z_1, Z_2, \ldots, Z_N$.

Let D_{III} be (26) with $A=\hat{B}$:

$$D_{\mathrm{III}} = D_C\left(\hat{B}\right). \tag{36}$$

The estimator D_{III} is N-consistent, but because $A=\hat{B}$ depends on S_θ, D_{III} is not unbiased (Fishman, 1996, p. 279); however, the bias is of order N^{-1} (cf. Cochran, 1977, pp. 198–199). □

3.2.3. *Comparison: estimators $D_0(\varepsilon)$, D_{I}, D_{II}, and D_{III}*

It is evident from Lemma 2 that D_{I} and D_{II} are unbiased and N-consistent, in contrast to $D_0(\varepsilon)$, while D_{III} is an N-consistent estimator whose bias is of order N^{-1}.

To demonstrate how the LR estimators D_{I}, D_{II}, and D_{III} are interrelated, let $L=1$. It was argued that D_{II} may be superior to D_{I} in terms of variance, and that if the standard deviation of U is small relative to the absolute value of the expectation $E_{\hat{\theta}}U=u$ and hence U is "almost constant", then $\mathrm{Var}_\theta\left(D_{\mathrm{II}}\right) \approx \mathrm{Var}_\theta\left(D_{\mathrm{I}}\right) - u^2\,\mathrm{Var}_\theta\left(S_\theta\right)/N$. For large N, $\hat{B}$ is expected to be close to B as given by (32) and hence the variance of D_{III} is roughly $\left(1-\varrho^2\right)\mathrm{Var}_\theta\left(D_{\mathrm{I}}\right)$, where ϱ is the correlation between US_θ and S_θ. Furthermore, if the standard deviation of U is small relative to $\left|E_{\hat{\theta}}U\right|$, then (33) and (34) imply that $\Gamma_{21} \approx u\Gamma_{11}$ and thus $B=\Gamma_{21}\Gamma_{11}^{-1} \approx u$, suggesting that $D_{\mathrm{II}} \approx D_{\mathrm{III}}$ and that the variance of D_{II} and D_{III} is of a similar order of magnitude. In sum, if the standard deviation of U is small relative to $\left|E_{\hat{\theta}}U\right|$, then it is thought that D_{II} and D_{III} are close in terms of variance and outperform D_{I}.

An important practical motivation for considering D_{I}, D_{II}, and in particular D_{III} as alternative estimators of the Jacobian matrix $\Delta(\theta)$ is that using one of the LR estimators instead of $D_0(\varepsilon)$ roughly cuts down the computation time by a factor $L+1$, where L is the dimension of θ.

4. Applications

In the present section, the derivative estimators $D_0(\varepsilon)$, D_{I}, D_{II}, and D_{III} are compared in a situation where the true Jacobian matrix is known (Section 4.1) and in addition in the common situation where the true Jacobian matrix is unknown (Section 4.2).

Please cite this article as: Michael Schweinberger, Tom A.B. Snijders, Markov models for digraph panel data: Monte Carlo-based derivative estimation, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.014

In each subsection, one real-world data set is studied, the moment estimate $\hat{\theta}$ of θ is obtained, and the Jacobian matrix $\varDelta\left(\hat{\theta}\right)$ is estimated from 1000 Monte Carlo samples of size $N = 1000$, where N corresponds to the number of terms on which the derivative estimators—which are averages—are based, implying that for D_{I}, D_{II}, and D_{III} the Markov process is N times simulated, while for $D_0(\varepsilon)$ the Markov process is $(L+1) \times N$ times simulated, because each term requires $L+1$ simulations.

4.1. Application: Jacobian matrix known

A simple, classical model where Jacobian matrices can be derived analytically is the independent arcs (IA) model (see Snijders and Van Duijn, 1997), which is in most empirical applications inadequate because of its simplicity, but provides an opportunity to compare the simulation-based derivative estimators.

The IA model is a continuous-time Markov model, where the rate of change $q_\theta\left(x^\star, i, j\right)$ depends only on $x^\star_{ij}$, so that all arc variables $x_{ij}(t)$ follow independent Markov processes. To keep the parametrization consistent with the general model of Section 2, the rate of change is written as

$$q_\theta\left(x^\star, i, j\right) = \theta_1 \frac{\left(1 - x^\star_{ij}\right)\exp\left(\theta_2\right) + x^\star_{ij}\exp\left(-\theta_2\right)}{n-1}. \tag{37}$$

The rate function λ_i follows from (3) and (37), and is given by

$$\lambda_i\left(\theta, x^\star\right) = \theta_1 \frac{\left((n-1) - x^\star_{i+}\right)\exp\left(\theta_2\right) + x^\star_{i+}\exp\left(-\theta_2\right)}{n-1}, \tag{38}$$

where $x^\star_{i+} = \sum_{h \neq i}^{n} x^\star_{ih}$. The conditional probability mass function r_i is, using (4), (37), and (38), given by

$$r_i(\theta, x, j) = \frac{\left(1 - x^\star_{ij}\right)\exp\left(\theta_2\right) + x^\star_{ij}\exp\left(-\theta_2\right)}{\left((n-1) - x^\star_{i+}\right)\exp\left(\theta_2\right) + x^\star_{i+}\exp\left(-\theta_2\right)}. \tag{39}$$

Let

$$M_{kl} = \#\left\{(i,j) \mid x_{ij}\left(t_0\right) = k, x_{ij}\left(t_1\right) = l\right\}, \tag{40}$$

where $k, l = 0, 1$. The statistic $U = \left(M_{01} + M_{10}, M_{01} + M_{11}\right)'$ is a sufficient statistic, and thus, to estimate $\theta = \left(\theta_1, \theta_2\right)'$ by the method of moments, $E_\theta U - u$ is a natural estimating function; note that, in the case of the IA model, the sufficiency implies that the moment estimator based on estimating equation $E_\theta U - u = 0$ coincides with the maximum likelihood (ML) estimator (see Snijders and Van Duijn, 1997). The expectation $E_\theta U$, the variance–covariance matrix $\Sigma(\theta) = \mathrm{Cov}_\theta\, U$, and the Jacobian matrix $\varDelta(\theta) = \partial E_\theta U / \partial\theta'$ are derived analytically in Appendix C.

Snijders and Van Duijn (1997) applied the IA model to a well-known data set called the EIES data, corresponding to the "communication" among $n = 32$ scholars observed at two time points, where $x_{ij} = 1$ if scholar i met scholar j, and $x_{ij} = 0$ otherwise. The moment (and ML) estimate of $\theta = \left(\theta_1, \theta_2\right)'$ is $\hat{\theta} = (2.418, 1.557)'$, the expectation of U is $E_{\hat{\theta}} U = (154, 653)'$, and, using the results of Appendix C, the exact values of $\Sigma\left(\hat{\theta}\right)$ and $\varDelta\left(\hat{\theta}\right)$ are given by

$$\Sigma\left(\hat{\theta}\right) = \begin{pmatrix} 108.80 & 95.00 \\ 95.00 & 108.80 \end{pmatrix} \quad \text{and} \quad \varDelta\left(\hat{\theta}\right) = \begin{pmatrix} 52.18 & 114.55 \\ 47.44 & 130.85 \end{pmatrix}. \tag{41}$$

The Jacobian matrix $\varDelta\left(\hat{\theta}\right)$ is estimated from each Monte Carlo sample by $D_0(\varepsilon)$ with $\varepsilon = .2$ and $\varepsilon = 1.0$; the two values of ε are motivated by the fact that most values of ε used in practice are between .2 and 1.0. In addition, $\varDelta\left(\hat{\theta}\right)$ is estimated from each Monte Carlo sample by D_{I}, D_{II}, and D_{III}; the complete-data efficient score, on which D_{I}, D_{II}, and D_{III} are based, is derived in Appendix B.

The computation time required to evaluate $D_0(.2)$ and $D_0(1.0)$ was on average approximately 20 s on a PC with Intel Pentium 3.06 GHz processor and 1021 MB RAM. While 20 s are negligible, note that in general the computation

Table 1
EIES data: average estimates of $\Delta\left(\hat{\theta}\right) = (\delta_{ij})$ across 1000 Monte Carlo samples

	True value	$D_0(.2)$	$D_0(1.0)$	D_{I}	D_{II}	D_{III}
δ_{11}	52.18	51.34	48.22	52.60	52.03	51.92
δ_{21}	47.44	46.67	43.83	49.72	47.30	47.21
δ_{12}	114.55	122.98	151.00	115.42	114.29	114.06
δ_{22}	130.85	137.76	161.08	135.43	130.67	130.41

Table 2
EIES data: Monte Carlo standard deviations of estimates of $\Delta\left(\hat{\theta}\right) = (\delta_{ij})$ based on 1000 Monte Carlo samples

	$D_0(.2)$	$D_0(1.0)$	D_{I}	D_{II}	D_{III}
δ_{11}	.53	.21	26.36	2.39	2.40
δ_{21}	.53	.22	111.36	2.28	2.28
δ_{12}	.87	.32	62.27	5.53	5.54
δ_{22}	1.01	.33	262.31	5.84	5.87

time is roughly proportional to $N \times L^2 \times C_1 \times C_2$, where $C_1 = \sum_{g=1}^{G} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \left|x_{ij}\left(t_g\right) - x_{ij}\left(t_{g-1}\right)\right|$ and $C_2 = (1/G)\sum_{g=0}^{G-1} \sum_{i=1}^{n} \sum_{j\neq i}^{n} x_{ij}\left(t_g\right)$. Here, $L = 2$ and $n = 32$, which are very small values: in practice, it is frequently the case that $L > 10$ and $n > 50$ (in fact, n may be in the hundreds), and then computation time is an issue and the computational advantage of D_{I}, D_{II}, and D_{III} over $D_0(\varepsilon)$ is appreciated.

The average estimates of the Jacobian matrix $\Delta\left(\hat{\theta}\right)$ are presented in Table 1. The average estimate of $\Delta\left(\hat{\theta}\right)$ based on the biased estimator $D_0(.2)$ is close to the true value for the first column, but overestimates the elements of the second column by 7.4% and 5.3%, respectively; the average estimate of $D_0(1.0)$ underestimates the elements of the first column by roughly 7.6%, and overestimates the elements of the second column by 31.8% and 23.1%, respectively. While the bias of $D_0(.2)$ may be tolerable, the bias of $D_0(1.0)$ clearly is not. The average estimates of $\Delta\left(\hat{\theta}\right)$ based on D_{II} and D_{III} are very close to the true value of $\Delta\left(\hat{\theta}\right)$, but D_{I}, which is known to be unbiased, seems to overestimate the elements of the second row of $\Delta\left(\hat{\theta}\right)$; this is an inaccuracy stemming from the huge variance of D_{I} (see below).

The Monte Carlo standard deviations (MC SDs) of the estimates of $\Delta\left(\hat{\theta}\right)$ are shown in Table 2. Note that the efficiency of the estimators cannot be evaluated by inspecting the MC SDs alone, because $D_0(\varepsilon)$ requires $(L+1) \times N = 3000$ simulations, while D_{I}, D_{II}, and D_{III} require $N = 1000$ simulations; the efficiency of the estimators is discussed in Section 5. The MC SDs of D_{II} and D_{III} for given N seem to be of a similar order of magnitude, whereas $D_0(.2)$ and $D_0(1.0)$ seem to do considerably better and D_{I} considerably worse.

The MC SDs of the elements of $D_0(.2)$ are 2.4 to 3.1 times as large as the corresponding MC SDs of $D_0(1.0)$, which is in line with the fact that the variance of $D_0(\varepsilon)$ is at least of order ε^{-1} (see Lemma 1), and if independent random numbers are used, it is of order ε^{-2}; hence the ratio of MC standard deviations of $D_0(.2)$ to $D_0(1.0)$ is expected to be between $\sqrt{5} = 2.24$ and 5, and because common random numbers are used, closer to 2.24 than to 5, which is indeed the case.

Concerning D_{I}, note that D_{I} estimates the first row of $\Delta\left(\hat{\theta}\right)$ much more accurately than the second row, which is not surprising. By (41), the standard deviations of coordinates U_1 and U_2 of U both equal 10.43, while the expectation of U is given by $E_{\hat{\theta}}U = (154, 653)'$; therefore, the standard deviations are small relative to the expectations and U can be considered to be "almost constant" for practical purposes. Then $\mathrm{Var}_{\hat{\theta}}\left(D_{\mathrm{I}}\right) = \mathrm{Var}_{\hat{\theta}}\left(U S_{\hat{\theta}}\right)\Big/ N \approx u^2\, \mathrm{Var}_{\hat{\theta}}\left(S_{\hat{\theta}}\right)\Big/ N$ (in case $L = 1$, with obvious extension to $L > 1$), and the ratio of the MC SDs of the second row to the first row is expected to be roughly $\frac{653}{154} = 4.24$, which is indeed the case.

Concerning the LR control variate estimators D_{II} and D_{III}, it is obvious that the introduction of the complete-data score as a control variate reduces the variance considerably as compared to D_{I}. The small advantage of D_{II} over D_{III}

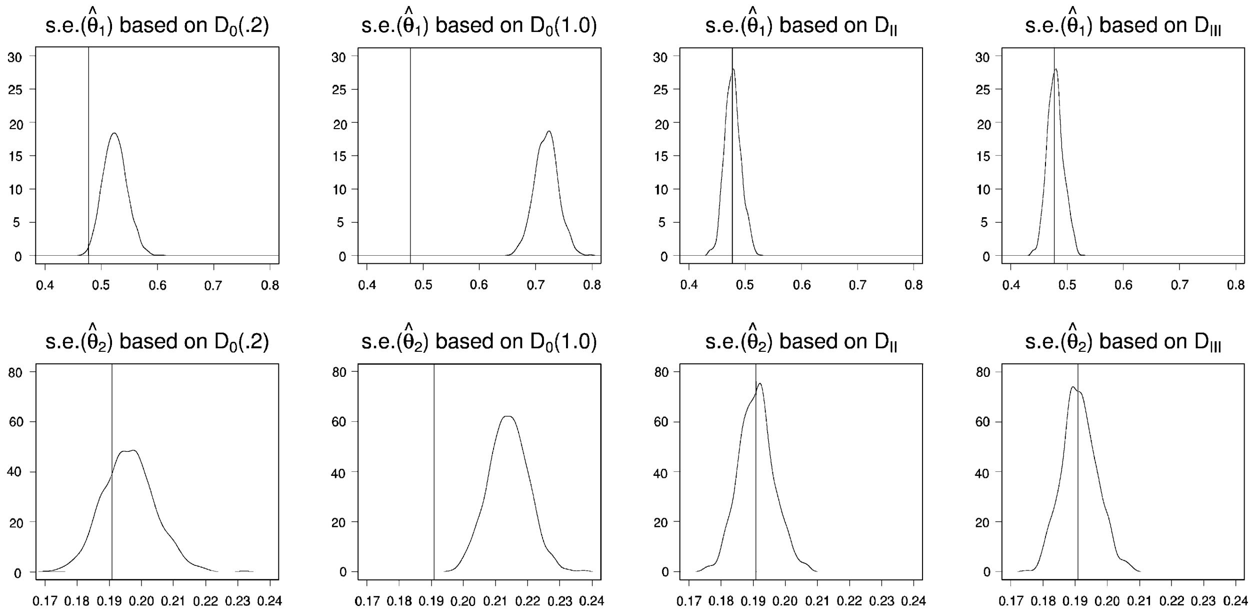

Fig. 1. EIES data: kernel density plots of Monte Carlo estimates of s.e.$\left(\hat\theta_1\right)$ and s.e.$\left(\hat\theta_2\right)$ based on 1000 Monte Carlo samples. The exact values of the standard errors, s.e.$\left(\hat\theta_1\right)$ = .477 and s.e.$\left(\hat\theta_2\right)$ = .191, are represented by vertical lines.

can be explained as follows. Since the standard deviations of the coordinates of U are small relative to the expectations, $\hat{B} \approx u$ and $D_{\mathrm{II}} \approx D_{\mathrm{III}}$ as argued in Section 3.2.3, and therefore it is not surprising that the MC SDs of D_{II} and D_{III} are of a similar order of magnitude. The fact that D_{II} seems to outperform D_{III} slightly may be due to the additional sampling variance resulting from the estimation of B given by (32).

As was pointed out above, the ultimate aim of the derivative estimators is to produce Monte Carlo estimates of

$$\varDelta^{-1}\left(\hat\theta\right)\varSigma\left(\hat\theta\right)\left[\varDelta^{-1}\left(\hat\theta\right)\right]', \tag{42}$$

which is an approximation of the variance–covariance matrix of $\hat\theta$ (see (10)). For the IA model, (42) can be evaluated analytically by using (41). The resulting standard errors of $\hat\theta_1$ and $\hat\theta_2$ are given by s.e.$\left(\hat\theta_1\right) = .477$ and s.e.$\left(\hat\theta_2\right) = .191$, respectively; these values will be referred to as the exact standard errors, exact in the sense that they are based on analytical evaluation of (42) and not estimated from Monte Carlo simulations. It is of interest to study the behavior of the Monte Carlo estimators of the standard errors, which are obtained by plugging in estimators $D_0(.2)$, $D_0(1.0)$, D_{I}, D_{II}, and D_{III} for $\varDelta\left(\hat\theta\right)$ and a Monte Carlo estimator for $\varSigma\left(\hat\theta\right)$. Fig. 1 shows Gaussian kernel density plots of the Monte Carlo estimates of the standard errors based on $D_0(.2)$, $D_0(1.0)$, D_{II}, and D_{III}; the plots corresponding to D_{I} are omitted because the estimated standard errors have huge MC SDs. All distributions are fairly symmetric, but the MC SDs of the standard errors are larger for $D_0(.2)$ and $D_0(1.0)$ than for the LR estimators D_{II} and D_{III}. The estimators of the standard errors based on $D_0(.2)$ and $D_0(1.0)$ seem to be upwards biased, in particular $D_0(1.0)$ leads to a large bias in estimated standard errors. In contrast, the standard errors based on D_{II} and D_{III} seem to be (almost) unbiased.

As a side remark, Snijders and van Duijn (1997) report for the IA model applied to the EIES data standard errors s.e.$\left(\hat\theta_1\right) = .22$ and s.e.$\left(\hat\theta_2\right) = .24$, which are Monte Carlo estimates based on $D_0(\varepsilon)$; the value of ε is not reported. Thus, the exact standard error s.e.$\left(\hat\theta_1\right) = .477$ is underestimated by 53.9%, which can be explained by (a) bias (if ε was large) or (b) large MC SD (if ε was small) or both; (a) is not too plausible, because Fig. 1 suggests an upwards bias; however, no matter what explanation is applicable here, the case illustrates that choosing ε is hard and can have great practical implications.

Please cite this article as: Michael Schweinberger, Tom A.B. Snijders, Markov models for digraph panel data: Monte Carlo-based derivative estimation, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.014

Table 3
Van de Bunt data: average estimates of $\Delta\left(\hat{\theta}\right)$ across 1000 Monte Carlo samples

$D_0(.2)$				$D_0(1.0)$			
78.96	23.93	17.14	2.84	51.86	19.26	22.85	.85
28.88	30.33	37.48	6.23	16.00	26.36	40.10	2.13
150.35	64.81	270.23	15.64	91.13	57.63	342.82	7.79
3.41	4.53	5.33	17.82	4.82	3.65	4.97	17.52
D_{I}				D_{II}			
88.38	25.63	−6.80	2.31	86.74	25.04	7.26	3.24
34.30	31.36	22.27	6.75	33.52	31.08	28.94	7.20
172.60	67.40	203.58	15.29	170.02	66.47	225.66	16.76
2.85	4.60	6.53	17.61	2.93	4.63	5.83	17.57
D_{III}							
86.55	24.98	7.27	3.23				
33.44	31.02	28.88	7.18				
169.65	66.31	225.25	16.70				
2.93	4.62	5.80	17.54				

4.2. Application: Jacobian matrix unknown

In the present section, the derivative estimators $D_0(.2)$, $D_0(1.0)$, D_{I}, D_{II}, and D_{III} are compared in the common situation where the Jacobian matrix is unknown.

Snijders (2001) studied data collected by Van de Bunt (1999), concerning a friendship relation among $n = 32$ university freshmen enrolled in a common study program. The digraph was observed at 7 time points. Here, the digraph evolution between observation points $t_2 < t_3 < t_4$ is modeled.

A simple model is specified by constant rate functions and objective function

$$f_i(\beta, x, j) = \sum_{k=1}^{4} \beta_k s_{ik}(x, j), \tag{43}$$

where the statistics s_{ik} are given by

$s_{i1}(x, j) = \sum_{h=1}^{n} x_{ih}$: the number of outgoing arcs (“outdegree”),
$s_{i2}(x, j) = \sum_{h=1}^{n} x_{ih} x_{hi}$: the number of reciprocated arcs,
$s_{i3}(x, j) = \sum_{h=1}^{n} (1 - x_{ih}) \max_l x_{il} x_{lh}$: the number of indirect connections,
$s_{i4}(x, j) = c_i\, s_{i1}(x, j)$: interaction of outdegree and gender of student i,
where s_{i1} is the outdegree of i and $c_i = 1$ if i is male and 0 otherwise.

Conditioning, as described in Snijders (2001), on the observed number of changes, which is 60 in $[t_2, t_3]$ and 51 in $[t_3, t_4]$, the parameter θ to be estimated by the method of moments reduces to $\theta = \left(\beta_1, \beta_2, \beta_3, \beta_4\right)'$. The coordinate U_k of statistic vector U, corresponding to coordinate θ_k of θ, is given by $U_k = \sum_{i=1}^{n} s_{ik}$ $(k = 1, \ldots, 4)$. The moment estimate of θ turns out to be $\hat{\theta} = (-1.058, 2.507, -.535, -.562)'$.

The average estimates of the Jacobian matrix $\Delta\left(\hat{\theta}\right)$ are presented in Table 3. As was pointed out above, estimators $D_0(.2)$ and $D_0(1.0)$ are biased, whereas D_{I} and D_{II} are unbiased and D_{III} is approximately unbiased. An unbiased estimator of the bias of $D_0(.2)$ is the average of $D_0(.2) - D_{\mathrm{II}}$ across Monte Carlo samples; the bias of $D_0(1.0)$ can be estimated accordingly. The bias of $D_0(1.0)$ seems to be large; the bias of $D_0(.2)$ also is non-negligible. The average estimates of D_{II} and D_{III} agree closely, while the average estimate of D_{I} differs slightly from D_{II} and D_{III}.

The MC SDs of the estimates of $\Delta\left(\hat{\theta}\right)$ are shown in Table 4; to save space, attention is restricted to the diagonal elements of $\Delta\left(\hat{\theta}\right)$. Note that $D_0(\varepsilon)$ requires $(L+1) \times N = 5000$ simulations, while D_{I}, D_{II}, and D_{III} require $N = 1000$ simulations, so that the efficiency of the estimators cannot be the evaluated on the basis of the MC SDs alone; see

Please cite this article as: Michael Schweinberger, Tom A.B. Snijders, Markov models for digraph panel data: Monte Carlo-based derivative estimation, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.014

Table 4
Van de Bunt data: Monte Carlo standard deviations of the estimated diagonal elements δ_{ii} of $\Delta\left(\hat{\theta}\right)$ based on 1000 Monte Carlo samples

	$D_0(.2)$	$D_0(1.0)$	D_{I}	D_{II}	D_{III}
δ_{11}	1.55	.32	108.17	4.69	4.70
δ_{22}	1.54	.32	24.67	1.96	1.97
δ_{33}	5.63	1.43	267.89	19.84	19.95
δ_{44}	.73	.16	2.46	.94	.95

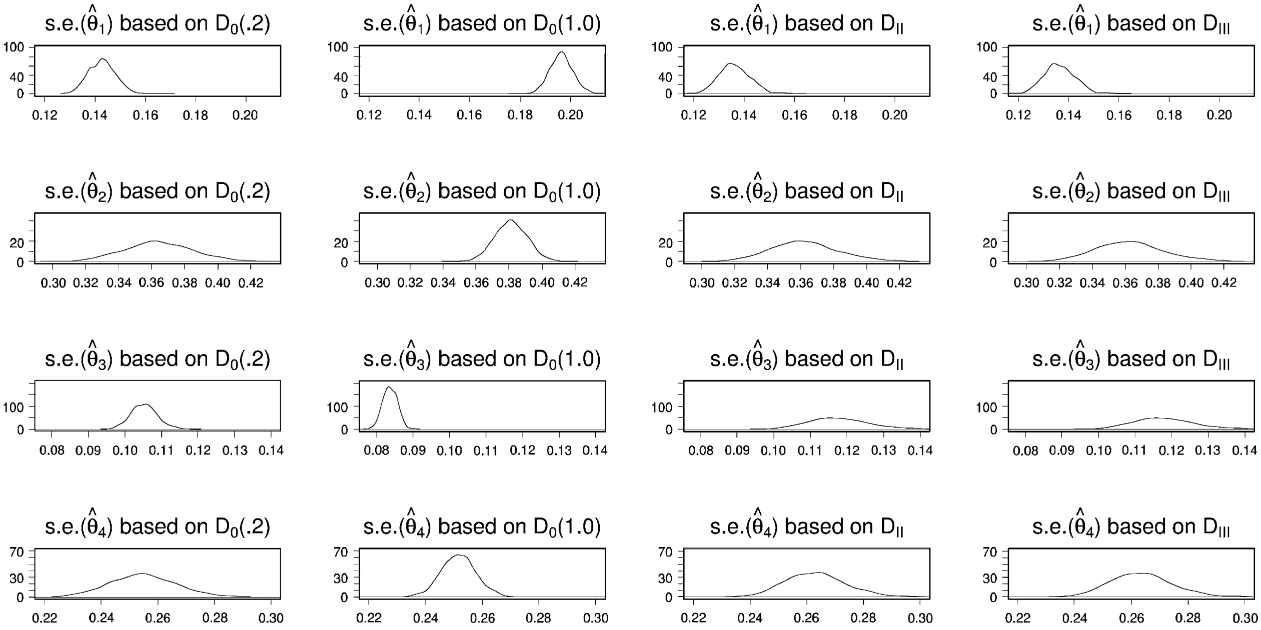

Fig. 2. Van de Bunt data: kernel density plots of Monte Carlo estimates of s.e.$\left(\hat{\theta}_k\right)$ $(k = 1, \ldots, 4)$ based on 1000 Monte Carlo samples.

Section 5. The MC SDs of D_{II} and D_{III} for given N seem to be of a similar order of magnitude, whereas $D_0(.2)$ and $D_0(1.0)$ seem to do considerably better and D_{I} considerably worse. Once again, the introduction of the complete-data score as a control variate reduces the variance of the LR estimator greatly.

It is of practical interest to assess how the methods perform in terms of the standard errors of the coordinates $\hat{\theta}_k$ of $\hat{\theta}$. Fig. 2 shows Gaussian kernel density plots of the estimated standard errors s.e.$\left(\hat{\theta}_k\right)$ $(k = 1, \ldots, 4)$ based on $D_0(.2)$, $D_0(1.0)$, D_{II}, and D_{III}; the estimates based on D_{I} have huge MC SDs and are omitted. The distributions of estimated standard errors based on D_{II} and D_{III} appear to be very similar; $D_0(.2)$, in turn, produces distributions that are similar to the ones based on D_{II} and D_{III}, apart from the distribution of estimates of s.e.$\left(\hat{\theta}_3\right)$ which has a smaller mean and a smaller MC SD than the corresponding distributions based on D_{II} and D_{III}. The estimator $D_0(1.0)$ gives rise to distributions that deviate more (see s.e.$\left(\hat{\theta}_1\right)$ and s.e.$\left(\hat{\theta}_3\right)$) or less (see s.e.$\left(\hat{\theta}_2\right)$ and s.e.$\left(\hat{\theta}_4\right)$) from the distributions produced by the other three methods; the estimated standard errors based on $D_0(1.0)$ seem to be biased.

Table 5 shows the MC SDs of the standard errors s.e.$\left(\hat{\theta}_k\right)$ $(k = 1, \ldots, 4)$ for Monte Carlo samples of size $N = 100$, 200, 500, and 1000; the Monte Carlo samples of size $N = 100$, 200, and 500 are obtained by taking the first 100, 200, and 500 observations of each Monte Carlo sample of size $N = 1000$, respectively; once again, the standard errors based on D_{I} are omitted. Table 5 indicates that the performance gap between $D_0(.2)$ on one hand and D_{II} and D_{III} on the other hand is relatively small, leaving aside s.e.$\left(\hat{\theta}_3\right)$. The MC standard deviations are roughly proportional to $N^{-1/2}$. However, it should be noted that, when using $D_0(.2)$ and in particular D_{II} and D_{III} and when N is small ($N = 100$ or 200), increasing N by factor c clearly reduces the MC SDs by more than $c^{1/2}$.

Table 5
Van de Bunt data: Monte Carlo standard deviations of Monte Carlo estimates of s.e.$(\hat{\theta}_k)$ $(k = 1, \dots, 4)$ based on 1000 Monte Carlo samples of size $N = 100$, 200, 500, and 1000

N		Monte Carlo standard deviations			
		Using $D_0(.2)$	Using $D_0(1.0)$	Using D_{II}	Using D_{III}
100	s.e.$\left(\hat{\theta}_1\right)$	.0188	.0143	.0257	.0293
	s.e.$\left(\hat{\theta}_2\right)$	.0813	.0322	.0864	.0925
	s.e.$\left(\hat{\theta}_3\right)$	.0134	.0061	.0392	.0408
	s.e.$\left(\hat{\theta}_4\right)$	.0409	.0191	.0454	.0508
200	s.e.$\left(\hat{\theta}_1\right)$	.0122	.0100	.0155	.0162
	s.e.$\left(\hat{\theta}_2\right)$	.0514	.0221	.0505	.0534
	s.e.$\left(\hat{\theta}_3\right)$	.0084	.0044	.0203	.0206
	s.e.$\left(\hat{\theta}_4\right)$	.0272	.0136	.0265	.0279
500	s.e.$\left(\hat{\theta}_1\right)$	.0074	.0065	.0090	.0091
	s.e.$\left(\hat{\theta}_2\right)$	.0296	.0138	.0289	.0294
	s.e.$\left(\hat{\theta}_3\right)$	.0051	.0028	.0119	.0120
	s.e.$\left(\hat{\theta}_4\right)$	.0167	.0085	.0156	.0160
1000	s.e.$\left(\hat{\theta}_1\right)$	.0052	.0046	.0062	.0062
	s.e.$\left(\hat{\theta}_2\right)$	.0207	.0099	.0208	.0210
	s.e.$\left(\hat{\theta}_3\right)$	.0036	.0020	.0080	.0081
	s.e.$\left(\hat{\theta}_4\right)$	.0114	.0060	.0108	.0110

A minimum requirement is that the MC SD should be less than 5% of the estimate to be useful in practice; the average estimated standard errors s.e.$\left(\hat{\theta}_k\right)$ $(k = 1, \dots, 4)$ based on D_{II} are .136, .363, .118, and .263, respectively. According to Table 5, for $N = 1000$ each of the estimators $D_0(.2)$, D_{II}, and D_{III} meets the 5%-standard for two standard errors, and "almost" meets it for the other two standard errors; therefore, it is sensible to slightly increase the sample size N.

5. Conclusion

Three likelihood ratio/score function (LR) estimators of the Jacobian matrix of the estimating function, D_{I}, D_{II}, and D_{III}, were proposed, and compared with the conventional estimator $D_0(\varepsilon)$ based on finite differences.

Based on theoretical and empirical evidence, it is safe to say that the finite differences estimator $D_0(\varepsilon)$ should be used with much care; in fact, the difficult choice of ε and the associated bias-variance dilemma, together with the computational disadvantage, are strong arguments against using $D_0(\varepsilon)$.

Concerning the LR estimators, the huge variance of LR estimator D_{I} renders D_{I} useless for practical applications. The efficiency of the LR control variate estimators D_{II} and D_{III} relative to $D_0(\varepsilon)$ can be evaluated by the classical efficiency ratio of Hammersley and Handscomb (1964, p. 51), which can be written as

$$\frac{\mathrm{Var}_\theta\left[D_0(\varepsilon)\right]}{\mathrm{Var}_\theta\left[D_{\mathrm{II}} \text{ or } D_{\mathrm{III}}\right]} \times \frac{\mathrm{time}\left(D_0(\varepsilon)\right)}{\mathrm{time}\left(D_{\mathrm{II}} \text{ or } D_{\mathrm{III}}\right)} = \frac{\mathrm{Var}_\theta\left[D_0(\varepsilon)\right]}{\mathrm{Var}_\theta\left[D_{\mathrm{II}} \text{ or } D_{\mathrm{III}}\right]} \times (L+1), \tag{44}$$

Please cite this article as: Michael Schweinberger, Tom A.B. Snijders, Markov models for digraph panel data: Monte Carlo-based derivative estimation, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.014

where $\mathrm{Var}_\theta[D]$ refers to some element of derivative estimator D—that is, to some partial derivative—and "time(D)" refers to the amount of computation time required to evaluate D. It is evident that the efficiency ratio tends to favor D_{II} and D_{III} if L is moderate or large. In Section 4.1, where the simple IA model with $L=2$ parameters was considered, the efficiency ratio of D_{II} relative to $D_0(.2)$ turns out to be .1 for each of the two elements on the main diagonal of the Jacobian matrix. In Section 4.2, where $L=4$, the efficiency ratio of D_{II} relative to $D_0(.2)$ is .5, 3.1, .4, and 3.0 for the four elements on the main diagonal of the Jacobian matrix. However, in practice L is frequently larger than 10, which tends to favor D_{II} and D_{III}.

Overall, the conclusion is that using D_{II} or D_{III} is preferable to using $D_0(\varepsilon)$.

An alternative approach to derivative estimation is to combine the estimation of θ and $\Delta(\theta)=\partial E_\theta U/\partial\theta'$ as follows: if the interim estimate $\hat{\theta}_N$ generated by the stochastic approximation algorithm for solving estimating Eq. (9) is in a small neighborhood of the solution $\hat{\theta}$, then the simulations from the distributions corresponding to the interim estimates $\hat{\theta}_N$ can be used to estimate the surface of $E_\theta U$ as a function of θ, and to produce derivative estimates of $\Delta\left(\hat{\theta}\right)$ based on fitting a linear model. It will be useful to investigate the computational benefits obtainable by such an approach.

The proposed estimators are implemented in the Windows-based computer program `Siena` embedded in the program collection `StOCNET`, which can be downloaded free of charge from `http://stat.gamma.rug.nl/stocnet`.

Acknowledgments

We are grateful for valuable comments of the reviewers.

Appendix A. Complete-data efficient score

Let M be the total number of changes of $x(t)$ in time interval $[t_0, t_1]$, and let x_0 be the digraph $x\,(t_0)$ observed at time point t_0. The Markov process corresponds to holding times $h_1,\ldots,h_M$ and some sequence $x_0, x_1,\ldots,x_M$ of digraphs, where x_m is the digraph to which the process moves by the mth transition, which takes place at time $t_0+\sum_{i=1}^m h_i$. Conditional on $x\,(t_0)$, the complete-data probability density $p_\theta = p_\theta(z)$ based on an observed outcome of the Markov process is given by

$$\prod_{m=1}^{M}\left[q_\theta\,(x_{m-1})\exp\left[-q_\theta\,(x_{m-1})\,h_m\right]\times\frac{q_\theta(x_{m-1},x_m)}{q_\theta\,(x_{m-1})}\right]\exp\left[-q_\theta\,(x_M)\,h_{M+1}\right], \tag{A.1}$$

where $h_{M+1}=(t_1-t_0)-\sum_{m=1}^M h_m>0$. The parameter $q_\theta\,(x_{m-1})$ of the negative exponential distribution is given by

$$q_\theta\,(x_{m-1})=-q_\theta\,(x_{m-1},x_{m-1})=\sum_{j=1}^{n}\lambda_j\,(\theta,x_{m-1}), \tag{A.2}$$

while $q_\theta\,(x_{m-1},x_m)$ is decomposed according to

$$q_\theta\,(x_{m-1},x_m)=\lambda_{i_m}\,(\theta,x_{m-1})\,r_{i_m}\,(\theta,x_m,j_m), \tag{A.3}$$

where the rate function λ_{i_m} and the conditional probability mass function r_{i_m} are given by (6) and (7), respectively.

Section A.1 gives the complete-data efficient score with respect to $\theta=\left(\rho,\alpha',\beta'\right)'$. Section A.2 considers briefly some more general models and the associated complete-data efficient score.

A.1. Complete-data efficient score with respect to $\theta=(\rho,\alpha',\beta')'$

If η represents either ρ or an element of $\alpha=(\alpha_k)$, then the complete-data score corresponding to η can be written as

$$\sum_{m=1}^{M}\left[\frac{\partial\ln\lambda_{i_m}\,(\theta,x_{m-1})}{\partial\eta}-\frac{\partial q_\theta\,(x_{m-1})}{\partial\eta}h_m\right]-\frac{\partial q_\theta\,(x_M)}{\partial\eta}h_{M+1}, \tag{A.4}$$

where

$$\frac{\partial \ln \lambda_{i_m}(\theta, x_{m-1})}{\partial \rho} = \frac{1}{\rho}, \tag{A.5}$$

$$\frac{\partial q_\theta(x_{m-1})}{\partial \rho} = \frac{q_\theta(x_{m-1})}{\rho}, \tag{A.6}$$

$$\frac{\partial \ln \lambda_{i_m}(\theta, x_{m-1})}{\partial \alpha_k} = a_{i_m k}\left(x_{m-1}, c_{i_m}\right), \tag{A.7}$$

and

$$\frac{\partial q_\theta(x_{m-1})}{\partial \alpha_k} = \sum_{j=1}^{n} a_{jk}\left(x_{m-1}, c_j\right) \lambda_j(\theta, x_{m-1}). \tag{A.8}$$

The complete-data score with respect to β_k is given by

$$\sum_{m=1}^{M}\left[s_{i_m k}(x_m, j_m) - \sum_{j=1}^{n} s_{i_m k}(x_m, j)\, r_{i_m}(\theta, x_m, j)\right]. \tag{A.9}$$

A.2. *Other, more general cases*

The present section discusses briefly how the complete-data efficient score for some selected, more general models deviates from the simple case above.

A.2.1. *$G > 1$ time intervals*

If the digraph is observed at more than two time points, that is, $G > 1$, then the complete-data score corresponding to parameter coordinates that are constant across time intervals is obtained by simply summing the complete-data score across the time intervals.

A.2.2. *“Thinning”*

Modeling social science network data sometimes makes it desirable to give nodes (“actors”), when designated to change something, the freedom not to change anything. In formal terms, the additional freedom is represented by replacing the constraint

$$\sum_{j \neq i}^{n} r_i(\theta, x_m, j) = 1 \tag{A.10}$$

by the constraint

$$\sum_{j \neq i}^{n} r_i(\theta, x_m, j) \leqslant 1 \tag{A.11}$$

so that

$$r_i(\theta, x_m, i) = 1 - \sum_{j \neq i}^{n} r_i(\theta, x_m, j) \geqslant 0. \tag{A.12}$$

Therefore, in step (2) of the simulation method described in Remark 2.1 of Section 2, it is admissible that $j = i$, and if $j = i$, then $x^{(M)} = x^{(M-1)}$. The parameter of the negative exponential distribution after thinning the Markov process (that is, after omitting the “changes” which do not lead to a change of state because $j = i$) is given by

$$q_\theta(x_{m-1}) = \sum_{j=1}^{n} \lambda_j(\theta, x_{m-1}) \sum_{h \neq j}^{n} r_j(\theta, x_m, h), \tag{A.13}$$

which leads to inconvenient derivatives due to the dependence of $q_\theta\left(x_{m-1}\right)$ on β through the conditional probabilities r_j. It is therefore more appealing to consider the Markov process before thinning. The parameter of the negative exponential distribution before thinning is

$$q_\theta\left(x_{m-1}\right)=\sum_{j=1}^{n}\lambda_j\left(\theta,x_{m-1}\right). \tag{A.14}$$

Let the complete data correspond to all the holding times, the events which do not result in change, and the events which do result in change. The complete-data likelihood then is proportional to the joint probability of the complete sequence of events, where an event may or may not lead to some change. The same formulae can be used as in Section A.1.

A.2.3. Co-evolution of digraphs and other outcome variables

The model can be extended to include, in addition to the Markov process that shapes the digraph, continuous-time Markov processes that shape other outcome variables; see Snijders et al. (2006). It is beyond the scope of the present paper to describe such models in detail, but it should be noted that convenient parametrizations lead to simple derivatives.

Appendix B. Complete-data efficient score: IA model

The same notation is used as in Appendix A. Formula (A.1) concerning the complete-data probability density is valid, but, using (A.2) and (38),

$$\begin{aligned}q_\theta\left(x_{m-1}\right)&=\sum_{j=1}^{n}\lambda_j\left(\theta,x_{m-1}\right)\\&=\theta_1\frac{\left(n(n-1)-x^\star_{++}\right)\exp\left(\theta_2\right)+x^\star_{++}\exp\left(-\theta_2\right)}{n-1},\end{aligned} \tag{B.1}$$

and, by (37),

$$q_\theta\left(x_{m-1},x_m\right)=\theta_1\frac{\left(1-x^{(m-1)}_{i_m j_m}\right)\exp\left(\theta_2\right)+x^{(m-1)}_{i_m j_m}\exp\left(-\theta_2\right)}{n-1}, \tag{B.2}$$

where $x^{(m-1)}_{i_m j_m}$ is the arc variable of digraph x_{m-1} that is changed by the mth move of the Markov process, and $x^{(m-1)}_{++}=\sum_{i=1}^{n}\sum_{h\neq i}^{n}x^{(m-1)}_{ih}$.

The complete-data score with respect to θ_k $(k=1,2)$ can be written as

$$\sum_{m=1}^{M}\left[\frac{\partial\ln q_\theta\left(x_{m-1},x_m\right)}{\partial\theta_k}-\frac{\partial q_\theta\left(x_{m-1}\right)}{\partial\theta_k}h_m\right]-\frac{\partial q_\theta\left(x_M\right)}{\partial\theta_k}h_{M+1}, \tag{B.3}$$

where

$$\frac{\partial\ln q_\theta\left(x_{m-1},x_m\right)}{\partial\theta_1}=\frac{1}{\theta_1}, \tag{B.4}$$

$$\frac{\partial q_\theta\left(x_{m-1}\right)}{\partial\theta_1}=\frac{q_\theta\left(x_{m-1}\right)}{\theta_1}, \tag{B.5}$$

$$\frac{\partial\ln q_\theta\left(x_{m-1},x_m\right)}{\partial\theta_2}=\frac{\left(1-x^{(m-1)}_{i_m j_m}\right)\exp\left(\theta_2\right)-x^{(m-1)}_{i_m j_m}\exp\left(-\theta_2\right)}{\left(1-x^{(m-1)}_{i_m j_m}\right)\exp\left(\theta_2\right)+x^{(m-1)}_{i_m j_m}\exp\left(-\theta_2\right)}, \tag{B.6}$$

and

$$\frac{\partial q_\theta\left(x_{m-1}\right)}{\partial\theta_2}=\theta_1\frac{\left(n(n-1)-x^{(m-1)}_{++}\right)\exp\left(\theta_2\right)-x^{(m-1)}_{++}\exp\left(-\theta_2\right)}{n-1}. \tag{B.7}$$

Appendix C. Variance–covariance matrix $\Sigma(\theta)$ and Jacobian matrix $\Delta(\theta)$: IA model

Let $M_{0+} = M_{00} + M_{01}$ and $M_{1+} = M_{10} + M_{11}$, where M_{kl} is defined by (40). It can be shown (see Snijders and van Duijn, 1997) that M_{01} and M_{11} are independent random variables with distribution

$$M_{01} \sim \text{Binomial}\left(M_{0+}, \xi_0(T)\right) \tag{C.1}$$

and

$$M_{11} \sim \text{Binomial}\left(M_{1+}, \xi_1(T)\right). \tag{C.2}$$

The parameters $\xi_0(T)$ and $\xi_1(T)$ of the binomial distributions are given by

$$\xi_0(T) = p\left(\theta_2\right)\left(1 - \exp\left[-r\left(\theta_1, T\right) q\left(\theta_2\right)\right]\right) \tag{C.3}$$

and

$$\xi_1(T) = p\left(\theta_2\right) + \left(1 - p\left(\theta_2\right)\right) \exp\left[-r\left(\theta_1, T\right) q\left(\theta_2\right)\right], \tag{C.4}$$

where $T = t_1 - t_0$,

$$p\left(\theta_2\right) = \frac{\exp\left(\theta_2\right)}{\exp\left(\theta_2\right) + \exp\left(-\theta_2\right)}, \tag{C.5}$$

$$q\left(\theta_2\right) = \exp\left(\theta_2\right) + \exp\left(-\theta_2\right), \tag{C.6}$$

and

$$r\left(\theta_1, T\right) = \frac{\theta_1 T}{n-1}. \tag{C.7}$$

Due to the conditioning on digraph $x\left(t_0\right)$ observed at time point t_0, M_{0+} and M_{1+} are known constants. Thus, for given $x(t_0)$ and $\theta = (\theta_1, \theta_2)'$, the expectation $E_\theta U$ is known,

$$E_\theta U = E_\theta \begin{pmatrix} M_{01} + M_{10} \\ M_{01} + M_{11} \end{pmatrix} = \begin{pmatrix} M_{0+}\xi_0(T) + M_{1+}\left(1 - \xi_1(T)\right) \\ M_{0+}\xi_0(T) + M_{1+}\xi_1(T) \end{pmatrix}, \tag{C.8}$$

the variance–covariance matrix $\Sigma(\theta) = \text{Cov}_\theta\, U$ is given by

$$\Sigma(\theta) = \begin{pmatrix} \text{Var}_\theta\, M_{01} + \text{Var}_\theta\, M_{11} & \text{Var}_\theta\, M_{01} - \text{Var}_\theta\, M_{11} \\ \text{Var}_\theta\, M_{01} - \text{Var}_\theta\, M_{11} & \text{Var}_\theta\, M_{01} + \text{Var}_\theta\, M_{11} \end{pmatrix}, \tag{C.9}$$

and the Jacobian matrix $\Delta(\theta) = \partial E_\theta U / \partial \theta'$ with respect to $\theta' = (\theta_1, \theta_2)$ is given by

$$\Delta(\theta) = \begin{pmatrix} M_{0+}\dfrac{\partial \xi_0(T)}{\partial \theta_1} - M_{1+}\dfrac{\partial \xi_1(T)}{\partial \theta_1} & M_{0+}\dfrac{\partial \xi_0(T)}{\partial \theta_2} - M_{1+}\dfrac{\partial \xi_1(T)}{\partial \theta_2} \\ M_{0+}\dfrac{\partial \xi_0(T)}{\partial \theta_1} + M_{1+}\dfrac{\partial \xi_1(T)}{\partial \theta_1} & M_{0+}\dfrac{\partial \xi_0(T)}{\partial \theta_2} + M_{1+}\dfrac{\partial \xi_1(T)}{\partial \theta_2} \end{pmatrix}, \tag{C.10}$$

where

$$\frac{\partial \xi_0(T)}{\partial \theta_1} = p\left(\theta_2\right) q\left(\theta_2\right) \exp\left[-r\left(\theta_1, T\right) q\left(\theta_2\right)\right] \frac{T}{n-1}, \tag{C.11}$$

$$\frac{\partial \xi_1(T)}{\partial \theta_1} = -\left(1 - p\left(\theta_2\right)\right) q\left(\theta_2\right) \exp\left[-r\left(\theta_1, T\right)\, q\left(\theta_2\right)\right] \frac{T}{n-1}, \tag{C.12}$$

$$\begin{aligned} \frac{\partial \xi_0(T)}{\partial \theta_2} = {} & \left(1 - \exp\left[-r\left(\theta_1, T\right) q\left(\theta_2\right)\right]\right) \frac{2}{q^2\left(\theta_2\right)} \\ & + p\left(\theta_2\right) r\left(\theta_1, T\right) \exp\left[-r\left(\theta_1, T\right) q\left(\theta_2\right)\right] \left(\exp\left(\theta_2\right) - \exp\left(-\theta_2\right)\right), \end{aligned} \tag{C.13}$$

and

$$\frac{\partial \xi_1(T)}{\partial \theta_2} = \left(1 - \exp\left[-r\left(\theta_1, T\right) q\left(\theta_2\right)\right]\right) \frac{2}{q^2\left(\theta_2\right)} \\ - \left(1 - p\left(\theta_2\right)\right) r\left(\theta_1, T\right) \exp\left[-r\left(\theta_1, T\right) q\left(\theta_2\right)\right] \left(\exp\left(\theta_2\right) - \exp\left(-\theta_2\right)\right). \tag{C.14}$$

References

Aleksandrov, V.M., Sysoyev, V.I., Shemeneva, V.V., 1968. Stochastic optimization. Eng. Cybern. 5, 11–16.

Cochran, W.G., 1977. Sampling Techniques. Wiley, New York.

Ferguson, T.S., 1996. A Course in Large Sample Theory. Chapman & Hall, London.

Fieller, E.C., Hartley, H.O., 1954. Sampling with control variables. Biometrika 41, 494–501.

Fishman, G.S., 1996. Monte Carlo. Concepts, Algorithms, and Applications. Springer, New York.

Glynn, P.W., L'Ecuyer, P., 1995. Likelihood ratio gradient estimation for regenerative stochastic recursions. Adv. Appl. Probab. 27, 1019–1053.

Hammersley, J.M., Handscomb, D.C., 1964. Monte Carlo Methods. Methuen, London.

Holland, P.W., Leinhardt, S., 1977. A dynamic model for social networks. J. Math. Sociology 5, 5–20.

L'Ecuyer, P., 1991. An overview of derivative estimation. In: Proceedings of the 1991 Winter Simulation Conference, pp. 207–217.

Lehmann, E.L., 1999. Elements of Large Sample Theory. Springer, New York.

Lehmann, E.L., Romano, J.P., 2005. Testing Statistical Hypotheses. 3rd ed., Springer, New York.

Marsaglia, G., Zaman, A., 1991. A new class of random number generators. Ann. Appl. Probab. 1, 462–480.

Pearson, K., 1902a. On the systematic fitting of curves to observations and measurements. Biometrika 1 (3), 256–303.

Pearson, K., 1902b. On the systematic fitting of curves to observations and measurements II. Biometrika 2 (1), 1–23.

Robbins, H., Monro, S., 1951. A stochastic approximation method. Ann. Math. Statist. 22, 400–407.

Rubinstein, R.Y., 1986. The score function approach for sensitivity analysis of computer simulation models. Math. Comput. Simul. 28, 351–379.

Rubinstein, R.Y., 1989. Sensitivity analysis and performance extrapolation for computer simulation models. Oper. Res. 37, 72–81.

Rubinstein, R.Y., Marcus, R., 1985. Efficiency of multivariate control variates in Monte Carlo simulation. Oper. Res. 33, 661–677.

Snijders, T.A.B., 2001. The statistical evaluation of social network dynamics. In: Sobel, M., Becker, M. (Eds.), Sociological Methodology. Basil Blackwell, Boston and London, pp. 361–395.

Snijders, T.A.B., Van Duijn, M.A.J., 1997. Simulation for statistical inference in dynamic network models. In: Conte, R., Hegselmann, R., Terna, P. (Eds.), Simulating Social Phenomena. Springer, Berlin, pp. 493–512.

Snijders, T.A.B., Steglich, C.E.G., Schweinberger, M., 2006. Modeling the co-evolution of networks and behavior. In: Montfort, K., Oud, H., Satorra, A. (Eds.), Longitudinal Models in the Behavioral and Related Sciences. Lawrence Erlbaum, in press.

Van de Bunt, G.G., 1999. Friends by choice. An Actor-Oriented Statistical Network Model for Friendship Networks through Time. Thesis Publishers, Amsterdam.

Wasserman, S., 1979. A stochastic model for directed graphs with transition rates determined by reciprocity. In: Schuessler, K.F. (Ed.), Sociological Methodology. Jossey-Bass, San Francisco, CA, pp. 392–412.

Wasserman, S., 1980. Analyzing social networks as stochastic processes. J. Amer. Statist. Assoc. 75, 280–294.

Please cite this article as: Michael Schweinberger, Tom A.B. Snijders, Markov models for digraph panel data: Monte Carlo-based derivative estimation, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.014