Graphical Models and Modelling

Presented to 6th Robust Conference, Liblice, January 1990

By Joe Whittaker, Lancaster University

Introduction

Graphical models represent multivariate interaction by means of a finite graph, in which nodes represent variables and edges represent association; more precisely, the absence of an edge represents a conditional independence relationship. Such graphs were introduced by Darroch, Lauritzen and Speed (1980), where the variables are discrete; their continuous counterpart was examined by Speed and Kiiveri (1986); and more recently, the mixed discrete and continuous case has been discussed by Lauritzen and Wermuth (1984, 1989) and generalised by Edwards (1989). A theory for directed independence graphs, based on the ideas of Wermuth and Lauritzen (1983), and Kiiveri, Speed and Carlin (1984), is developed by limiting the conditioning set to the 'past'. This material is very much related to the original path analysis ideas of Sewell Wright (1921, 1923).

The essential ingredients of a theory of graphical models are the notion of conditional independence, standard graph theory, the conditional Gaussian distribution, and maximum likelihood estimation. Our intention is to outline the basic theory, without derivation, and to report some practical applications of the techniques that are to appear in Whittaker (1990). The positive benefits that graphical modelling techniques bring to data analysis are: a probability framework to assess interaction, phrased in terms of the original untransformed variables; visual display; a framework which naturally includes both discrete and continuous variables; and a statistical model to assess goodness of fit.

Conditional Independence

Let X denote a k-dimensional random vector, $X = (X_1, X_2, \ldots, X_k)$ where X_i is the i-th coordinate of X. Put $K = \{1, 2, \ldots, k\}$, the index set consisting of the full set of suffices, and let $a = \{i_1, i_2, \ldots, i_p\}$ denote an arbitrary subset of K. Define the random vector X_a as the ordered tuple

$$X_a = (X_{i_1}, X_{i_2}, \ldots, X_{i_p}) = (X_i;\ i \in a).$$

Then $X_{a\cup b}$ and $X_{a\cap b}$ are well-defined, and so are other set operations, in particular, $X_{K\setminus\{i\}}$ denotes the the sub-vector of X obtained by excluding X_i.

The density function f_a is defined as the marginal density function of X_a, for $a \subseteq K$; and, if a and b disjoint, the conditional density of X_b given X_a is $f_{b|a} = f_{a\cup b}/f_a$, defined as f_a is positive.

Using the independence notation due to Dawid (1979), for disjoint sets a, b and c of K, $X_b \perp\!\!\!\perp X_c | X_a$ if and only if the conditional density function $f_{bc|a} = f_{b|a} f_{c|a}$ for all values of x. We shall assume that the density is positive on its support, and we point out that the theory has to be extensively reworked if the assumption of positivity is not met.

Independence Graphs

Let $X = (X_1, X_2, \ldots, X_k)$ denote a vector of random variables, and consider a graph, in the sense of Harary (1969) or Berge (1973), with k vertices representing each variable.

Definition. The *conditional independence graph* of X is the undirected graph $G = (K, E)$ where $K = \{1, 2, .., k\}$ and (i, j) is *not* in the edge set E if and only if $X_i \perp\!\!\!\perp X_j | X_{K\setminus\{i,j\}}$. □

The graph is a conditional independence graph, or more loosely, an independence graph, if there is no edge between two vertices whenever the pair of variables is independent given all the remaining variables.

The vector of the remaining variables is sometimes referred to as the *rest*. This definition is the *pairwise Markov property*; and because of its Markov properties, a better name might be a *Markov* graph, but unfortunately this term is already used in the theory of random graphs. We use the shorthand, for example $1 \perp\!\!\!\perp 2 | \{3,4\}$ for $X_1 \perp\!\!\!\perp X_2 | (X_3, X_4)$, so that the independence of X_i and X_j given the rest can be written as $i \perp\!\!\!\perp j | K \backslash \{i,j\}$. The resulting undirected graph gives a picture of the pattern of dependence or association between the variables.

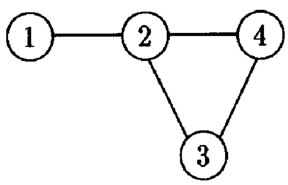
It is often easy to construct the independence graph if we are given the joint density function of X by repeated application of the factorisation criterion, Dawid (1979), for conditional independence. For example, take $k = 4$ and consider the density function

$$f_K(x) = \exp(u + x_1 + x_1 x_2 + x_2 x_3 x_4),$$

of $X = (X_1, X_2, X_3, X_4)$ on the 4 dimensional cube, $\{x; x = (x_1, x_2, x_3, x_4),\ 0 < x_i < 1,\ i = 1,2,3,4\}$, where the constant u ensures the density integrates to 1. The function factorises into $f_K(x) = g(x_1, x_2) h(x_2, x_3, x_4)$, and it is easy to verify that

$$X_1 \perp\!\!\!\perp X_4 | (X_2, X_3) \quad \text{and} \quad X_1 \perp\!\!\!\perp X_3 | (X_2, X_4).$$

Consequently the conditional independence graph is



The graph uses the fact that vertex 1 is not adjacent to either 3 or 4 for construction but highlights the fact that the *cliques* of the graph are $\{1,2\}$ and $\{2,3,4\}$. Note that this independence graph is identical to the interaction graph constructed by drawing an edge between variables that occur together in the interactions terms $x_1 x_2$ and $x_2 x_3 x_4$, of the log-linear expansion of the density function.

We remark on the importance of conditioning; there is no suitable theory for graphs constructed from pairwise marginal independences.

Markov Properties

Inspection of the independence graph in the example above suggests that conditionally on X_2, X_1 and (X_3, X_4) are independent. It is a remarkable feature of independence graphs that is true in general: more specifically that with respect to a given graph on a finite number of vertices, the following properties are equivalent:

- the pairwise independence property: that non-adjacent pairs of variables are independent conditional on the remaining variables;
- the local Markov property: that conditional only on the adjacent variables, any variable is independent of all the remaining variables; and
- the global Markov property: that any two subsets of variables separated by a third is independent conditionally only on variables in the third subset.

The strongest part of this assertion is the separation theorem, that the pairwise Markov property implies the global Markov property.

Since these properties are equivalent, any one could be taken as the definition of the conditional independence graph. But from the applied point of view of statistical modelling, it is most convenient to use the pairwise independence property. There are two essential reasons for this choice: firstly, the list of requirements to be verified in constructing the independence graph to model a given data set, is the least stringent, while the interpretations that follow are the strongest. Secondly, the pairwise definition of the independence relationship naturally corresponds to the pairwise graph-theoretic definition of the edge set in a graph.

There is now a fairly large published literature on the theory of graphical models and their applications, and some references are given here.

References

Andersen, A. H. (1974). Multidimensional contingency tables. *Scand. J. Statist.*, **1**, 115-127.

Asmussen, S. and Edwards, D. (1983). Collapsibility and response variables in contingency tables. *Biometrika*, **70**, 567-578.

Berge, C. (1973). *Graphs and Hypergraphs.* Transl. from French by E. Miniecka. North Holland: Amsterdam.

Birch, M. W. (1963). Maximum likelihood in three-way contingency tables. *J. Roy. Statist. Soc.*, **25**, 220-223.

Bishop Y. M., Fienberg, S. and Holland, P. (1975). *Discrete Multivariate Analysis.* M.I.T. Press: Cmabridge, Mass.

Darroch, J. N., Lauritzen, S. L. and Speed, T. P. (1980). Markov fields and log linear interaction models for contingency tables. *Ann. Stat.*, **8**, 522-539.

Dawid, A. P. (1979). Conditional independence in statistical theory (with discussion). *J. Roy. Statist. Soc. B*, **41**, 1, 1-31.

Dawid, A. P. (1980). Conditional independence for statistical operations. *Ann. Statist.*, **8**, 598-617.

Dempster, A. P. (1972). Covariance Selection. *Biometrics*, **28**, 157-175.

Edwards, D. E. (1987). A guide to MIM. Research Report 87/1. Statistical Research Unit, University of Copenhagen.

Edwards, D. E. (1989). Hierarchical interaction models, (with discussion). *J. R. Statist. Soc. B* **51**, 3.

Edwards, D. E. and Havranek, T. (1985). A fast procedure for model search in multidimensional contingency tables. *Biometrika*, **72**, 2, 339-51.

Edwards, D. E. and Kreiner, S. (1983). The analysis of contingency tables by graphical models. *Biometrika*. **70**, 3, 553-565.

Frydenberg, M. (1989). The chain graph Markov property. *Research Report, 186.* Department of Theoretical Statistics, University of Aarhus.

Golumbic, M. C. (1980). *Algorithmic Graph Theory and Perfect Graphs.* Academic Press: New York.

Goodman, L. A. (1970). The multivariate analysis of qualitative data: interaction among multiple classifications. *J. Amer. Statist. Assoc.* **65**, 226-256.

Goodman, L. A. (1973). The analysis of multidimensional contingency tables when some variables are posterior to others: a modified path analysis approach. *Biometrika*, **60**, 179-192.

Harary, F. (1969). *Graph Theory.* Addison-Wesley: Reading, Mass.

Kiiveri H. and Speed, T. P. (1982). Structural analysis of multivariate data: a review. *In* Leinhardt, S. (Ed). *Sociological Methodology.* Jossey Bass: San Francisco.

Kiiveri, H., Speed, T.P. and Carlin, J.B. (1984) Recursive causal models. *J. Aust. Math. Soc.*, **36** 30-52.

Kreiner, S. (1987). Analysis of multidimensional contingency tables by exact conditional tests: techniques and strategies. *Scand. J. Stat.*, **14**.

Lauritzen, S. L. (1982). *Lectures on Contingency Tables.* (2nd Ed.) University of Aalborg Press: Aalborg.

Lauritzen, S. L., Dawid, A. P., Larsen, B. N. and Leimer, H. G. (1988). Independence properties of directed Markov fields. Research Report. R-88-32. Institute of Electronic Systems. Aalborg University.

Lauritzen, S. L. and Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert sustems (with discussion). *J. Roy. Statist. Soc. B*, **50**, 2, 157-224.

Lauritzen, S. L. and Wermuth, N. (1984). Mixed interaction models. Research Report. R-84-8. Institute of Electronic Systems. Aalborg University.

Laúritzen, S. L. and Wermuth, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Statist.*, **17**, 31-57.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems.* Morgan and Kaufman: San Mateo.

Porteous, B. T. (1985). Improved likelihood ratio statistics for covariance selection models. *Biometrika*, **72**, 97-101.

Speed, T. P. (1978). *Graph-theoretic methods in the analysis of interaction.* Lecture Notes, Institute of Mathematical Statistics, University of Copenhagen.

Speed, T. P. and Kiiveri, H. (1986). Gaussian Markov distributions over finite graphs. *Ann. Statist*, **14**, 1, 138-150.

Spiegelhalter, D. J. (1986). Probabilistic reasoning in predictive expert systems. *In* Kanal, L. N. and Lemner, J. *Uncertainty in Artificial Intelligence.* 47-68. North-Holland.

Sundberg, R. (1975). Some results about decomposable (or Markov-type) models for multidimensional contingency tables: distribution of marginals and partitioning of tests. *Scand. J. Statist.*, **2**, 71-79.

Wermuth, N. (1976). Analogies between multiplicative models in contingency tables and covariance selection. *Biometrics*, **32**, 95-108.

Wermuth, N. (1976). Model search among multiplicative models. *Biometrics*, **32**, 253-264.

Wermuth, N. (1980). Linear recursive equations, covariance selection and path analysis. *J. Amer. Stat. Ass.*, **75**, 963-972.

Wermuth, N. and Lauritzen, S. L. (1983). Graphical and recursive models for contingency tables. *Biometrika*, **70**, 3, 537-552.

Wermuth, N. and Lauritzen, S. L. (1989). On substantive research hypotheses, conditional independence graphs and graphical chain models. *J. Roy. Statist. Soc. B*, **51**, 3.

Whittaker, J. (1982). Glim syntax and simultaneous tests for graphical log-linear models. *In* Gilchrist R. (Ed). *GLIM 82. Lecture notes in Statistics*, **14**, 98-108. Springer Verlag.

Whittaker, J. (1984). Fitting all possible decomposable and graphical models to multiway contingency tables. *In* Havranek, T. et al. (Eds). *Compstat*, Physica-Verlag: Vienna. 401-406.

Whittaker, J. (1989). Invited contribution to Edwards, and Wermuth and Lauritzen. *J. Roy. Statist. Soc. B*, 1989, **51**, 3.

Whittaker, J. (1990). *Independence Graphs in Applied Multivariate Statistics.* Wiley: Chichester.

Whittaker, J., Iliakopoulis., T. and Smith, P. (1988). Graphical modelling with large numbers of variables: a comparison with principal components. *In* Edwards, D.G. and Raun, N. E. (Eds). *Compstat 88.* Physica Verlag: Heidelberg. 73-80.

Wright, S. (1923). The theory of path coefficients: a reply to Niles' criticism. *Genetics*, **8**, 239-255.

Wright, S. (1934). The method of path coefficients. *Ann. Math. Statist.*, **5**, 161-215.