Stat 521, Graphical Models and Parsimonious Dependence
Prof Harry Joe, Fall 2015, 2015.10.26–2015.12.05

Outline

- Definitions for Bayesian network, independence graph, vine, path diagram
- Interpretation of models with zeros in the precision matrix.

Definition of Bayesian network graph and model

There is an ordering of variables, labelled $Y_1, \dots, Y_d$ after permutation. There are parent sets for $j = 2, \dots, d$. $P(Y_j) \subset \{Y_1, \dots, Y_{j-1}\}$ is the set of parents of Y_j; could be empty set or all predecessors. Similarly, let P_j as the indices of the variables in $P(Y_j)$.

The graph (directed acyclic graph) has directed edges from $Y_k \to Y_j$ for every $Y_k \in P(Y_j)$.

This leads to a joint density $f_{Y_1,\dots,Y_d} = f_{Y_1} \prod_{j=2}^d f_{Y_j|P(Y_j)}$, or $f_{1:d} = f_1 \prod_{j=2}^d f_{j|P_j}$.

There is parsimony when (there is a permutation of variables such that) parent sets have small cardinality, e.g., an ordering of the variables so that parent sets are bounded in size by m such as $m = 3$ or 4.
If all parent sets have cardinality of 1, the Bayesian network is a Markov tree (with direction).

In the case of multivariate Gaussian $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, its conditional density $f_{Y_j|P(Y_j)}$ is univariate Gaussian of the form:

$$[Y_j|Y_k = y_k, k \in P_j] \sim N(\mu_j + \boldsymbol{\Sigma}_{jP_j}\boldsymbol{\Sigma}_{P_jP_j}^{-1}(\boldsymbol{y}_{P_j} - \boldsymbol{\mu}_{P_j}), \boldsymbol{\Sigma}_{jP_j}\boldsymbol{\Sigma}_{P_jP_j}^{-1}\boldsymbol{\Sigma}_{P_jj}).$$

Here, $\boldsymbol{\Sigma}_{jP_j}$ is the submatrix of $\boldsymbol{\Sigma}$ with row j and columns $k \in P_j$, $\boldsymbol{\Sigma}_{P_jP_j}$ is the submatrix $\boldsymbol{\Sigma}$ with rows and columns $k \in P_j$, etc.

Definition of (conditional) independence graph

Consider d variables $(Y_1, \dots, Y_d)$ that are jointly multivariate Gaussian. The quadratic form in the density is $(\boldsymbol{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$. Let $\boldsymbol{\Sigma}^{-1} = (\sigma^{jk})_{1\le j,k\le d}$.
Variables Y_j and Y_k are connected in the (conditional) independence graph if $\sigma^{jk} \neq 0$.
Equivalently Y_j and Y_k are <u>not</u> connected if $\rho_{jk;\{i:i\neq j, i\neq k\}} = 0$.

More general definition in Kurowicka and Cooke (2006): Y_j and Y_k are not connected if Y_j, Y_k are conditionally independent given $\{Y_i : i \neq j, i \neq k\}$; otherwise they are connected.

However in general multivariate non-Gaussian, there is no way to construct a multivariate distribution with these conditional independence relations.

Definition of vine and truncated vine. [A tree is a graph with no cycles.]

(Regular vine). [Cooke, Bedford, Kurowicka]. $\mathcal{V}$ is a *regular vine on d elements*, with $\mathcal{E}(\mathcal{V}) = \mathcal{E}_1 \cup \cdots \cup \mathcal{E}_{d-1}$ denoting the set of edges of $\mathcal{V}$, if

1. $\mathcal{V} = \{\mathcal{T}_1, \dots, \mathcal{T}_{d-1}\}$ [consists of $d-1$ trees];
2. $\mathcal{T}_1$ is a connected tree with nodes $\mathcal{N}_1 = \{1, \dots, d\}$, and edges $\mathcal{E}_1$; for $\ell = 2, \dots, d-1$, $\mathcal{T}_\ell$ is a tree with nodes $\mathcal{N}_\ell = \mathcal{E}_{\ell-1}$ [edges in a tree becomes nodes in the next tree];
3. *(proximity)* for $\ell = 2, \dots, d-1$, for $\{a, b\} \in \mathcal{E}_\ell$, $\#(a \triangle b) = 2$, where $\triangle$ denotes symmetric difference and $\#$ denotes cardinality [nodes joined in an edge differ by two elements].

More simply, a d-dimensional vine is a set of edges/nodes divided into $d-1$ trees; an edge in tree ℓ becomes a node in tree $\ell+1$.
Tree 1: $d-1$ edges, each with 2 variables.
tree 2: $d-2$ edges, each with 2 conditioned variables and 1 conditioning variable;
tree 3: $d-3$ edges, each with 2 conditioned variables and 2 conditioning variables;
$\ldots$,
tree $d-1$: one edge, 2 conditioned variables and $d-2$ conditioning variables.
Two nodes can be joined in an edge in the next tree only if the symmetric difference of the **nodes** is 2.

***m*-truncated vine**: conditional independence after tree m. A 1-truncation vine is a Markov tree model.

The following result is due to Bedford and Cooke 2001. Let $\mathcal{V} = \{\mathcal{T}_1, \ldots, \mathcal{T}_{d-1}\}$ be a regular vine, then

1. the number of edges is $d(d-1)/2$ [equals number of pairs of variables];
2. each conditioned set is a doubleton, each pair of variables occurs exactly once as a conditioned set [matching a bivariate distribution with univariate conditional margins] **[or each pair of variables is joined in an edge in exactly one of the trees]**;
3. if two edges have the same conditioning set, then they are the same edge.

Sequentially constructing a vine with d variables

$\mathcal{T}_1$: $\binom{d}{2}$ possible edges in $\mathcal{E}_1^* = \{\{j,k\} : 1 \le j < k \le d\}$; choose $d-1$ edges and get set $\mathcal{E}_1$.

$\mathcal{T}_2$: $\mathcal{N}_2 = \mathcal{E}_1$; with the proximity conditions, at least $d-2$ edges can be formed and at most $\binom{d-1}{2}$. Let $\mathcal{E}_2^* = \{\{a,b\} \in \mathcal{N}_2 : a \cap b = 1\}$ be the set of possible edges and choose $d-2$ edges in subset $\mathcal{E}_2$ to form a tree. If $e \in \mathcal{E}_2$ is formed from $a, b \in \mathcal{N}_2$, then $a \cup b$ has cardinality 3 and the edge can be denoted as $a \cup b$ or $[a\Delta b | a \cap b]$ where $a \cap b$ is the set of conditioning variables and $a \Delta b$ is the conditioned set of 2 variables.

$\mathcal{T}_3$: $\mathcal{N}_3 = \mathcal{E}_2$; with the proximity conditions, at least $d-3$ edges can be formed and at most $\binom{d-2}{2}$. Let $\mathcal{E}_3^* = \{\{a,b\} \in \mathcal{N}_3 : a \cap b = 2\}$ be the set of possible edges and choose $d-3$ edges in subset $\mathcal{E}_3$ to form a tree. If $e \in \mathcal{E}_3$ is formed from $a, b \in \mathcal{N}_3$, then $a \cup b$ has cardinality 4 and the edge can be denoted as $a \cup b$ or $[a\Delta b | a \cap b]$ where $a \cap b$ is the set of conditioning variables and $a \Delta b$ is the conditioned set of 2 variables.

Etc. (example to be done on the whiteboard).

A vine graph can be coded compactly into a vine array, from which (in the case of multivariate Gaussian) a sequence of recursive regressions can be written. Because the vine array for a vine is not unique, neither is the recursive regression representation.

The vine array for d variables is $d \times d$ and only the diagonal and the superdiagonal are used. The diagonal has a permutation $i_1, \ldots, i_d$ of $1, \ldots, d$. In the jth column, $a_{1j}, \ldots, a_{j-1,j}$ is a permutation of $(i_1, \ldots, i_{j-1})$.

$$A = \begin{bmatrix} i_1 & a_{12} & \cdots & a_{1j} & \cdots & a_{1d} \\ & i_2 & \cdots & a_{2j} & \cdots & a_{2d} \\ & & \ddots & \vdots & \vdots & \vdots \\ & & & i_j & \cdots & a_{jd} \\ & & & & \ddots & \vdots \\ & & & & & i_d \end{bmatrix}$$

In column j ($j = 2, \ldots, d$), a_{1j} can be any of $i_1, \ldots, i_{j-1}$; there is a further condition that, for $\ell = 2, \ldots, j-1$, $\{a_{1j}, \ldots, a_{\ell,j}\} = \{a_{k1}, \ldots, a_{\ell-1,k}, a_{kk}\}$ for some $k = k_\ell < j$.

The edges in $\mathcal{T}_1$ are $[a_{12}i_2], [a_{13}i_3], \ldots, [a_{1d}i_d]$.
The edges in $\mathcal{T}_2$ are $[a_{23}i_3|a_{13}], \ldots, [a_{2d}i_d|a_{1d}]$.
The edges in $\mathcal{T}_3$ are $[a_{34}i_4|a_{14}, a_{24}], \ldots, [a_{3d}i_d|a_{1d}, a_{2d}]$.
$\cdots$. The one edge in $\mathcal{T}_{d-1}$ is $[a_{d-1,d}i_d|a_{1d}, \ldots, a_{d-2,d}]$.

Example from page 354 of *Dependence modeling with copulas*

$$A = \begin{bmatrix} 4 & 4 & 4 & 2 & 2 & 5 & 2 \\ & 2 & 2 & 4 & 4 & 2 & 4 \\ & & 6 & 6 & 6 & 4 & 1 \\ & & & 5 & & & \\ & & & & 1 & & \\ & & & & & 3 & \\ & & & & & & 7 \end{bmatrix}, \quad \begin{bmatrix} - & 42 & 46 & 25 & 21 & 53 & 27 \\ & - & 26|4 & 45|2 & 41|2 & 23|5 & 47|2 \\ & & - & 65|24 & 61|24 & 43|52 & 17|24 \\ & & & - & & & \\ & & & & - & & \\ & & & & & - & \\ & & & & & & - \end{bmatrix}.$$

Check that the requirements for a vine array are satisfied.
For example, 43|52 is an edge with nodes 23|5 (one row immediately above in vine array) and 45|2 (one row above and to the left).
What are possibilities to complete rows 4 to 6?

Definition of path diagram; origin Wright (1921), *J Agricultural Research*, 20, 557-585.
Page 120 of Mulaik (2009) has explanations of a path diagram (for a recursive model), but not a definition.

- Manifest (observed) variables are shown by rectangles, latent (unobserved) variables are shown by ellipses.
- Direct causal paths are represented by one-headed arrows pointing from the causal variable to effect variable (so each latent variable often points to one or more observed variable).
- Covariances between pairs of variables are shown by a two-headed curve (edge with arrows on both sides); these variables are conditionally dependent given all the variables that point to them.
- Exogenous variables (covariates) have one-headed arrows pointed away from them
- Endogenous variables (responses) have arrows pointing to them, but can also point to other variables.
- Manifest exogenous variables are denoted by x; manifest endogenous variables are denoted by y.
- Latent variables are often denoted by ξ and η.
- Optionally, a disturbance variable, denoted by e or ϵ, points to the observed variable it affects (typically, there is a disturbance variable for each y).
- There is a structural coefficient for each direct causal path.

Recursive path diagram matches(?) a Bayesian network even if edges with two arrows are replaced by a directed edge (arbitrary direction), provide this is done in a way so that there are no directed cycles in the graph.

Recursive because there are recursions to determine the covariance matrix; compare recursions for autocorrelations in Gaussian time series.

There are also non-recursive models, where there is a directed cycle (simplest case of two nodes Y_a, Y_b such that Y_a points to Y_b and Y_b points to Y_a in a feedback loop.

A path diagram for non-recursive structural equation model (SEM) cannot be converted to a Bayesian network (or truncated vine). The derivation of the covariance matrix is more complicated.

Interpretation of models with zeros in precision matrix

Most useful for those which are equivalent to a recursive sequence of regressions (i.e., Bayesian network representation).

See the document that is part of Homework 1, for when 0s in the precision matrix can imply the lower order partial correlations are also zero.

Wermuth's papers has conditions on zero positions of precision matrix that imply a recursive model.

More simply, consider recursive models that follow from Bayesian networks or truncated vines (subset of Bayesian networks)

Most useful results on partial correlations:

1. Regression Y on $\boldsymbol{W} = (W_1, \dots, W_m)$ where $(Y, W_1, \dots, W_m)$ are jointly multivariate normal. Then regression equation is

$$Y = \beta_0 + \beta_1 W_1 + \cdots + \beta_m W_m + \epsilon$$

and the conditional expectation is

$$\mathrm{E}\,(Y|W_1 = w_1, \dots, W_m = w_m) = \beta_0 + \beta_1 w_1 + \cdots + \beta_m w_m.$$

Then for $j = 1, \dots, m$,

$$\beta_j \overset{\text{sign}}{=} \rho_{YW_j;\boldsymbol{W}_{-j}} = \rho_{YW_j;\text{rest}}, \quad \boldsymbol{W}_{-j} = \{\boldsymbol{W} \backslash W_j\}$$

1b. Related result for multiple regression: variables $y, x_1, \dots, x_m$. Data $(y_i, x_{i1}, \dots, x_{im})$, $i = 1, \dots, m$. Regress y on $\boldsymbol{x} = (x_1, \dots, x_m)$. Get least squares regression coefficients $\hat\beta_0, \hat\beta_1, \dots, \hat\beta_m$. Then for $j = 1, \dots, m$,

$$\hat\beta_j \overset{\text{sign}}{=} r_{yx_j;\boldsymbol{x}_{-j}} = r_{yx_j;\text{rest}}, \quad \boldsymbol{x}_{-j} = \{\boldsymbol{x} \backslash x_j\}$$

2. $\boldsymbol{\Sigma}$ is covariance matrix of multivariate normal random vector $Y_1, \dots, Y_d$. $\boldsymbol{\Sigma}^{-1} = (\sigma^{ij})_{1 \le i,j \le d}$. Then

$$\sigma^{ij} \overset{\text{sign}}{=} -\rho_{ij;T(i,j)} = -\rho_{ij;\text{rest}}, \quad T(i,j) = \{1, \dots, d\} \backslash \{i,j\}$$

Put together items 1 and 2:

If $\sigma^{12} = 0$, consider regression of Y_1 on $Y_2, Y_3, \dots, Y_d$ to get

$$Y_1 = \beta_{10} + \beta_{12} Y_2 + \beta_{13} Y_3 + \cdots + \beta_{1d} Y_d + \epsilon_1,$$

and regression of Y_2 on $Y_1, Y_3, \dots, Y_d$ to get

$$Y_2 = \beta_{20} + \beta_{21} Y_1 + \beta_{23} Y_3 + \cdots + \beta_{2d} Y_d + \epsilon_2.$$

Then $\beta_{12} = \beta_{21} = 0$.