Biometrika (2000), **87**, 1, *pp.* 206–212

Printed in Great Britain

On the generation of the chordless four-cycle

By D. R. COX
Nuffield College, Oxford OX1 1NF, U.K.
david.cox@nuf.ox.ac.uk

and NANNY WERMUTH
Center of Survey Research (ZUMA), Postfach 122155, Mannheim, Germany
wermuth@zuma-mannheim.de

Summary

In the theory of Markov graphical representations of conditional independencies a special role is played by the chordless four-cycle, representing for four random variables the conditional independencies $X \perp\!\!\!\perp V \mid (U, W)$ and $W \perp\!\!\!\perp U \mid (X, V)$. It is not immediately clear how such systems are to be generated. Here we sketch some possible data-generating mechanisms.

Some key words: Concentration graph; Conditional independence; Covariance graph; Markov graph; Stochastic differential equation.

1. Introduction

So-called full line concentration graphs represent a set of random variables by the vertices of an undirected graph. That is, some, but in general not all, pairs of vertices are joined by edges and a missing edge between, say, vertices i and j implies that the corresponding random variables are conditionally independent given all remaining variables. If the joint distribution is multivariate Gaussian a missing edge corresponds to a zero in the concentration matrix, i.e. in the inverse covariance matrix, thus corresponding to the covariance selection models of Dempster (1972). The relation between a covariance matrix Σ of a random vector Y and the interpretation of the concentration matrix Σ^{-1} in terms of partial correlations is most directly seen (Cox & Wermuth, 1996, p. 69) by showing that the random vector $\Sigma^{-1}Y$ has covariance matrix Σ^{-1} and that its cross-covariance matrix with Y is the identity matrix, leading to an interpretation of the off-diagonal elements of Σ^{-1} as proportional to partial regression coefficients.

A general theory of fitting concentration graphs for Gaussian models is given by Speed & Kiiveri (1986) and for log-linear models for discrete variables by Darroch, Lauritzen & Speed (1980) and described more generally by Lauritzen (1996). For the connection between log-linear models and covariance selection, see Wermuth (1976).

In many cases it is possible to assign a direction to each edge leading to a directed acyclic graph and, better still for interpretation, to a univariate recursive regression graph, the new graphs representing the same set of conditional independencies as the given undirected graph (Wermuth, 1980; Cox & Wermuth, 1996, Ch. 2; Wermuth & Cox, 1998). A univariate recursive regression representation sets out the variables sequentially with Y_j considered conditionally on $Y_{j+1}, \ldots, Y_p$, each missing edge in the graph corresponding to just one conditional independency in such a system. If such a representation of the undirected graph exists it is typically not unique. Such forms are valuable partly because they indicate potential generating processes which may be confirmation of or suggestive of valuable subject-matter interpretations.

The condition that such a representation is possible is that the concentration graph has no chordless m-cycle ($m \geqslant 4$). Thus the simplest concentration graph not consistent with a univariate recursive regression is the chordless four-cycle. An example where such a graph is strongly indicated empirically as the simplest representation of the data is given in Table 1, as noted by Cox & Wermuth (1993) using data of Spielberger, Russel & Crane (1983). It gives the estimated correlations and partial correlations, the latter being directly derived from the sample concentration matrix.

Table 1. *Correlations among four psychological variables for* 684 *students. Marginal correlations in lower triangle. Partial correlations given other two variables in upper triangle*

	X	W	U	V
X, state anxiety	1	0·45	0·47	−0·04
W, state anger	0·61	1	0·03	0·32
U, trait anxiety	0·62	0·47	1	0·32
V, trait anger	0·39	0·50	0·49	1

Despite the simplicity of the structure, it is puzzling for interpretation in the absence of a potential generating process. That is, there is no sequence of univariate linear regression relations in the four variables that would represent directly the set of conditional independencies observed. Here we outline several possible data generating processes. We make no claim that they necessarily correspond to the illustrative data. They are intended as general explanations of this kind of data.

For some purposes it is reasonable to replace the chordless four-cycle of Fig. 1(a) by the Markov equivalent version of Fig. 1(b) in which (U, V) as trait variables are regarded as explanatory to (W, X) as state variables and in which two of the edges are therefore regarded as directed. We deal with Gaussian variables for simplicity and arrange that all random variables have zero mean.

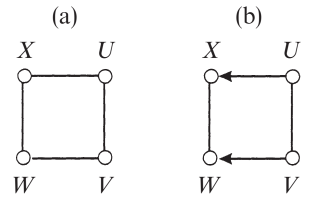

Fig. 1. (a) Chordless four-cycle; independencies $X \perp\!\!\!\perp V \mid U, W$, $W \perp\!\!\!\perp U \mid X, V$. (b) Markov equivalent chain graph in which U, V are explanatory to W, X.

2. Explanation via selection

We supplement the observed random variables by two latent variables, ξ, η represented by the nodes of the special graph of Fig. 2. In terms of linear relations we have that

$$U = \beta_{U\xi}\xi + \varepsilon_{U.\xi}, \quad V = \beta_{V\xi}\xi + \varepsilon_{V.\xi}, \quad W = \beta_{WV}V + \varepsilon_{W.V}, \quad X = \beta_{XU}U + \varepsilon_{X.U},$$

$$\eta = \beta_{\eta W.X}W + \beta_{\eta X.W}X + \varepsilon_{\eta.WX},$$

where the ε's are independently normally distributed with zero mean and the β's are all nonzero. This is a simple univariate recursive system.

Suppose now that we marginalise over the distribution of ξ and condition on the value of η. The first step induces a correlation between U and V and the second a conditional correlation between W and X given U and V. No other edge is introduced and, with the exception of very

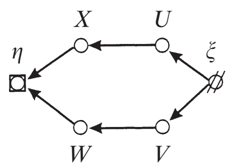

Fig. 2. Model with two latent variables, one, ξ, marginalised and the other, η, conditioned on, thus producing chordless four-cycle in observed variables U, V, W, X.

particular parameter values, no additional independency results, that is there is no parametric cancellation; for a further discussion of parametric cancellation, see Wermuth & Cox (1998). Thus a chordless four-cycle has been achieved. These results have been used previously by Wermuth (1980) and Pearl (1988, p. 118) and follow, for instance, from the general procedure for marginalising and conditioning in directed acyclic graphs, see an unpublished report by N. Wermuth, D. R. Cox and J. Pearl, or in this special case can be derived by direct calculation with the 4×4 covariance matrix (X, W, U, V) and its inverse.

In particular, with the standard notation for partial correlation coefficients,

$$\rho_{WX.\eta} = (\rho_{WX} - \rho_{W\eta}\rho_{X\eta})\{(1-\rho_{W\eta}^2)(1-\rho_{W\eta}^2)\}^{-\frac{1}{2}},$$

we have that $\rho_{WX} = 0$, $\rho_{W\eta} \neq 0$, $\rho_{X\eta} \neq 0$, implies that $\rho_{WX.\eta} \neq 0$. We apply this last result conditioning all the correlations also on (U, V). This shows that an edge is indeed induced between W and X by conditioning on η. Similar arguments show that in general no new edge for (X, V) or for (W, U) is introduced.

The representation of the dependence between U and V via an unobserved common explanatory variable is a common and plausible device. The notion of an unobserved conditioned upon response, η, is less familiar. It can, however, be taken as corresponding to a selection from a larger target population giving only those members of the larger population that show a particular value of the response η. In an unpublished Aalborg report S. L. Lauritzen has given some more general results on selection.

3. A stochastic process

3·1. *General formulation*

We now discuss several related but distinct interpretations based on a linear stochastic formulation. We start with a $p \times 1$ vector Y of response variables and a $q \times 1$ vector of explanatory variables, Z. Suppose that Z is constant in time but that the components of $Y(t)$ change in accordance with a linear system forced by a stochastic innovation process

$$dY_r(t) = \sum_{s=1}^{p} a_{rs} Y_s(t)\, dt + \sum_{j=1}^{q} b_{rj} Z_j\, dt + d\zeta_r(t), \tag{1}$$

where A, B are constant matrices with elements a_{rs}, b_{rj} and $d\zeta$ is a $p \times 1$ vector of stochastic innovations of zero mean and uncorrelated with the current value $Y(t)$ and with Z.

We discuss two different possibilities in § 3·2 and a further one in § 3·4.

3·2. *Two rather static versions*

We first follow Fisher (1970) although he worked in discrete time; a few details are formally simpler in continuous time. Suppose that A is a stability matrix (Bellman, 1997, p. 251), i.e. that its eigenvalues are either negative or if complex have negative real parts. This is necessary for statistical equilibrium, which we assume to be possible. If we cumulate over a long time period the left-hand side of (1) will be small compared with the right-hand side and there results the structural equation model

$$0 = AY + BZ + \varepsilon, \tag{2}$$

where now Y, Z, ε are time-aggregates, or averages, and the innovation term cumulated over time, ε, say, has zero mean, covariance matrix $\Sigma_{\varepsilon\varepsilon}$, say, and is independent of Z.

For a formal justification of (2) a more explicit formulation is needed. If we simply integrate (1) over a long time period t_0, the system being in statistical equilibrium, the left-hand side is $O_p(1)$, the first and third terms on the right-hand side are $O_p(\sqrt{t_0})$ and the second term is $O_p(t_0)$. Thus in the limit the second term on the right-hand side dominates and the time-aggregate Y becomes degenerate. To find the properties of a system observed for a time long compared with the damping time of the process and such that all three terms in (2) are comparable we thus embed the system in a family in which the matrix B is $O(1/\sqrt{t_0})$ as t_0 increases.

Post-multiply (2) by Z and take expectations. Then

$$0 = A\Sigma_{YZ} + B\Sigma_{ZZ},$$

where Σ_{YZ}, Σ_{ZZ} are respectively the covariance matrix of Y and Z and of Z. Further

$$Y = -A^{-1}(BZ + \varepsilon)$$

so that the covariance matrix of Y is

$$A^{-1}B\Sigma_{ZZ}B^{\mathrm{T}}(A^{-1})^{\mathrm{T}} + \mathrm{A}^{-1}\Sigma_{\varepsilon\varepsilon}(A^{-1})^{\mathrm{T}}.$$

Missing edges in the concentration graph of (Y, Z) correspond to zeros in the concentration or inverse covariance matrix of (Y, Z). The standard formula for the inverse of a partitioned matrix shows that the cross-concentration of (Y, Z) is $A^{\mathrm{T}}\Sigma_{\varepsilon\varepsilon}^{-1}B$. In particular, the condition for a missing edge between a Y and a Z component is the vanishing of the corresponding matrix element.

For a second interpretation suppose that the system (1) is subject to a step-function shock of amount ε constant for a long duration. The response will initially have a time-varying term. It will then come to equilibrium at a value of Y satisfying

$$0 = AY + BZ + \varepsilon$$

and the previous discussion applies. Each realisation of the system, for example each new subject in the psychological context, has an independent and constant innovation ε; see the unpublished Carnegie-Mellon doctoral thesis of T. Richardson for some other discussion of Fisher's process.

3·3. *A chordless four-cycle*

We now consider the special case of the chordless four-cycle in which the component matrices in all the above representations are 2×2. In the notation of § 2, we would have $Y = (X, W)$, $Z = (U, V)$. We shall assume that

$$\Sigma_{\varepsilon\varepsilon} = \mathrm{diag}\{\mathrm{var}(\varepsilon_1), \mathrm{var}(\varepsilon_2)\}.$$

Then it follows from the form of the cross-covariance matrix of (Y, Z) that the edge between Y_1 and Z_2 is missing if and only if

$$b_{12}/b_{22} + \{a_{21}\,\mathrm{var}(\varepsilon_1)\}/\{a_{11}\,\mathrm{var}(\varepsilon_2)\} = 0.$$

It aids interpretation to strengthen the condition on the eigenvalues of A by imposing the requirement that $a_{11} < 0$. To simplify the notation by working with positive quantities we therefore write $a_{ii} = -a'_{ii}$. Furthermore we choose standardised units such that the unit of time ensures that $a'_{11} = \alpha$, $a'_{22} = 1/\alpha$, the units of Y_1, Y_2 are such that $\mathrm{var}(\varepsilon_1) = \mathrm{var}(\varepsilon_2) = 1$ and the units of Z such that $b_{11} = b_{22} = 1$. If $\alpha = 1$ in isolation the two components decay at the same rate. That is, if we consider two simplified systems in which $Z = 0$, $a_{12} = a_{21} = 0$ and there is no noise, then the components Y_1

and Y_2 decay at the same rate if and only if $\alpha = 1$. In these standardised units we write

$$a_{12} = \alpha_{12}, \quad a_{21} = \alpha_{21}, \quad b_{12} = \beta_{12}, \quad b_{21} = \beta_{21}.$$

The system is thus specified by the covariance matrix of Z in the standardised units, by α and by the four parameters just defined and the correlation between the components of ε.

Our condition for $Y_1 \perp\!\!\!\perp Z_2 | (Y_2, Z_1)$, that is for $X \perp\!\!\!\perp V | (W, U)$, is that $\alpha\beta_{12} = \alpha_{21}$. There is a complementary condition $\alpha^{-1}\beta_{21} = \alpha_{12}$ for the other missing edge. In words the first condition can be stated as that the rate of self-dissipation of state anger divided by the rate of transfer from state anger to state anxiety is equal to the rate of transfer from trait anxiety to state anger divided by the rate of transfer form trait anxiety to state anxiety.

3·4. *A dynamic cross-section*

For our third interpretation we suppose the innovation process to be a Brownian motion and suppose that $Y(t)$ corresponds to an observation of the process in its stationary state.

It helps to write the defining equation (1) in the form

$$Y(t + dt) = (I + A\, dt)Y(t) + BZ\, dt + d\zeta(t). \tag{3}$$

On taking expectations of $Y(t + dt)Y^{\mathrm{T}}(t + dt)$ we have in statistical equilibrium that

$$A\Sigma_{YY} + \Sigma_{YY}\mathrm{A}^{\mathrm{T}} + \mathrm{B}\Sigma_{ZY} + \Sigma_{YZ}B^{\mathrm{T}} + \Sigma_{\zeta\zeta} = 0,$$

where now $\Sigma_{\zeta\zeta}\, dt$ is the covariance matrix of the innovation.

Similarly on post-multiplying by Z^{T} and taking expectations we have that

$$\Sigma_{YZ} = -A^{-1}B\Sigma_{ZZ},$$

so that

$$A\Sigma_{YY} + \Sigma_{YY}A^{\mathrm{T}} = B\Sigma_{ZZ}B^{\mathrm{T}}(A^{-1})^{\mathrm{T}} + A^{-1}B\Sigma_{ZZ}B^{\mathrm{T}} - \Sigma_{\zeta\zeta}.$$

For the present purpose we are interested especially in the concentration matrix partitioned with sections denoted by superscripts. In particular, defining Λ_{YY} by $\Sigma_{YY.Z}$, we have that

$$\Sigma^{YY} = \Lambda_{YY}^{-1},$$

where

$$\Lambda_{YY} = \Sigma_{YY.Z} = \Sigma_{YY} - A^{-1}B\Sigma_{ZZ}B^{\mathrm{T}}(A^{-1})^{\mathrm{T}}, \tag{4}$$

$$\Sigma^{ZY} = B^{\mathrm{T}}(A^{-1})^{\mathrm{T}}\Sigma^{YY}. \tag{5}$$

Direct calculation shows that Λ_{YY} satisfies the equation

$$A\Lambda_{YY} + \Lambda_{YY}A^{\mathrm{T}} = -\Sigma_{\zeta\zeta}.$$

We note, but will not here exploit, the solution (Bellman, 1997, p. 239)

$$\Lambda_{YY} = \int_0^\infty e^{At}\Sigma_{\zeta\zeta}a^{A^{\mathrm{T}}t}\, dt.$$

We use the alternative form involving a Kronecker sum, namely

$$(A \otimes I + I \otimes A)\,\mathrm{vec}\,\Lambda_{YY} = -\mathrm{vec}\,\Sigma_{\zeta\zeta}, \tag{6}$$

essentially a set of simultaneous linear equations for the elements of Λ_{YY} then leading to an expression for Σ^{YZ}.

3·5. *Another chordless four-cycle*

We return to the special case of the chordless four-cycle. The condition for conditional independence is from (3) and (4) that

$$b_{21}/b_{11} + \{\mathrm{var}(\varepsilon_2)a_{12} - \mathrm{var}(\varepsilon_1)a_{21}\}/\{\mathrm{var}(\varepsilon_1)(a_{11} + a_{22})\} = 0.$$

In standardised units we require respectively that

$$\alpha_{12} - \alpha_{21} = (\alpha + 1/\alpha)\beta_{21}, \quad \alpha_{21} - \alpha_{12} = (\alpha + 1/\alpha)\beta_{12}.$$

In particular they are satisfied by

$$\alpha_{12} = \alpha_{21}, \quad \beta_{12} = \beta_{21} = 0.$$

This formulation in its simplified form requires only that, in the terminology of the example, trait anger feeds just into state anger and trait anxiety just into state anxiety, and that in standardised units the flows from state anger to state anxiety and vice versa are at equal rates. This in some ways is the simplest explanation directly in terms of the observed variables of all those considered here.

3·6. *A symmetrical special case*

We now explore in a little more detail the symmetrical case in which (X, U) and (W, V) can be interchanged without altering the joint distribution. Thus in standardised units $\alpha = 1$ and the adjustable parameters are

$$a_{12} = a_{21} = a, \quad b_{12} = b_{21} = b, \quad \text{var}(U) = \text{var}(V) = \sigma^2, \quad \text{corr}(U, V) = \rho.$$

Then, in the discussion in § 3·3, we have that $a = b$ and there is thus for each given Σ_{ZZ} a one-parameter family of covariance and concentration matrices having the chordless four-cycle structure. Similarly in the process of § 3·5 the condition for a chordless four-cycle is $b = 0$, leading to a different one-parameter family, emphasising the distinction between the processes.

Finally, we make, as noted in § 1, no claim that any of the above processes is indeed the generating process for the particular example. It would be interesting to know if there are other plausible types of explanation of the chordless four-cycle and other structures which cannot be transformed into an equivalent univariate recursive regression form in the observed variables.

As a check on these results a number of simulations were run of discrete time versions of these models and the requisite independence properties verified by computing the estimated covariance and concentration matrices. The calculations were programmed in MATLAB.

4. Some more constrained structures

In the above discussion we have concentrated on systems that can generate a chordless four-cycle in the concentration matrix, i.e. having two special conditional independencies and no other. We now discuss briefly two further possibilities. For simplicity we restrict ourselves to the symmetric case of § 3·6 in which (X, U) can be interchanged with (W, V).

First there is the possibility that in addition to a chordless four-cycle in concentrations there is a chordless four-cycle in covariances; i.e. that, for example, in addition to $W \perp\!\!\!\perp U \mid (X, V)$ and $X \perp\!\!\!\perp V \mid (X, U)$, there are the marginal independencies $W \perp\!\!\!\perp U$ and $X \perp\!\!\!\perp V$. In general simultaneous simplification of both covariance and concentration matrix arises only exceptionally. For an example and a formulation directly in terms of marginal correlations, see Cox & Wermuth (1993, p. 213). This structure cannot be achieved via the conditioning process of § 2.

We work with the dynamic model of § 3·4 and use the standardised units, in which $b = 0$, to achieve the property in concentrations and then evaluate the cross-covariance matrix

$$\Sigma_{YZ} = -A^{-1}B\Sigma_{ZZ}.$$

The required condition for example to achieve $W \perp\!\!\!\perp U \mid (X, V)$ is that $a + \rho = 0$. That is, the correlation, ρ, between U and V has to have the opposite sign and in standardised units to have the same magnitude as the parameter defining the rate of flow between W and X. The numerical equality is an instance of so-called parametric cancellation.

A second possibility, in some ways of more interest from an interpretational point of view, is that $U \perp\!\!\!\perp V$, that is in the general formulation that Σ_{ZZ} is diagonal. This structure cannot be achieved via the conditioning process of § 2. More generally it has been shown that this structure cannot arise from marginalising or conditioning in any directed graphical model, cyclic or acyclic (Richardson, 1998). In the symmetric case, again with $b = 0$, it can be shown that

$$\Sigma^{YY} = 2(1 - a^2)J(-a), \quad \Sigma^{YZ} = -2(1 - a^2)I, \quad \Sigma^{ZZ} = 1/\sigma^2 - 2J(a),$$

where $J(a)$ is the 2×2 matrix with diagonal elements one and off-diagonal elements a.

Thus in particular the partial correlation between W and V given X and U, obtained via the standardised off-diagonal element of Σ^{YZ}, is $(1 + 1/\sigma^2)^{-\frac{1}{2}}$, showing that positive partial correlations up to $1/\sqrt{2}$ can be achieved under this model.

Acknowledgement

It is a pleasure to thank Ingram Olkin for some valuable pointers over the matrix algebra. We are grateful also to the referees for helpful comments.

References

Bellman, R. (1997). *Introduction to Matrix Analysis*, 2nd ed. Philadelphia, PA: SIAM.

Cox, D. R. & Wermuth, N. (1993). Linear dependencies represented by chain graphs (with Discussion). *Statist. Sci.* **8**, 204–83.

Cox, D. R. & Wermuth, N. (1996). *Multivariate Dependencies*. London: Chapman and Hall.

Darroch, J. N., Lauritzen, S. L. & Speed, T. P. (1980). Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.* **8**, 522–39.

Dempster, A. P. (1972). Covariance selection. *Biometrics* **28**, 157–75.

Fisher, F. M. (1970). A correspondence principle for simultaneous equation models. *Econometrica* **38**, 73–92.

Lauritzen, S. L. (1996). *Graphical Models*. Oxford: Oxford University Press.

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

Richardson, T. (1998). Chain graphs and symmetric associations. In *Learning in Graphical Models*, Ed. M. I. Jordan, pp. 231–59. Dordrecht: Kluwer.

Speed, T.P. & Kiiveri, H. T. (1986). Gaussian Markov distributions over finite graphs. *Ann. Statist.* **14**, 138–50.

Spielberger, C. D., Russell, S. & Crane, R. (1983). Assessment of anger. In *Advances in Personality Assessment*, **2**, Ed. J. D. Butcher and C. D. Spielberger, pp. 159–87. Hillsdale, NJ: Elbaum.

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

Wermuth, N. (1980). Linear recursive equations, covariance selection, and path analysis. *J. Am. Statist. Assoc.* **75**, 963–72.

Wermuth, N. & Cox, D. R. (1998). On association models defined over independence graphs. *Bernoulli* **4**, 477–95.

[*Received March* 1999. *Revised July* 1999]