In: Encyclopedia of Behavioral Statistics, II. B. Everitt and David C. Howell (eds). Wiley, Chichester, 755-759).

Graphical chain models

Graphical Markov models represent relations, most frequently among random variables, by combining simple yet powerful concepts: data generating processes, graphs and conditional independence. The origins can be traced back to independent work in genetics (S. Wright, 1921, [30]), in physics (W. Gibbs, 1902, [10]) and in probability theory (A. A. Markov, 1912, [20]). Wright used directed graphs to describe processes of how his genetic data could have been generated and to check consistency of such hypotheses with observed data. He called his method path analysis. Gibbs described total energy of systems of particles by the number of nearest neighbors for nodes in undirected graphs. Markov suggested how some seemingly complex structures can sometimes be explained in terms of a chain of simple dependencies using the notion of conditional independence.

Development of these ideas continued largely independently in mathematics, physics, and engineering. In the social sciences and econometrics an extension of path analysis was developed, called simultaneous equation models, which does not directly utilize the notion of conditional independence and which does not incorporate nonlinear relations or time-dependent variation. In decision analysis, computer science, and philosophy extensions of path analysis have been called influence diagrams, belief networks, or Bayesian networks, and are used among others for constructing so-called expert systems and systems with learning mechanisms.

A systematic development of graphical Markov models for representing multivariate statistical dependencies for both discrete and continuous variables started in the 1970's with work on fully undirected graph models for purely discrete and for Gaussian random variables and on linear models with graphs that are fully directed and have no cycles. This work was extended to models permitting sequences of response variables to be considered on equal footing, that is without specifications of a direction of dependence. Joint responses can be modeled in quite different ways, some define independence structures of distinct types of graphical chain model. Properties of corresponding types of graph have been studied intensively, so that, in particular, all independencies, implied by a given graph, can be derived by so-called separation criteria.

Several books give overviews of theory, analyses and interpretations of graphical

Markov models in statistics, based on developments on this work during the first few decades, see [7], [15], [2], [29], and a wide range of different applications has been reported, see e.g. [11], [16]. For some compact descriptions and for references see [26], [27].

Applicability of fully directed graph models to very large systems of units has been emphasized recently, see e.g. [6] and is simplified by free source computational tools within the framework of the R-project, see [19], [18], [1].

Special extensions to time series have been developed ([5],[8],[9]) and it has been shown that the independence structure defined with any structural equation model (SEM) can be read off a corresponding graph [13]. The result does not extend to the interpretation of SEM parameters. Extensions to point processes and to multilevel models are in progress. Graphical criteria for deciding on the identifiability of special linear models including hidden variables have been derived [23], [21], [25], [12], [24].

A new approach to studying properties of graphical Markov models is based on binary matrix forms of graphs [28]. This uses analogies between partial inversion of parameter matrices for linear systems and partial closing of directed and of undirected paths in graphs. The starting point for this is are stepwise generating processes either for systems of linear equations or for joint distributions.

In both cases the graph consists of a set of nodes, with node i representing random variable Y_i and a set of directed edges. Each edge is drawn as an arrow outgoing from what is called a parent node and pointing to an offspring node. The graph is acyclic if it is impossible to return to any starting node by following arrows pointing in the same direction. The set of parent nodes of node i is denoted by par_i and the graph is called a parent graph if there is a complete ordering of the variables as $(Y_1, \ldots, Y_d)$. Either a joint density is given by a recursive sequence of univariate conditional densities or a covariance matrix is generated by a system of recursive equations.

The joint density f_N, generated over a parent graph with nodes $N = (1, \ldots, d)$ and written in a compact notation for conditional densities in terms of nodes, is

$$f_N = \prod_i f_{i|i+1,\ldots,d} = \prod_i f_{i|\mathrm{par}_i}. \tag{1}$$

The conditional independence statement $i \perp\!\!\!\perp j|\mathrm{par}_i$ is equivalent to the factorization $f_{i|\mathrm{par}_i,j} = f_{i|\mathrm{par}_i}$ and it is represented by a missing ij-arrow in the parent graph for $i < j$.

The joint covariance matrix Σ of mean-centered and continuous variables Y_i, generated over a parent graph with nodes $N = (1, \ldots, d)$, is given by a system of linear recursive equations with uncorrelated residuals, written as

$$AY = \varepsilon, \tag{2}$$

where A is an upper-triangular matrix with unit diagonal elements and ε is a residual vector of zero-mean uncorrelated random variables ε. A diagonal form of the residual covariance matrix $\mathrm{cov}(\varepsilon) = \Delta$ is equivalent to specifying that each row of A in (2) defines a linear least squares regression equation ([4], p.302) for response Y_i regressed on $Y_{i+1}, \ldots, Y_d$. For the regression coefficient of Y_j in this regression it holds for $i < j$:

$$-a_{ij} = \beta_{i|j.\{i+1,\ldots,d\}\setminus j} = \beta_{i|j.\mathrm{par}_i\setminus j}. \tag{3}$$

The vanishing contribution of Y_j to the linear regression of Y_i on $Y_{i+1}, \ldots, Y_d$ is represented by zero value in position (i, j) in the upper triangular part of A.

The types of question that can be answered for these generating processes are: which independencies (either linear or probabilistic) are preserved if the ordering the variables is modified or if some of the variables are considered as joint instead of univariate responses or if some of variables are explicitly omitted or if a subpopulation is selected? [28]. Joint response models which preserve exactly the independencies of the generating process after omitting some variables and conditioning on others form a slightly extended subclass of SEM models [22], [14].

Sequences of joint responses occur in different types of chain graphs. All these chain graphs have in common that the nodes are arranged in a sequence of say d_{CC} chain components g, each containing one or more nodes. For partially ordered nodes $N = (1, \ldots, g, \ldots, d_{CC})$ a joint density is considered in the form

$$f_N = \prod\nolimits_g f_{g|g+1,\ldots,d_{CC}}. \tag{4}$$

Within this broad formulation of chain graphs one speaks of multivariate-regression chains whenever for a given chain component g, variables at nodes i and j are considered conditionally given all variables in chain components $g + 1, \ldots, d_{CC}$. Then the univariate and bivariate densities

$$f_{i|g+1,\ldots,d_{CC}}, \qquad f_{ij|g+1,\ldots,d_{CC}} \tag{5}$$

determine the presence or absence of a directed ij-edge, which points to node i in chain component g from a node j in $g+1, \ldots, d_{CC}$, or of an undirected ij-edge within g when j itself is in g.

The more traditional form of chain graphs results if for a given chain component g variables at nodes i and j are considered conditionally given all other variables in g and the variables in $g+1, \ldots, d_{CC}$. Then the univariate and bivariate densities

$$f_{i|g\setminus\{i\},g+1,\ldots,d_{CC}}, \qquad f_{ij|g\setminus\{i,j\},g+1,\ldots,d_{CC}} \tag{6}$$

are relevant for a directed ij-edge which points to node i in chain component g from a node j in $g+1, \ldots, d_{CC}$, as well as for an undirected ij-edge within g.

These traditional chain graphs are called blocked-concentration graphs or sometimes LWF (Lauritzen, Wermuth, Frydenberg) graphs. Chain graphs with the undirected components as in blocked-concentration graphs and the directed components as in multivariate regressions graphs are called concentration-regression graphs or sometimes AMP (Andersson, Madigan, Perlman) graphs. The statistical models corresponding to the latter for purely discrete variables are the so-called marginal models. These belong to the exponential family of models and have canonical parameters for the undirected components and moment parameters for the directed components.

Stepwise generating processes in univariate responses arise both in observational and in intervention studies. With an intervention the probability distribution is changed so that the intervening variable is decoupled from all variables in the past that relate directly to it in an observational setting, see [17]. Two main assumptions distinguish "causal models with potential outcomes" (or counterfactual models) from general generating processes in univariate responses. These are (1) unit-treatment additivity and (2) a notional intervention. These two assumptions taken together assure that there are no unobserved confounders and that there is no interactive effect on the response by an unobserved variable and the intervening variable. One consequence of these assumptions is for linear models that the effect of the intervening variable on the response averaged over past variables coincides with its conditional effects given past unobserved variables. Some authors have named this a causal effect. For a comparison of different definitions of causality from a statistical viewpoint, including many references, and for the use of graphical Markov models in this context see [3].

References

[1] Badsberg, J.H. (2004). DynamicGraph: interactive graphical tool for manipulationg graphs. URL: http://cran.r-project.org.

[2] Cox, D. R. & Wermuth, N. (1996). *Multivariate dependencies: models, analysis, and interpretation.* Chapman and Hall, London.

[3] Cox, D.R. & Wermuth, N. (2004). Causality a statistical view. *Int. Statist. Rev.*. **72**, 285-305. .

[4] Cramér, H. (1946). *Mathematical methods of statistics.* Princeton, N.J.: Princeton University Press.

[5] Dahlhaus, R. (2000). Graphical interaction models for multivariate time series. *Metrika*, 157–172.

[6] Dobra, A. (2003). Markov bases for decomposable graphical models. *Bernoulli*, **9**, 1093–1108.

[7] Edwards, D. (2000). *Introduction to graphical modelling.* 2nd ed. Springer, New York.

[8] Eichler, M, Dahlhaus R. & Sandkühler J. (2003). *Partial correlation analysis for the identification of synaptic connections.* Biological Cybernetics. **89**, 289-302.

[9] Fried R. & Didelez, V. (2003). Decomposability and selection of graphical models for time series. *Biometrika.* **90**, 251-267.

[10] Gibbs, W. (1902). *Elementary Principles of Statistical Mechanics.* Yale Univ. Press, New Haven.

[11] Green, P.J., Hjort, N.L. & Richardson, S. (2003). *Highly Structured Stochastic Systems.* Oxford: University Press.

[12] Grzebyk M. & Wild, P. & Chouaniére, D. (2003). On identification of multi-factor models with correlated residuals. *Biometrika.* **91**, 141-151.

[13] Koster, J.T.A. (1999). On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors. *Scand. J. Statist.*, **26**, 413–431.

[14] Koster, J.T.A. (2002). Marginalizing and conditioning in graphical models. *Bernoulli*, **8**, 817–840.

[15] Lauritzen, S. L. (1996). *Graphical models.* Oxford University Press, Oxford.

[16] Lauritzen S.L. & N. A. Sheehan (2003). Graphical models for genetic analyses. *Statistical Science*, 18, 489–514.

[17] Lindley, D.V. (2002). Seeing and doing: the concept of causation. *Int. Statist. Rev.* **70**, 191-214.

[18] Marchetti, G. M. (2004). R functions for computing graphs induced from a DAG after marginalization and conditioning. Proceedings of the Amer. Statist. Ass. Alexandria, VA.

[19] Marchetti, G. M. & Drton, M. (2003). GGM: an R package for Gaussian graphical models. URL: http://cran.r-project.org.

[20] Markov, A.A. (1912). *Wahrscheinlichkeitsrechnung* (German translation of 2nd Russian edition: Markoff, A.A., 1908). Teubner, Leipzig.

[21] Pearl J. (1998). Graph, causality and structural equation models. *Sociological Methods and Research* **27**, 226-284.

[22] Richardson, T.S. & Spirtes, P. (2002). Ancestral Markov graphical models. *Ann. Statist.* **30**, 962–1030.

[23] Stanghellini, E. (1997). Identification of a single-factor model using graphical Gaussian rules. *Biometrika*, **84**. 241-254.

[24] Stanghellini, E. & Wermuth, N. (2004). On the identification of path analysis models with one hidden variable. *Biometrika.* To appear.

[25] Vicard, P. (2000). On the identification of a single-factor model with correlated residuals. *Biometrika*, **84**. 241-254.

[26] Wermuth, N. (1998). Graphical Markov models. *Encyclopedia of Statistical Sciences.* S. Kotz, C. Read and D. Banks (eds). Wiley, New York, Second Update Volume, 284-300.

[27] Wermuth, N. & Cox, D.R. (2001). Graphical models: overview. In: *International Encyclopedia of the Social and Behavioral Sciences* P.B. Baltes and N.J. Smelser (eds), Elsevier, Amsterdam, **9**, 6379–86.

[28] Wermuth, N. & Cox, D.R. (2004). Joint response graphs and separation induced by triangular systems. *J. Roy. Statist. Soc. B.* **66**, 687-717.

[29] Whittaker, J. (1990). *Graphical models in applied multivariate statistics.* Wiley, Chichester.

[30] Wright, S. (1921). Correlation and causation. *J. Agric. Res.* **20**, 162–177.