eighteenth century social scientists that Beniger and Robyn (1978) describe. Whatever the reason, there is little justification for this to continue. Software packages such as S-Plus, SAS, SPSS, and SYSTAT (used for the graphics in this article) now offer statistical graphics and publishing standards such as Adobe PDF format now make it possible to display graphics as easily as text.

See also: Bayesian Graphical Models and Networks; Graphical Models: Overview

Bibliography

Andrews D F, Herzberg A M 1985 *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. Springer-Verlag, New York

Becker R A, Cleveland W S, Shyu M-J 1996 The design and control of Trellis display. *Journal of Computational and Graphical Statistics* **5**: 123–55

Beniger J R, Robyn D L 1978 Quantitative graphics in statistics: A brief history. *The American Statistician* **32**: 1–11

Bertin J 1967 *Semiologie Graphique*. Editions Gauthier-Villars, Paris [English translation Berg W J, Wainer H 1983 *Semiology of Graphics*. University of Wisconsin Press, Madison, WI]

Bertin J 1981 *Graphics and Graphic Information-Processing*. [English translation Berg W J, Scott P. Walter de Gruyter, New York]

Bishop Y M, Fienberg S E, Holland P W 1975 *Categorical Data Analysis*. MIT Press, Cambridge, MA

Bureau of the Census 1986 *State and Metropolitan Area Data Book*. US Government Printing Office, Washington, DC

Chambers J M, Cleveland W S, Kleiner B, Tukey P A 1983 *Graphical Methods for Data Analysis*. Wadsworth, Monterey, CA

Cleveland W S 1984 Graphs in scientific publications. *The American Statistician* **38**: 19–26

Cleveland W S 1993 A model for studying display methods of statistical graphics (with discussion). *Journal of Computational and Graphical Statistics* **2**: 323–43

Cleveland W S 1994 *The Elements of Graphing Data*, rev. edn. Hobart Press, Summit, NJ

Cleveland W S 1995 *Visualizing Data*. Hobart Press, Summit, NJ

Cleveland W S, Devlin S 1988 Locally weighted regression analysis by local fitting. *Journal of the American Statistical Association* **83**: 596–640

Cleveland W S, McGill R 1984 Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association* **79**: 531–54

Collins B M 1993 Data visualization—has it all been seen before? In: Earnshaw R A, Watson D (eds.) *Animation and Scientific Visualization: Tools and Applications*. Academic Press, New York, pp. 3–28

Davis J A, Smith T W, Marsden P V 1993 *The General Social Survey*. National Opinion Research Center, Chicago

Fienberg S 1979 Graphical methods in statistics. *The American Statistician* **33**: 165–78

Funkhouser H G 1936 Note on a tenth century graph. *Osiris* **1**: 260–2

Funkhouser H G 1937 Historical development of the graphical representation of statistical data. *Osiris* **3**: 269–404

Gonnelli S, Cepollaro C, Montagnani A, Monaci G, Campagna M S, Franci M B, Gennari C 1996 Bone alkaline phosphatase measured with a new immunoradiometric assay in patients with metabolic bone diseases. *European Journal of Clinical Investigation* **26**: 391–6

Herdeg W 1981 *Graphis Diagrams: The Graphic Visualization of Abstract Data*. Graphis Press, Zürich, Switzerland

Holmes N 1991 *Designer's Guide to Creating Charts and Diagrams*. Watson-Guptill Publications, New York

Kosslyn S M 1994 *Elements of Graph Design*. W. H. Freeman, New York

McGill R, Tukey J W, Larsen W A 1978 Variations of box plots. *The American Statistician* **32**: 12–6

Mosteller F, Tukey J W 1968 Data analysis, including statistics. In: Lindzey G, Aronson E (eds.) *Handbook of Social Psychology*, 2nd edn. Addison-Wesley, Reading, MA, pp. 80–203

Pinker S 1997 *How the Mind Works*. Norton, New York

Robinson A H 1982 *Early Thematic Mapping in the History of Cartography*. University of Chicago Press, Chicago

Scott D W 1992 *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, New York

Silverman B W 1986 *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London

Simkin D, Hastie R 1987 An information processing analysis of graph perception. *Journal of the American Statistical Association* **82**: 454–65

Stigler S 1983 *The History of Statistics*. Harvard University Press, Cambridge, MA

Tilling L 1975 Early experimental graphics. *The British Journal for the History of Science* **8**: 193–213

Tufte E R 1983 *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT

Tufte E R 1990 *Envisioning Data*. Graphics Press, Cheshire, CT

Tufte E R 1997 *Visual Explanations*. Graphics Press, Cheshire, CT

Tukey J W 1977 *Exploratory Data Analysis*. Addison-Wesley, Reading, MA

Wainer H 1997 *Visual Revelations: Graphical Tales of Fate and Deception from Napoleon Bonaparte to Ross Perot*. Springer-Verlag, New York

Wainer H, Spence I 1997 Who was Playfair? *Chance* **10**: 35–7

Wainer H, Velleman P F 2001 Statistical graphs: Mapping the pathways of science. *The Annual Review of Psychology* **52**

Wilkinson L 1999a Dot plots. *The American Statistician* **53**: 276–81

Wilkinson L 1999b *SYSTAT*. SPSS, Chicago

Wilkinson L 1999c *The Grammar of Graphics*. Springer-Verlag, New York

Wilkinson L, Blank G, Gruber C 1996 *Desktop Data Analysis with SYSTAT*. Prentice-Hall, Upper Saddle River, NJ

L. Wilkinson

Graphical Models: Overview

1. Some General and Historical Remarks

Graphical models aim to describe concisely the possibly complex interrelationships between a set of variables. Moreover, from the description key, proper-

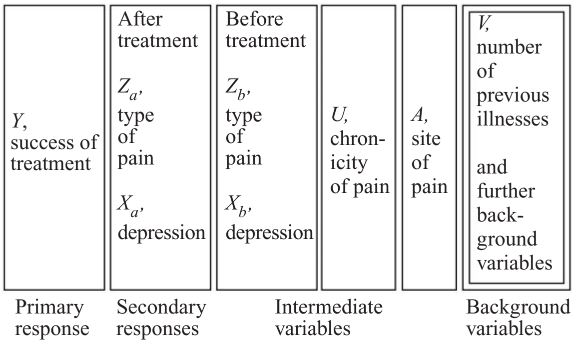

Figure 1
Initial partial ordering of variables in chronic pain study for a multivariate regression chain. Variables in the same box are treated on an equal footing. Variables in any one box are considered conditionally given all variables in boxes listed to their right. Associations among background variables in the double-lined box are not analyzed but taken as given

ties can be read directly. The central idea is that each variable is represented by a node in a graph. Any pair of nodes may be joined by an edge. For most types of graph a missing edge represents some form of independency between the pair of variables. Because the independency may be either marginal or conditional on some or all of the other variables, a variety of types of graph are needed.

A particularly important distinction is between directed and undirected edges. In the former an arrow indicates the direction of dependency from an explanatory variable to a response. If, however, the two variables are to be interpreted on an equal footing then an edge between them is undirected, or cyclic dependencies are permitted. For instance, systolic and diastolic blood pressure would typically be treated on an equal footing because they are two aspects of a single phenomenon, namely the blood pressure wave.

Graphical models started to be developed by Darroch et al. (1980) and Wermuth (1976) as special subclasses of loglinear models for contingency tables and of multivariate Gaussian distributions which are interpretable in terms of conditional independencies and can be represented by undirected graphs. See *Multivariate Analysis*: *Overview*; *Multivariate Analysis*: *Discrete Variables* (*Loglinear Models*).

The first extension was to problems in which the variables can be arranged in sequence so that each variable is regarded as a response to all variables to the right of it in the sequence. This leads to representation by a directed acyclic graph. Then there are joint response models with graphs having blocks of undirected components, the blocks being again connected in a directed and acyclic way. Much recent and ongoing research is for more specialized situations such as models for confounded dependencies, for event history data, for time series data, and for simultaneous dependencies in multivariate Gaussian distributions. The latter correspond, for instance, to independence graphs having directed cyclic components.

As an example of some of the above ideas we report here on a study concerning chronic pain. One main research question is: when investigating the development of this illness, is it necessary to include information on the patient's site of main pain? Figure 1 shows a first partial ordering of a set of variables within boxes. It reflects some of the prior knowledge on the development of chronic pain and its treatment. It also determines important aspects of the types of analysis to be carried out: relations among several background variables are taken jointly as given, univariate conditional distributions are studied when there is a single variable within a box, and joint conditional distributions are studied when there are several variables within a box. We return to a graph resulting from such statistical analyses in Sect. 6

We concentrate in this article on the use of models for the analysis and interpretation of empirical data, and mention some of their other uses in Sect. 7 on suggested further reading. The more recent statistical work quoted here by authors only is referenced in Edwards (2000), Studený and Bouckaert (1998), Richardson (1998), or Wermuth (1998).

2. *Models for Contingency Tables and the Need for Extensions*

For a q-dimensional contingency table, i.e., for counts on q categorical variables, the most complex undirected graph model has no constraints. It is called the saturated model and is represented by a graph in q vertices or nodes, with each node pair having exactly one edge, i.e., being connected by a full line. Each full line represents the conditional association of the connected variable pair given all remaining $q-2$ variables. Removing a line from the graph introduces a particular independence constraint for the corresponding variable pair: the pair are to be conditionally independent given all remaining variables. In a loglinear model formulation this is achieved by setting the two-factor and all higher-order interaction terms involving this variable pair to zero. If all full lines are removed, the graph of q disconnected nodes results. It represents the simplest model in the class, the one of mutual independence of all q variables.

This development built on previous work (quoted in monographs on graphical models) by M. S. Bartlett, M. W. Birch, L. A. Goodman, Y. Y. Bishop, S. E. Fienberg and P. Holland and much earlier work by A. A. Markov, who used the notion of conditional independence around 1900 to formulate simple multivariate models, now called Markov chains. A Markov chain for q categorical variables can be represented by an undirected graph which is a single path of full lines, i.e., a sequence of nodes $i_1, i_2, \ldots, i_q$ where just each consecutive pair is connected by a full line.

Compared to the larger class of loglinear models the graphical models are attractive for three main reasons: (a) the interpretation of a model can be much simpler if independencies are taken into account compared to only loglinear interactions; (b) parameter estimation can be strengthened by basing it on a sequence of possible small contingency tables which depends on the unique decomposition of a given graph into its prime graphs, i.e., into graphs which in a graphical sense cannot be split any further; and (c) a powerful separation criterion permits one to read all independence statements implied by a model directly off the graph.

This separation criterion for undirected graphs takes any three nonoverlapping subsets a, b, c of nodes in the graph of which the set c may be empty. Then the corresponding variables X_a are conditionally independent of X_b given X_c if every path from a to b has a node in c. This holds provided two conditions are satisfied: all pairwise independence statements specified with the missing edges of the undirected graph hold, and a technical positivity condition is fulfilled. One sufficient condition for checking the latter from an observed contingency table is that there are no zero frequencies.

In graphical models represented by undirected graphs all variables are treated on an equal footing, i.e., response variables and possible explanatory variables are not distinguished. This is appropriate for various different types of data, for instance for symptoms of a given disease, for items which are to measure slightly varying aspects of a particular attitude or of a behavior of persons, or for several aspects which all might contribute to a specific risk factor. However, for much multivariate research some response variables are of primary interest. There is in addition a set of background or context variables, and there are possibly sequences of intermediate variables which play the role of potential response variables to some and of potential explanatory variables to other variables under investigation. Typically, not all variables are only categorical; some are nominal, some are ordinally scaled, and still others are quantitative measurements. Subsets of variables may have to be considered as joint responses instead of univariate responses.

3. *Questions Regarding Applications and Statistical Research*

The original formulation of graphical models has already been extended in various directions to accommodate the need posed by different substantive research questions, to describe properties of different classes of independence models, and to prove corresponding results.

For research in different substantive fields such as genetics, economics, social, and life sciences, some important questions arising in applying graphical Markov models are:

(a) How do graphical Markov models relate to techniques applied traditionally in the given field? What additional possibilities do they offer to gain better insight into any given research question?

(b) How can a given graphical Markov model be best fitted to data? Is software available which permits fast application as well as comparison and integration into existing techniques and strategies of data analysis?

(c) Which of the associated new results in statistical theory are most useful for the substantive research at hand? Are case studies available in which graphical Markov models have been used fruitfully?

(d) Are criteria available to decide among alternative types of model to be considered for a given set of data?

The answers must depend partly on the specific substantive research questions. Also, the kinds of research questions which may be analyzed in the context of graphical Markov models widen as more theoretical results become available, and as new classes of model are developed. But many positive answers have become available, especially since, as described in Sect. 2 for general loglinear models, representations interpretable in terms of conditional independencies often result as a subclass of widely used tools for data analysis.

The following are some of the associated research questions in statistical theory.

(a) Which sets of independence statements can be specified within different classes of independence graphs? In particular, what are the defining independencies, i.e., what are (possibly alternative) sets of independence statements by which a given graph can be characterized? Which independence statements are implied in addition to those in a defining set?

(b) Under what conditions are two different graphs independence-equivalent; that is, when do they characterize the same independence structure, so that they imply an identical list of independence statements?

(c) Under what distributional assumptions do all models in a given class of independence graphs exist? When are they not only independence-equivalent but when are they also distribution-equivalent; that is, when do they define the same joint distribution?

(d) For which graphical Markov models do unique maximum-likelihood estimates exist? With which types of algorithm can maximum-likelihood estimates or parameter estimates in posterior distributions be obtained? When can large estimation problems be decomposed into smaller, simpler ones? When can complicated estimation problems be embedded in larger, simpler ones?

(e) Can the implied independence structure be predicted that results, for instance, after marginalizing over some of the variables and after conditioning on other variables? When can even the strength and direction of implied conditional associations be derived explicitly?

(f) How can deterministic transformations of variables be utilized systematically in model building? Should definitions of variables, i.e., measurement questions, be integrated into the formulation of a graphical Markov model, or should they be introduced and studied separately?

While the last set of questions involve ongoing research, at least partial answers are available to all the others for most subclasses of graphical Markov models known until now. Here we give only examples of results, and point to the researchers who developed them.

4. *Types of Independence Graph and Corresponding Markov Models*

As mentioned before, there are a number of different types of graph which differ with respect to their defining sets of independencies. To describe them we denote by Y_V a random vector variable of d_V individual components. A graph for the node set V captures an independence structure for the joint distribution of all d_V components. For nonoverlapping subsets S and C of V, with S having at least two nodes, a graph for the node set S gives an independence structure for the joint conditional distribution of Y_S, given Y_C. When C is empty the graph is for the joint marginal distribution of Y_S. Conditional independence of Y_i and Y_j, given Y_C, is written more compactly as $i \perp\!\!\!\perp j \mid C$.

For each member of the following types of graph, there exists the corresponding independence structure for mixtures of discrete and continuous variables in some exponential family distribution, defined either locally for conditional Gaussian regressions, or globally for a general conditional Gaussian distribution or for one having no higher than two-factor interactions. For simplicity we illustrate here the different types of model only for a mean-centered vector variable Y_V having a joint Gaussian distribution and being partitioned into at most three components $V = (a, b, c)$. The overall covariance matrix as well as its inverse, the overall concentration matrix, are written accordingly, partitioned as

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} & \Sigma_{ac} \\ \cdot & \Sigma_{bb} & \Sigma_{bc} \\ \cdot & \cdot & \Sigma_{cc} \end{pmatrix}, \quad \Sigma^{-1} = \begin{pmatrix} \Sigma^{aa} & \Sigma^{ab} & \Sigma^{ac} \\ \cdot & \Sigma^{bb} & \Sigma^{bc} \\ \cdot & \cdot & \Sigma^{cc} \end{pmatrix} \tag{1}$$

where the dots denote symmetric entries.

The concentration matrix in the marginal distribution of Y_c is denoted by $\Sigma^{cc\cdot ab}$, and the one in the joint marginal distribution of Y_b and Y_c has components $\Sigma^{bb\cdot a}$, $\Sigma^{bc\cdot a}$, $\Sigma^{cc\cdot a}$. In the latter, the notation reminds one that marginalizing is over a. Similarly, the conditional covariance matrix of Y_a is denoted by $\Sigma_{aa\cdot bc}$, and the conditional covariance matrix of Y_a and Y_b given Y_C has components $\Sigma_{aa\cdot c}$, $\Sigma_{ab\cdot c}$, $\Sigma_{bb\cdot c}$. Here the notation reminds one that conditioning is on c. Finally, the matrix of regression coefficients of Y_c when regressing Y_b on Y_c is $\Pi_{b|c}$, while in the regression of Y_a on both Y_b and Y_c the regression coefficient matrix is in partitioned as $(\Pi_{a|b.c}, \Pi_{a|c.b})$. For example, $\Pi_{a|b\cdot c}$ are the coefficients of Y_b. The notation reminds one that regression is also on c.

4.1 *Directed Acyclic Graph*

A *directed acyclic graph* of Y_V is a graph of arrows in d_V nodes without directed cycles, i.e., starting from any one node it is impossible to return to this node by following any path in the direction of the arrows. The (i, j) arrow is missing in it if

$$i \perp\!\!\!\perp j \mid \quad \text{parents of } i. \tag{2}$$

Nodes from which an arrow points directly to node i are called the parents of i.

For a Gaussian distribution, the independencies show as zero coefficients in recursive linear equations having independent residuals. More precisely, we have equations $AY_V = \varepsilon$, with residuals ε having zero mean and covariance matrix, cov (ε), being a diagonal matrix T. Further, A is upper triangular with ones along the diagonal, giving $\Sigma^{-1} = A^T T^{-1} A$. Therefore (A, T^{-1}) is a triangular decomposition of the concentration matrix and off-diagonal elements in row i and A are negative values of regression coefficients when regressing Y_i on all its potentially explanatory variables, i.e., variables ordered to have indices larger than i. Thus, the independencies show as zeros in A. If only a directed acyclic graph is given, typically more than one set of recursive equations is compatible with it, i.e., with several different orderings of the variables, an upper triangular matrix can result which reflects the same independence structure as the given graph.

However, often a unique full ordering is provided from substantive knowledge about how the data could have been generated. This was how the geneticist S. Wright introduced and used path diagrams for recursive linear relations. For other than linear relations, a generating process in terms of univariate conditional distributions still determines uniquely the edge matrix in a directed acyclic graph and a corresponding factorization of the joint density. Separation criteria for directed acyclic graphs apply therefore to any joint distribution generated in this way over a directed acyclic graph. A first path criterion was given by J. Pearl, who called it d-separation (for separation in directed graphs). Two other equivalent criteria are due to S. L. Lauritzen and coworkers, and D. R. Cox and N. Wermuth.

4.2 Types of Undirected Graph

A *conditional concentration graph* of Y_S, given Y_C, is an undirected graph of full lines in d_S nodes. The full (i, j) line is defined to be missing in it if

$$i \perp\!\!\!\perp j \mid C, S \quad \text{excluding } i, j. \tag{3}$$

For a Gaussian distribution, the independencies show as zeros in the conditional concentration matrix $\Sigma^{-1}{}_{SS.C}$. Such models have been introduced by A. P. Dempster under the name of covariance selection. The name concentration graph model for Gaussian variables is more common now. It records that for Gaussian distributions the above independence statement is equivalent to the vanishing of a concentration, which is a multiple of the partial correlation of Y_i and Y_j given all remaining variables in S and C.

The name concentration graph model does not imply that for general distributions concentrations vanish, but only that the distribution satisfies a set of independence statements of the above type. The separation criterion in Sect. 2 holds for such general concentration graph models and was given by J. Darroch, S. L. Lauritzen and T. P. Speed.

A *conditional covariance graph* of Y_S, given Y_C, is an undirected graph of dashed lines in d_S nodes. The dashed (i, j) line is missing in it if

$$i \perp\!\!\!\perp j \mid C. \tag{4}$$

For a Gaussian distribution, the independencies show as zeros in the conditional covariance matrix $\Sigma_{SS.C}$. The models are a special case of T. W. Anderson's linear in covariance models. They may not have a unique maximum likelihood estimate.

In general a covariance graph structure can capture all independencies in a joint distribution only if it has, just as a Gaussian distribution, no higher than two-factor interactions. For binary variables this holds, for instance, for the quadratic binary exponential distribution. For mixed variables it can be achieved for the conditional Gaussian distribution with all three and higher-order interactions being zero. Separation criteria were given by G. Kauermann and by Richardson (2001).

4.3 Types of Chain Graph

A *chain graph* of Y_V is a graph in d_V nodes which may be arranged in a sequence of ordered boxes $(1,\dots, k, k+1,\dots, K)$ such that there are undirected edges within boxes and arrows between boxes all pointing in one direction, e.g., from k to $1,\dots, k-1$; the arrows form no directed cycle, i.e., starting from any one box it is impossible to return to this box by following the direction of the arrows.

A *chain graph of block regressions* contains full lines within boxes and full arrows between boxes. The full (i, j) line is missing within box k or the full (i, j) arrow pointing from a node in one of the boxes $k+1,\dots, K$ to a node in box k is missing if

$$i \perp\!\!\!\perp j \mid \text{ all nodes in boxes } k,\dots, K \text{ excluding } i, j. \tag{5}$$

For three chain components in a Gaussian distribution, each independence statement corresponds to a zero entry (possibly symmetric) in one of the following concentration matrices: Σ^{aa}, Σ^{ab}, Σ^{ac}, $\Sigma^{bb\cdot a}$, $\Sigma^{bc\cdot a}$, $\Sigma^{cc\cdot ab}$. The name block regression arises in spite of this concentration matrix representation since dividing all elements in row i of, say $(\Sigma^{bb\cdot a}, \Sigma^{bc\cdot a})$, by the element in position (i, i), a transformation into corresponding block-regression coefficients results. Two different equivalent separation criteria were given by M. Frydenberg and by M. Studený.

Models of this type for the special kind of model for mixtures of continuous and discrete variables called CG-regressions are known to have unique maximum likelihood estimates, but it took more than 20 years before an algorithm was made available by D. Edwards and S. L. Lauritzen that converges in reasonable time.

A *chain graph of multivariate regressions* contains dashed lines within boxes and dashed arrows between boxes. The dashed (i, j) line is missing within box k if

$$i \perp\!\!\!\perp j \mid \text{ all nodes in boxes } k+1,\dots, K \tag{6}$$

the full (i, j) arrow pointing from a node in one of the boxes $k+1,\dots, K$ to a node in box k is missing if

$$i \perp\!\!\!\perp j \mid \text{ all nodes in boxes } k+1,\dots, K \text{ excluding } j. \tag{7}$$

For three chain components in a Gaussian distribution each independence statement corresponds to a zero entry (possibly symmetric) in one of the following matrices of covariances or of regression coefficients: $\Sigma_{aa\cdot bc}$, $\Pi_{a|b\cdot c}$, $\Pi_{a|c\cdot b}$, $\Sigma_{bb\cdot c}$, $\Pi_{b|c}$, Σ_{cd}. A separation criterion was given by N. Wermuth and D. R. Cox and by T. S. Richardson.

A *chain graph of concentration regressions* is a mixture of a concentration graph and a multivariate regression graph. It contains full lines within boxes and dashed arrows between boxes. The full (i, j) line is missing within box k if

$$i \perp\!\!\!\perp j \mid \text{ all nodes in boxes, } k,\dots, K \text{ excluding } i, j \tag{8}$$

the dashed (i, j) arrow pointing to box k is missing if

$$i \perp\!\!\!\perp j \mid \text{ all nodes in boxes } k+1,\dots, K \text{ excluding } j. \tag{9}$$

For three chain components in a Gaussian distribution, each independence statement corresponds to a zero entry (possibly symmetric) in one of the following matrices of concentrations or of regression coefficients: Σ^{aa}, $\Pi_{a|b\cdot c}$, $\Pi_{a|c\cdot b}$, $\Sigma^{bb\cdot a}$, $\Pi_{b|c}$, $\Sigma^{cc\cdot ab}$. For

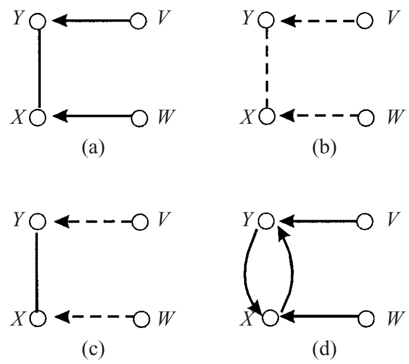

Figure 2
Simple examples of the four types of joint response graphs in Sects. 4.3 and 4.4. In each the joint responses are Y and X, and the explanatory variables are marginally independent: $V \perp\!\!\!\perp W$. The only additional independencies represented by each graph are in the (a) chain graph of block regressions: $Y \perp\!\!\!\perp W \mid (X, V)$ and $X \perp\!\!\!\perp V \mid (Y, W)$, (b) chain graph of multivariate regressions: $Y \perp\!\!\!\perp W \mid V$ and $X \perp\!\!\!\perp V \mid W$, (c) chain graph of concentration-regressions: $Y \perp\!\!\!\perp W \mid V$ and $X \perp\!\!\!\perp V \mid W$; and (d) cyclic graph: $V \perp\!\!\!\perp W \mid (Y, X)$.
Graphs (b) and (c) are examples of two independence equivalent graphs. In fact a distinction in independencies would emerge only with a third response Z in the left-hand box. Then a missing edge between, say X and Z, in (c) would refer to independency conditional on Y, V, W, and in (b) conditional only on V, W

exponential families the concentration graph components refer to canonical parameters and the arrows correspond to (conditional) mean parameters. One intensively studied special case is for categorical data in which joint responses are modeled by loglinear interactions and dependencies via so-called marginal models. A separation criterion was given by S. A. Andersson, D. Madigan and M. Perlman.

Of these three types of chain graph only the second class, the graph of multivariate regressions, contains the general linear models in which every regression is of each component of the response separately on the same set of explanatory variables and the residual covariance matrix contains the remaining parameters. The class also contains seemingly unrelated regressions.

4.4 *Cyclic Independence Graphs*

Some fully directed graphs may contain directed cycles among a set of different variables and even for the same pair of variables. It has been shown by P. Spirtes and by J. Koster that the d-separation criterion can be applied to these graphs in unchanged form. Graphs of traditional structural equation models may be slightly modified (Koster 1999) to have in our notation dashed lines, arrows, or combinations of both, as well as cyclic dependencies. Then a slightly extended version of the d-separation criterion can be applied to them as well. However, missing edges in these graphs need not point to any independence statement for the variable pair involved, and the meaning of edges in terms of distributional properties may have to be derived from scratch for a given graph.

Figure 2 shows examples of graphs for all four types of joint response models of Sects. 4.3 and 4.4. They have in common two connected joint responses, two marginally independent explanatory variables, and missing edges for two further pairs: (X, V) and (Y, W).

5. *Independence Graphs and Generating Processes*

As mentioned above, multivariate models which have a univariate recursive generating process can be expressed in a stepwise fashion in terms of univariate conditional distributions. They are sometimes called association models generated over a directed acyclic graph. Because they can be compatible with causal interpretations they are attractive for much substantive research driven by causal hypotheses. See *Causal Inference and Statistical Fallacies.*

Multivariate regression chains are closest to a recursive generating process, since residual association may be regarded as a secondary feature. Also, after replacing each dashed line separately by two arrows starting from a common additional node, called a synthetic common source node, a directed acyclic graph results which implies the correct independence structure for the variables, in the given multivariate regression chain. It also implies covariance graph components which do not have interactions involving more than two variables.

In general, no similar simple generating processes are available for the other two types of chain graph containing chordless undirected cycles. The reasons were that (a) it is not possible to order the nodes in a chordless cycle of a concentration graph to generate it recursively in the given variables and (b) it is also not possible to enlarge the graphs by synthetic common sake nodes to generate undirected chordless cycles. At present the only ways known to generate such cycles are by conditioning on synthetic common responses or by postulating a dynamic process observed cross-sectionally in statistical equilibrium.

Experience suggests that multivariate regression chain structures fit well data from panel studies in several waves. Also, they permit prediction of individual responses without involving response variables considered on an equal footing, but only variables prior to all the joint responses. For many applications in medicine and social science this is an explicit goal of the investigation.

6. A Chain Graph as Partial Summary of Data Analysis

We now return to the variables in Fig. 1. The multivariate regression chain in Fig. 3 is derived by statistical analyses based on data provided by Judith Kappesser of the Department of Psychology at the University of Mainz. The data are for 201 chronic pain patients who have been given a three-week stationary treatment at a chronic pain clinic.

The response of primary interest in Fig. 1 is self-reported success of treatment. It is measured three months after discharge by a score which comprises several aspects of successful treatment. The background or context variables are age, gender, marital status, years of formal schooling, number of previous other illnesses, and duration of pain. There are a number of intermediate variables. Before and directly after treatment, questionnaire scores are available of depression and of self-reported type of pain, ranging from 'no pain' to 'pain as strong as imaginable.' The chronification index is a score incorporating different aspects of time, spreading of pain, use of pain relievers, and the patient's pain treatment history. The main site of pain has two categories: 'back pain' and pain on the 'head, face, or neck.'

The chain graph of Fig. 3 summarizes some important aspects of the results of analysis. It shows, in particular, which of the variables are needed for each response such that adding one more of the potentially explanatory variables does not improve prediction. Site of pain is an important intermediate variable and information on it should therefore be included in studies of chronic pain.

Some of the directions and types of dependency which cannot be read off the graph are as follows. Patients with many years of formal schooling (13 years or more) are more likely to be headache patients, while the others are more likely to be backache patients, while possibly because more of them have jobs involving physical work, and they are more likely to reach higher stages of intensity of pain before treatment. Backache patients reach higher stages of chronification and report higher intensity of pain after treatment. Figure 4 shows that treatment is more successful the lower the intensity of pain, but for scores of intensity of pain of about six or higher this dependence vanishes. One path of development is that patients with shorter periods of formal schooling may get chronic backache, and patients with chronic backache get help too late and respond less well to the type of treatment offered.

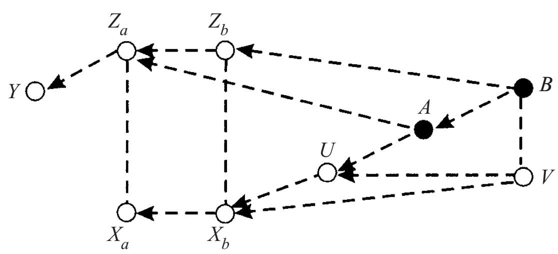

Figure 3
Multivariate regression chain of significant relations, well compatible with the data. Discrete variables are drawn as full circles, continuous ones as open circles

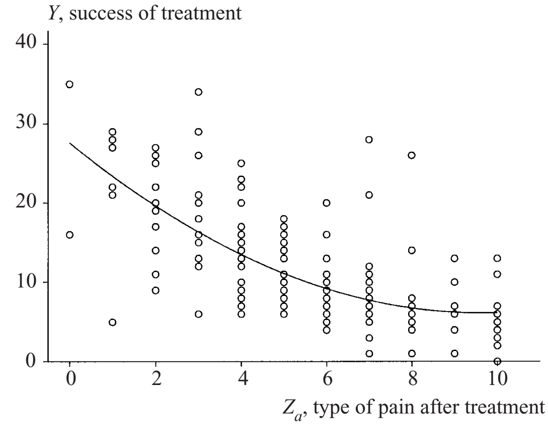

Figure 4
Form of dependence of primary response Y on Z_a

Screening for nonlinear relations and interactive effects did not point to important other relations. The model is said to fit the data well because for each response taken separately no indication was found that adding a further variable would improve prediction. Had there been no nonlinear relations and no categorical variables as responses the overall model fit could have been tested within the context of available software for models of linear structural relations.

7. Suggested Further Reading

We have described some of the statistical aspects of graphical Markov models. Statistical monographs documenting the development are by Wermuth (1978), Whittaker (1990), Edwards (2000), Lauritzen (1996), Cox and Wermuth (1996). For an exposition of important different types of independence structure present in small data sets containing only linear relations, see Cox and Wermuth (1993). Models for single and joint response chain graphs can be viewed as a generalization of Sewall Wright's path analysis different from the generalization to models of linear structural relations (Bollen 1989). The two model classes have some subclasses in common, for instance linear systems represented by directed acyclic graphs and linear multivariate regression chains.

Monographs treating the use of graphical Markov models as a basis for probabilistic calculations in expert systems and for artificial intelligence applications are by Pearl (1988) and by Cowell et al. (1999). For their use in a specifically decision-making context, see Oliver and Smith (1990). For discussions in relation to causal reasoning, see the monographs by Spirtes et al. (1993) and by Pearl (2000).

See also: Bayesian Graphical Models and Networks; Graphical Methods: Presentation

Bibliography

Bollen K A 1989 *Structural Equations with Latent Variables*. Wiley, New York

Cowell R G, Dawid A P, Lauritzen S L, Spiegelhalter D J 1999 *Probabilistic Networks and Expert Systems*. Springer, New York

Cox D R, Wermuth N 1993 Linear dependencies represented by chain graphs (with discussion). *Statistical Science* **8**: 204–18

Cox D R, Wermuth N 1996 *Multivariate Dependencies*. Chapman and Hall, London

Darroch J N, Lauritzen S L, Speed T P 1980 Markov fields and log-linear models for continuous tables. *Annals of Statistics* **8**: 552–39

Edwards D 2000 *Introduction to Graphical Modelling*, 2nd edn. (1st edn. 1995) Springer, New York

Koster J T A 1999 On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors. *Scandinavian Journal of Statistics* **26**: 413–31

Lauritzen S L 1996 *Graphical Models*. Oxford University Press, Oxford, UK

Oliver R M, Smith J Q 1990 *Influence Diagrams, Belief Nets and Decision Analysis*. Wiley, Chichester

Pearl J 1988 *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, CA

Pearl J 2000 *Causality: Models, Reasoning and Inference*. Cambridge University Press, Cambridge, UK

Richardson T S 1998 Chain graphs and symmetric associations. In: Jordan M I (ed.) *Learning in Graphical Models*. Kluwer, Dordrecht, The Netherlands, pp. 231–60

Richardson T S 2001 A local Markov property for acyclic directed mixed graphs. *Scandinavian Journal of Statistics*.

Spirtes P, Glymour C, Scheines R 1993 *Causation, Prediction and Search*. Springer, New York

Studený M, Bouckaert R R 1998 On chain graph models for description of conditional independence structures. *Annals of Statistics* **26**: 1434–95

Wermuth N 1976 Analogies between multiplicative models in contingency tables and covariance selection. *Biometrics* **32**: 95–108

Wermuth N 1978 *Zusammenhangsanalysen medizinischer Daten*. Springer, Heidelberg, Germany

Wermuth N 1997–99 Graphical Markov models. In: Kotz S, Read C, Banks D (eds.) *Encyclopedia of Statistical Sciences*. Wiley, New York, Vol. 2, pp. 284–300

Whittaker J 1990 *Graphical Models in Applied Multivariate Statistics*. Wiley, Chichester, UK

N. Wermuth and D. R. Cox

Green Parties

Green parties have become established political players in most industrialized democracies and are emerging in several industrializing ones. The electoral strides of green parties, especially in Western Europe, have resulted in a situation unimaginable a decade ago: green parties in many national parliaments, green ministers sharing power in several European national governments, greens in the European Parliament forming one of the largest party groups, as well as sustained green representation at subnational levels of government. Yet greens' electoral fortunes are decidedly uneven: electoral success in northern countries contrasts with near invisibility in the South. Moreover, electoral success has brought new strategic dilemmas for the greens as they struggle to maintain their alternative 'green' credentials while joining parliaments and supporting, if not forming, national governments.

1. Emergence

The emergence of green parties needs to be understood in the context of a broader shift from industrial to 'postindustrial' politics. For younger, better-educated citizens in particular, the prosperity and rise in living standards enjoyed in the postwar era formed the backdrop for a change from materialist to 'postmaterialist' values, or from 'old politics' to 'new politics' (see Inglehart 1977, Poguntke 1993). As the basic material needs (food, shelter, etc.) of a far higher share of the population were being satisfied, political attention shifted from materialist concerns to 'quality of life' issues, such as enhanced political participation, gender and racial equality, and, perhaps above all, environmental protection.

Growing environmental concern among the public was heightened by the continued degradation of the environment and increased media attention of environmental issues. Toxic chemical spills in major rivers, air pollution alerts in cities, and oil slicks in pristine coastal waters further increased public anxiety, especially in Europe and the USA. Meanwhile, the accelerated construction of nuclear plants in the wake of the energy crisis, as well as the increased deployment of nuclear weapons, awakened or reinforced fears about the safety of nuclear power, the problems of disposing nuclear waste, and fears of nuclear annihilation.

To many citizens in advanced democracies, none of these concerns was adequately addressed by existing parliamentary structures. A proliferation of extra-parliamentary, nonpartisan citizens' movements emerged in the 1960s and 1970s to protest issues of peace, ecology, and women's rights. These new social movements served as the antecedents for a new type of

 ISBN: 0-08-043076-7