

On the Estimation of Path-Analysis Models with Hidden Variables

Stima di Modelli di Path-Analysis con Variabili Latenti

Fulvia Pennoni¹

Dipartimento di Statistica G. Parenti
Università degli Studi di Firenze
pennoni@ds.unifi.it

Riassunto: Il presente lavoro illustra un metodo per ottenere le stime di massima verosimiglianza utilizzando l'algoritmo EM per la classe dei modelli grafici direzionati aciclici (DAG) Gaussiani, in cui si marginalizza rispetto ad un nodo. Gli errori standard dei parametri stimati vengono calcolati utilizzando una formula esplicita che permette di ottenere la matrice di informazione osservata.

Keywords: Directed acyclic graphs, recursive structural equations, latent variables, EM algorithm, standard errors.

1. Introduction

The idea of graphical modelling in studying multivariate dependencies is to represent conditional independence structure of a multivariate random vector by a graph where the vertices correspond to variables and the absence of an edge between vertices stands for conditional or marginal independence. In recent times the combinations of ideas from the area of graphical Markov models with one from path analysis and, more generally, from structural equation models, has formed a stimulating ground for reinterpreting and enlarging existing results. These models generalize the theory of recursive path models Wright (1934), Wermuth (1980), and covariance selection models for continuous variables Dempster (1972). Examples of implementation of a unifying language, based on graphs, to permit a list of equivalent models can be found in Frydenberg (1989), to evaluate the state of global identifiability using graphical rules for a single factor model Stanghellini (1997), Vicard (2000).

Kiiveri, Speed and Carlin (1984) studied the factorization of the joint normal density of a set of random variables satisfying a recursive system of linear equations (e.g. a LISREL model), and they showed that some of the Markov properties of such a system can be read off the associated path diagram. For such types of models Stanghellini and Wermuth (2003) have studied graphical criteria for identification of all parameters of the model when one variable is hidden.

The motivation of the present work is to study linear structural equation models with uncorrelated residuals as directed acyclic graph (DAG) models with hidden variables. We focus on dependence structures derived from multivariate normal random variables by marginalizing with respect to some components; they are denoted in the graph by synthetic nodes to show that they are not observed. The problem of what can be learnt from the distribution of the observed variables about the joint distribution specified by the DAG

¹Indirizzo per corrispondenza: Dipartimento di Statistica G. Parenti, Viale Morgagni 59, 50134 Firenze.

will be assessed. The EM algorithm for finding solutions to the likelihood equations of the class of Gaussian models with such marginalization is considered. The EM algorithm does not provide the matrix of the second derivatives, we show how to get an explicit formula for the observed information matrix using the missing information principle (Little and Rubin, 2002).

2. Gaussian Directed Acyclic Graph with Latent Variables

Suppose $X = (X_1, X_2, \dots, X_k)$ is a finite set of substantive variables of interest ordered in certain way, such that there exist a subset of indices $pa(i) \subseteq \{i + 1, \dots, k\}$, $i = 1, \dots, k$, some independent random variables $\epsilon_1, \epsilon_2, \dots, \epsilon_k$ and linear functions $f_1, f_2, \dots, f_k$ such that

$$X_i = f_i(X_{pa(i)}, \epsilon_i), \quad i = 1, \dots, k \quad [X_{pa(i)} \equiv \{X_j : j \in pa(i)\}].$$

The set of equations $X_i = f_i(X_{pa(i)}, \epsilon_i)$ is called the *recursive* structural equations system.

The vector X has a distribution P on a directed acyclic graph (DAG) denoted by $G = (V, E)$ which consists of a set of vertices $V \equiv \{1, \dots, k\}$ representing $X = (X_1, X_2, \dots, X_k)$ and a set E of arrows $i \leftarrow j \in E$ iff $j \in pa(i)$; such that there are no direct paths that start and end at the same variable. We say that the distribution P is Markov with respect to G if considering pairwise disjoint subsets $A, B, C \subseteq V$ the following conditional independence relations hold

$$X_A \perp\!\!\!\perp X_B | X_C$$

whenever C separates A from B in G . For the Gaussian case that interests us, the global, local and pairwise Markov properties are identical (Lauritzen, 1996).

If the distribution P is Markov relative to G with density function p , the joint density factors into k univariate conditional densities

$$p(x_1, \dots, x_k) = \prod_{i=1}^k p(x_i | x_{pa(i)}).$$

In this work we will assume X to be a vector of mean centered random variables such that the recursive structural equations system (SEM) can be written in the *reduced form* equations:

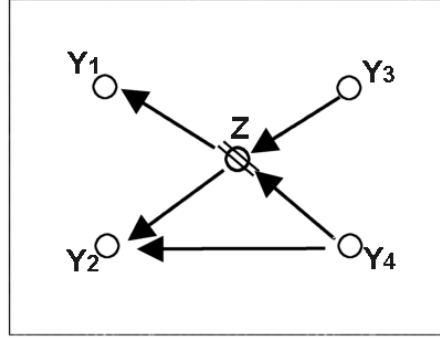
$$(A - I)X = \epsilon \quad \text{cov}(\epsilon) = \Delta \quad (1)$$

where A is $k \times k$ upper triangular matrix with ones along the diagonal and the elements $-a_{rs} = \beta_{rs.pa(r)\setminus s}$ are minus partial regression coefficients associated with an edge $X_s \leftarrow X_r$; $\text{cov}(\epsilon) = \Delta$ is a nonsingular diagonal covariance matrix of the residuals with elements of partial variances $\delta_{rr} = \sigma_{rr.pa(r)}$ along the diagonal. If the recursive system is specified to have a multivariate normal distribution over the error terms, covariances and concentrations matrices are respectively

$$\Sigma = B\Delta B' \quad \Sigma^{-1} = A'\Delta^{-1}A$$

where A and $B = A^{-1}$ are upper triangular; (A, Δ^{-1}) is called the triangular decomposition of the concentration matrix.

Figure 1: Directed acyclic graph with one latent variable Z



Supposing that we observe only a subset $Y = (Y_1, \dots, Y_p)$ of the variables, the complete data can be seen as $X = (Y, Z)$, where Y denotes the observed components of X and Z denotes the unobserved components; the corresponding DAG contains hidden nodes.

The following example of recursive SEM with one latent variable can naturally be associated with the DAG containing a syntetic node z of Figure 1, where the error terms are not represented in the graph

$$\begin{aligned} y_4 &= \epsilon_4 \\ y_3 &= \epsilon_3 \\ z &= a_{z3}y_3 + a_{z4}y_4 + \epsilon_z \\ y_2 &= a_{2z}z + a_{24}y_4 + \epsilon_2 \\ y_1 &= a_{3z}z + \epsilon_1 \end{aligned} \quad \Delta = \begin{pmatrix} \delta_{44} & 0 & 0 & 0 & 0 \\ 0 & \delta_{33} & 0 & 0 & 0 \\ 0 & 0 & \delta_{zz} & 0 & 0 \\ 0 & 0 & 0 & \delta_{22} & 0 \\ 0 & 0 & 0 & 0 & \delta_{11} \end{pmatrix}.$$

The relevant log-likelihood function based on the observed components can be written as

$$l_Y(\Sigma) = \frac{n}{2} [\log |\Sigma_{yy}^{-1}| - \text{tr}(\Sigma_{yy}^{-1} S_{yy})] \quad (2)$$

where Σ_{yy} is the observed covariance matrix.

The problem of what can be learnt from the distribution of the observed variables about the joint distribution specified by the DAG involves identifiability conditions. The model (1) is *globally* identified if the elements of A and Δ can be uniquely reconstructed from the parameters of the observable variables. The problem of identification of SEM with latent variables has been studied in econometric literature (Bollen, 1989) but general results are not available. Stanghellini and Wermuth (2003) gave criteria based on the properties of the graph to assess whether a path-analysis with one latent variable is identified.

Following Kiiveri (1987) we describe the maximum likelihood method for fitting such DAGs using the EM algorithm (Dempster *et al.*, 1977). The computations required are particularly straightforward: at the E-step we must compute $Q(\Sigma|\Sigma_r)$ the conditional expected value of the complete data log-likelihood to the observed data Y and a guessed initial value of complete data covariance matrix Σ_r

$$Q(\Sigma|\Sigma_r) = E(l(\Sigma, |Y_1, \dots, Y_p, \Sigma_r)). \quad (3)$$

Therefore at the M-step we maximize $Q(\Sigma|\Sigma_r)$ as a function of Σ to produce an improved estimate Σ_{r+1} .

Using the incomplete data approach a simple expression for the second derivative matrix of the observed log-likelihood can be derived in terms of the criterion function invoked by the EM algorithm. Minus the second derivative of the log-likelihood, when Σ is a function of a vector of parameters θ , is made of two parts

$$-\frac{\partial^2 l_Y}{\partial \theta \partial \theta} = -\frac{\partial^2 Q(\theta|\theta')}{\partial \theta \partial \theta} - \left(-\frac{\partial^2 H(\theta|\theta')}{\partial \theta \partial \theta} \right)$$

where Q is as in (3) and H is the expected value of the conditional density of the complete data X given the observed data Y (Tanner, 1996). Referring to $-Q$ as the complete information and to $-H$ as the missing information we have that the observed information is equal to the complete information minus the missing information due to the unobserved components (Little and Rubin, 2002). A basic result due to Louis (1982) is that if the complete data has distribution in a regular exponential family then $-\partial^2 H/\partial \theta \partial \theta = \text{Var}_{X|Y}(\partial l_X/\partial \theta)$, where l_X denotes the log-likelihood of the complete data. The missing information is determined by the conditional variance of the complete data log-likelihood l_X given Y . It follows that the second derivative of the log-likelihood of the observed data, can be expressed entirely in terms of the complete data log-likelihood as

$$-\frac{\partial^2 l_Y}{\partial \theta \partial \theta} = -\text{E}_{X|Y} \left[\frac{\partial^2 l_X}{\partial \theta \partial \theta} \right] - \text{Var}_{X|Y} \left(\frac{\partial l_X}{\partial \theta} \right).$$

Bibliography

- Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley, New York.
- Dempster, A. P. Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc. B*, 39, 1-38.
- Kiiveri, H. T., (1987). An incomplete data approach to the analysis of covariance structure. *Psychometrika*, 52, 539-554.
- Kiiveri, H. T., Speed T. P. and Carlin, J. B. (1984). Recursive causal models. *J. Austral. Math. Soc. Ser. A* 36-52.
- Lauritzen, S. L., (1996). *Graphical Models*. Oxford: Clarendon Press.
- Little, R. J. A. & Rubin, D. B. (2002). *Statistical Analysis with missing data*. 2nd Edition, Wiley.
- Louis T. A. (1982). Finding the observed information matrix when using the EM-algorithm. *Journal of the Royal Statistical Society, B*, 44, 226-233.
- Wright, S. (1934). The method of path coefficients. *Annals of Statistics*, 5, 161-215.
- Wermuth N. (1980). Linear recursive equations, covariance selection, and path analysis. *Amer. Statist. Assoc.*, 75, 963-972.
- Dempster, A. P. (1972). Covariance Selection. *Biometrics*, 28, 157-175.
- Stanghellini E. (1997). Identification of single-factor model using graphical Gaussian rules. *Biometrika*, 84, 241-244.
- Stanghellini E., Wermuth N. (2003). On the identification of path-analysis models with one hidden variable. *Submitted for publication in Biometrika*.
- Tanner, M.A. (1996). *Tools for statistical inference*. New York: Springer.