Linear Dependence represented by chain graphs

Cox and Wermuth (1993), Statistical Science, 8 (3), 204–219,

(maybe there is also discussion of this article).

Introduction section mentions:

- reduction in dimensionality from $p(p-1)/2$ correlations
- path analysis in genetics
- simultaneous equations in econometrics
- linear structural models in psychometrics
- conditional independence in expert systems

$\boldsymbol{Y}$ = vector of observations, partition into $\boldsymbol{Y}_a, \boldsymbol{Y}_b$.

Multivariate regression: Variables in $\boldsymbol{Y}_b$ are regressed on those on $\boldsymbol{Y}_b$. In general, this is same as writing

$$f_{\boldsymbol{Y}_a,\boldsymbol{Y}_b}\ f_{\boldsymbol{Y}_b} f_{\boldsymbol{Y}_\alpha|\boldsymbol{Y}_b};$$

Each of $f_{\boldsymbol{Y}_b}$ and $f_{\boldsymbol{Y}_\alpha|\boldsymbol{Y}_b}$ can be decomposed into a product of conditional probabilities. So this is a special case of a Bayesian network, where from the context, some variables come before others (time-wise, causal-wise). This extends to 3 or more components, called the *regression chain model*.

Block regression: Each component of $\boldsymbol{Y}_a$ is regressed on other components of $\boldsymbol{Y}_a$ in additional to $\boldsymbol{Y}_b$. There is no probability decomposition. If $\boldsymbol{Y}_b$ is empty, (and $\boldsymbol{Y} = \boldsymbol{Y}_a$) then each component of $\boldsymbol{Y}$ is regressed on others (simultaneous equations?), then this is the situation where the precision or concentration matrix is most relevant. It also means that the variables of interest are in $\boldsymbol{Y}$ and there is a theory for the simultaneous equations?

For multivariate regressions, other partial correlations are more relevant, as indicated in the data examples in Section 5.

Graphical structures: covariance graph, concentration graph, direct graphs within boxes.

The definition of *non-decomposable* is not clear in this paper.

In Section 6, there are:

(a) nondecomposable hypotheses in block regression

(b) nondecomposable hypotheses in concentrations (chordless 4-cycle?)

(c) nondecomposable hypotheses in multivariate regression chain models

(d) nondecomposable hypotheses in covariances

Section 2: a *nondecomposable independence hypotheses* consists of a set of k distinct variable pairs that cannot, in its entirety, be re-expressed in terms of vanishing coefficients in the form of *univariate recursive regressions*: that is, no ordering of the variables would produce a decomposable independence hypothesis with the same implications from the same distributional assumptions.

SEM (structural equation model), general form

Threshold model for the latent outcomes.

$\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ observed. $\tilde{\mathbf{Y}} = (\tilde{Y}_1, \ldots, \tilde{Y}_d)$ latent.

Factor model for latent is $\tilde{\mathbf{Y}} = \boldsymbol{\mu} + \Lambda\boldsymbol{\eta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\mu}, \boldsymbol{\epsilon}$ are vectors of length d, $\boldsymbol{\eta}$ is a vector of length k for the number of factors, and Λ is an $d \times k$ matrix. The covariance matrix of $\boldsymbol{\eta}$ is Σ and the covariance matrix of $\boldsymbol{\epsilon}$ is Ψ. Residual dependence implies that Ψ is not diagonal?

Structural model for latent is

$$\boldsymbol{\eta} = \Gamma\boldsymbol{\eta} + \boldsymbol{\zeta}.$$

Γ is $k \times k$, with diagonal elements being 0. The covariance matrix of $\boldsymbol{\zeta}$ is Ξ. The structure is usually summarized in a path diagram? For vines, Γ would be (lower) triangular.

In complete generality, this model might not be identifiable in Γ, Ξ

Examples in Cox&Wermuth (1993), Statistical Science

They are of the form for 4 variables Y, X, V, W.

Consider regression of (Y, X) on (V, W). If any of

$$\rho_{YV;W}, \rho_{YW;V}, \rho_{XV;W}, \rho_{XW;V}$$

is 0, then there is a regression coefficient that is zero and there is some parsimony relative to a saturated model.
For example, $\beta_{V:Y\sim V,W} \overset{\text{sgn}}{=} \rho_{YV;W}$.

Section 5, example 2: y, x, v, w: log(syst/diast)bp, log(diastolicbp), bodymass, age

$$R = \begin{pmatrix} 1 & - & - & - \\ -.544 & 1 & - & - \\ -.253 & .336 & 1 & - \\ -.131 & .510 & .608 & 1 \end{pmatrix}.$$

Then $r_{XV;W} = 0.038$, $r_{YW;V} = 0.030$. But $r_{XY;VW} = -0.566$. So this could be a $Y - V - W - X$ D-vine with 0s only in the second tree.

Section 5, example 6: y, x, v, w: long-interval-response, short-interval-response, strongest-short-interval-response, response-to-innocuous-stimulus

$$R = \begin{pmatrix} 1 & - & - & - \\ .72 & 1 & - & - \\ .30 & .54 & 1 & - \\ .19 & .43 & .71 & 1 \end{pmatrix}.$$

Then $r_{YW;X} = -0.19$, $r_{YV;X} = -0.15$, and $r_{YW;XV} = -0.12$ and Cox/Wermuth suggest the conditional indep of $Y \perp (V, W)|X$. This is a C-vine rooted at X with some conditional independence in tree 2.

This could be a 1-truncated $Y - X - V - W$ D-vine with (.72, .54, .71) on edges of tree 1, and $r_{YV;X} = -.15$, $r_{XW;V} = 0.08$, $r_{YW;XV} = -0.12$. Not clear that regression (Y, X) on (V, W) is reasonable in this case.

Section 5, example 1: y, x, v, w: state anxiety, state anger, trait anxiety, trait anger.

$$R = \begin{pmatrix} 1 & - & - & - \\ .61 & 1 & - & - \\ .62 & .47 & 1 & - \\ .39 & .50 & .49 & 1 \end{pmatrix}.$$

Then $r_{YW;XV} = -0.04$, $r_{XV;YW} = 0.03$.

This doesn't fit parametrization of vine or regression of (Y, X) on (V, W).

Wermuth 1980, JASA, 75, 963–972.

A covariance selection model with reducible zero pattern in the concentration matrix can equivalently be described by a linear recursive system (A, T) [same as Bayesian network] that has the same zero pattern for the regression coefficients, and vice versa.

Definition of reducible. Let $P = \{(i, j) : 1 \le i < j \le p\}$. $I \subset P$ is reducible if $(i, j) \in I$ and $h \in \{1, \ldots, i-1\}$ implies $(h, i) \in I$ or $(h, j) \in I$ or both. A matrix M has a reducible zero pattern with respect to I if the zeros of M are in I and the set I is reducible.

Speed and Kiiveri 1986, Annals of Statistics, 14 (1), 138–150

Assume multivariate normal. They have an algorithm for finding the MLE of $\boldsymbol{\Sigma}$ given known positions of 0s in the precision/concentration matrix $\boldsymbol{\Sigma}^{-1}$.

Also there is code in Wermuth and Scheidt (1977). Algorithm AS 105: Fitting a Covariance Selection Model to a Matrix Journal of the Royal Statistical Society. Series C (Applied Statistics), 26 (1), pp. 88–92.

Original idea of MLE with fixed zero positions in precision matrix is due to: Dempster (1972). Covariance selection. *Biometrics*, 28(1), 157–175.

Suppose we use graphical lasso to get 0 positions and then apply the MLE algorithm. Check if result is better for the discrepancy measure:

$$D_{\text{model}} = \log(\det[\boldsymbol{R}_{\text{model}}(\hat{\boldsymbol{\delta}})]) - \log(\det[\boldsymbol{R}_{\text{data}}]) + \text{tr}[\boldsymbol{R}^{-1}_{\text{model}}(\hat{\boldsymbol{\delta}})\boldsymbol{R}_{\text{data}}] - d,$$