On the generation of the chordless four-cycle

Cox, D.R.; Wermuth, Nanny

On the generation of the chordless four-cycle *

D.R.Cox
Department of Statistics and Nuffield College, Oxford, UK
and
Nanny Wermuth
ZUMA-Mannheim, Germany

February 26, 1999

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,

*Version submitted to Biometrika

of the variables thus corresponding to the covariance selection models of Dempster (1972). The relation beween 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 and 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 and Kiiveri (1986) and for log linear models for discrete variables by Darroch, Lauritzen and 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 and Wermuth, 1996). 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 ($m \geq 4$). 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 and Wermuth (1993) using data of Spielberger, Russell and Crane (1983). It gives the estimated correlations and partial correlations, the latter being directly derived from the sample

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

Variables		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

concentration matrix.

Despite the simplicity of the structure, it is puzzling for interpretation in the absence of a potential generating process. Here we outline several such. We make no claim that they necessarily correspond to the illustrative data. They are intended as general explanations of this kind of data.

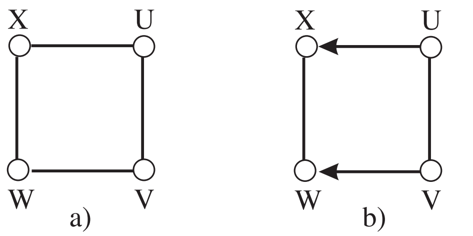

Figure 1: a) Chordless four cycle. Independencies $X \perp\!\!\!\perp V \mid (U, W)$, $W \perp\!\!\!\perp U \mid (X, V)$; b) Markov equivalent chain graph in which U, V are explanatory to X, W.

For some purposes it is reasonable to replace the chordless four-cycle of Fig.1a by the Markov equivalent version of Fig.1b in which (U, V) as trait variables are regarded as explanatory to (X, W) 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.

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.

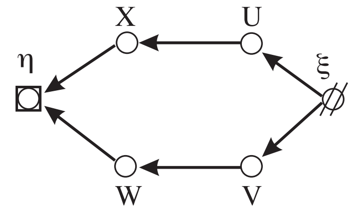

Figure 2: Model with two latent variables, one, ξ, marginalized over and the other, η, conditioned on and producing chordless four cycle in observed variables U, V, W, X.

In terms of linear relations we have that

$$\begin{aligned} U &= \beta_{U\xi}\xi + \epsilon_{U.\xi}, & V &= \beta_{V\xi}\xi + \epsilon_{V.\xi}, \\ X &= \beta_{XU}U + \epsilon_{X.U}, & W &= \beta_{WV}V + \epsilon_{W.V}, \end{aligned}$$

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

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 marginalize 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 edges are introduced and, with the exception of very particular parameter values, no edges are deleted, that is there is no parametric cancellation; for a further discussion of parametric cancellation, see Wermuth and 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 marginalizing and conditioning in directed acyclic graphs (Wermuth, Cox and Pearl, 1999) or in this special case can be derived by direct calculation with the 4×4 covariance matrix of (X, W, U, V) and its inverse.

In particular, with the standard notation for partial correlation coefficients

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

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 edges for (X, V) and (W, U) are 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 of the target population as corresponding only to those members of a larger population that show a particular response. 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 but that the components of $Y(t)$ change in accordance with a linear system forced by a stochastic innovation process

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

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

We discuss two different possibilities in Section 3.2 and a further in Section 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. 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

$$0 = AY + BZ + \epsilon,$$

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

Postmultiply 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 with Z and of Z. Further

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

so that the covariance matrix of Y is

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

Now 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 $A^T\Sigma_{\epsilon\epsilon}^{-1}B$. 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 and then will come to equilibrium at a value of Y satisfying

$$0 = AY + BZ + \epsilon$$

and the previous discussion applies. Each realization 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.

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 Section 3.2, we would have $Y = (X, W),\ Z = (U, V)$. We shall assume that

$$\Sigma_{\epsilon\epsilon} = \mathrm{diag}\{\mathrm{var}(\epsilon_1), \mathrm{var}(\epsilon_2)\}.$$

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

$$b_{12}/b_{22} + \{a_{21}\mathrm{var}(\epsilon_1)\}/\{a_{11}\mathrm{var}(\epsilon_2)\} = 0.$$

It aids interpretation to strengthen the condition on the eigenvalues of A by imposing the requirement that $a_{11} = -a'_{11} < 0$ and also to choose standardized units such that the unit of time ensures that $a'_{11} = \alpha, \quad a'_{22} = 1/\alpha$, the units of Y_1, Y_2 are such that $\mathrm{var}(\epsilon_1) = \mathrm{var}(\epsilon_2) = 1$ and the units of Z such that $b_{11} = b_{22} = 1$. If $\alpha = 1$ the two components decay on their own at the same rate. In these standardized 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 standardized units and by the four parameters just defined and the correlation between the components (U, V) of Z.

Our condition is that $\alpha\beta_{12} = 1$. In words the condition can be stated as that 'the rate of selfdissipation of state anger divided by the rate of transfer from state anxiety to state anger is equal to the rate of transfer from trait anxiety to state anxiety divided by the rate of transfer from trait anxiety to state anger.'

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

$$Y(t+dt) = (I + Adt)Y(t) + BZdt + d\zeta(t). \tag{2}$$

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

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

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

Similarly on postmultiplying by Z^T and taking expectations we have that

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

so that

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

For the present purpose we are interested especially in the concentration matrix partitioned with sections denoted by superscripts. In particular

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

where

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

and

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

Direct calculation shows that Λ_{YY} satisfies the equation

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

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

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

We use the alternative form involving a Kronecker sum, namely

$$(A \otimes I + I \otimes A)\text{vec}\Lambda_{YY} = -\Sigma_{\zeta\zeta}, \tag{3}$$

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 (2) and (3) that

$$b_{21}/b_{11} + \{\text{var}(\epsilon_2)a_{12} - \text{var}(\epsilon_1)a_{21}\}/\{\text{var}(\epsilon_1)(a_{11} + a_{22})\} = 0.$$

In standardized 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 that in standardized 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 standardized 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 of Section 3.3 we have $a = b$ 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 Section 3.5 the condition for a chordless four-cycle is $b = 0$, leading to a different one-parameter family, emphasizing the distinction between the processes.

Finally, we make, as noted in Section 1, no claim that any of the above processes are 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 involved. The calculations were programed 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 others. We now discuss briefly two further possibilities. For simplicity we restrict ourselves to the symmetric case of Section 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, in addition to $W \perp\!\!\!\perp U \mid (X, V)$ and $X \perp\!\!\!\perp V \mid (W, U)$, there are the marginal independences $W \perp\!\!\!\perp U$ 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 and Wermuth (1993, p.213).

We work with the dynamic model of Section 3.4 and use the standardized 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 is that

$$a + \rho = 0.$$

That is, the correlation, ρ, between U and V has to have the opposite sign and in standardized units 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 in the graph.

A second possibility, in some ways of more interest from an interpretational point of view, is that in addition to the chordless four-cycle in the concentration graph we have $U \perp\!\!\!\perp V$, i.e in the general formulation that Σ_{ZZ} is diagonal. This structure cannot be achieved via the conditioning process of Section 2.

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 offdiagonal elements a.

Thus in particular the partial correlation between W and V given X and U, obtained via the standardized offdiagonal element of Σ_{YZ}, is

$$(2 + 1/\sigma^2)^{-1/2},$$

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 helpful pointers over the matrix algebra.

REFERENCES

Bellman, R. (1997). *Introduction to matrix analysis.* 2nd edition. Philadelphia: SIAM.

Cox, D.R. and Wermuth, N. (1993). Linear dependencies represented by chain graphs (with discussion). *Statist. Sci.* **8**, 204-283.

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

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

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

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

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

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems.* San Mateo, Calif: Morgan Kaufmann.

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

Spielberger, C. D., Russell, S. and Crane, R. (1983). Assessment of anger. In *Advances in Personality Assessment*, J.D. Butcher and C.D. Spielberger (editors) **2**, 159-187. Hillsdale, N.J.: Erlbaum.

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**, 963-972.

Wermuth, N. and Cox, D.R. (1998). On association models defined over independence graphs. *Bernoulli* **4**, 477-495.

Wermuth, N., Cox, D.R. and Pearl, J. (1999). Explanations for multivariate structures derived from univariate recursive regressions. Submitted.