

Conditional independence copula models with graphical representations

Harry Joe (University of British Columbia)

For multivariate Gaussian with a large number of variables, there are several approaches to obtain **parsimonious dependence models**. Sometimes the variables can be divided into non-overlapping groups, and it can be desirable to first construct models for each group of variables followed by a second stage of combining the groups. Various constructions with parsimonious dependence are done via conditional independence relationships. Examples are: **common factor models**, **structured factor models**, **structural equation models**, **factor models with residual dependence**, **Bayesian networks**, for which graphical representations can be made with **path diagrams**, **vines** and **directed acyclic graphs**. Through some concrete examples, it is indicated how these **Gaussian models can be extended to copula models** with vine or similar representations. The extension from Gaussian models to copula models is typically needed when there is tail dependence or tail asymmetry relative to Gaussian joint tails.

(i) **Partial correlation parametrizations for Gaussian dependence models**

(ii) **The mixing of conditional distributions as the copula extension of partial correlations.**

(iii) **The substitution of a bivariate copula for each partial correlation and combined with the sequential mixing of conditional distributions to get the vine or Bayesian network copula extension of multivariate Gaussian, even if there are latent variables.**

Program and big picture

Big data: can copula modelling be used for high dimensions, say $100\text{--}1000^+$.

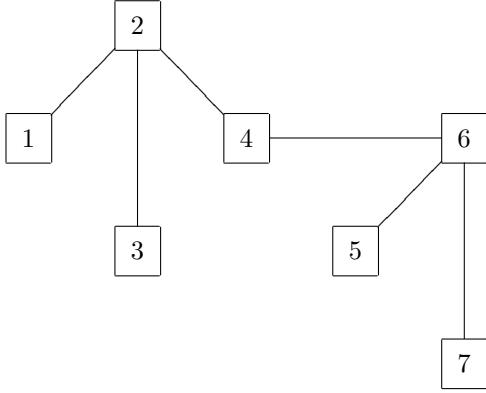
- Make use of linear Gaussian methods to get parsimonious dependence structures.
- These Gaussian models are first-order (linear) approximations.
- Modify the models with copulas to handle non-linear dependence (tail dependence, tail asymmetry).
- Hope that this leads to good useable “suboptimal” models.

Borrow ideas from machine learning, psychometrics, numerical methods,

Our software: some copula models have been tested in dimensions exceeding 100. Some of the (first-order) copula models can work in dimensions 1000^+ .

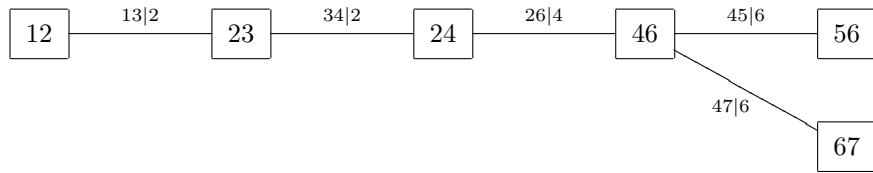
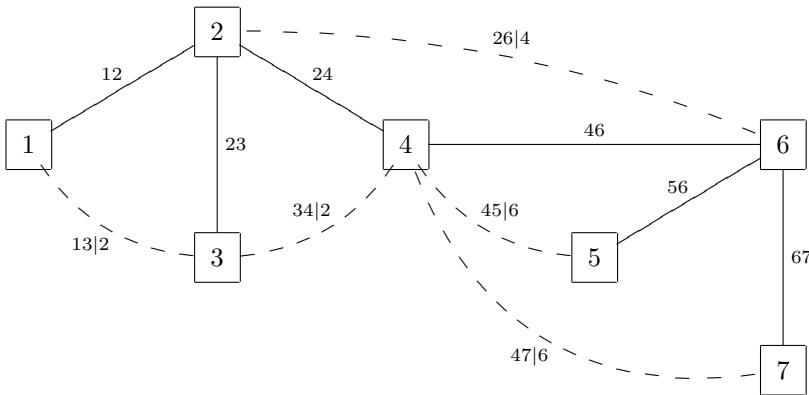
Outline:

- Pictures of graphical models via path diagrams, truncated vines and directed acyclic graphs.
 1. Most parsimonious graphical models: Markov tree and 1-factor; $d - 1$ or d parameters/edges for d variables [the context helps for which is better]
 2. Best Markov tree from minimum spanning tree with appropriate weights; best 1-factor from factor analysis.
 3. If latent variable model is desirable but 1-factor model does not fit well enough, next most parsimonious models to consider are 2-factor, bi-factor and 1-factor with residual dependence (structural equation models).
 4. Otherwise if Markov tree is better than 1-factor, and Markov tree does not fit well enough, next most parsimonious models to consider are truncated vines and Bayesian networks.
- Gaussian models, parametrization via partial correlations. Comparison of parsimonious dependence structures based on fit statistics applied to the correlation matrix of normal scores.
- Copula extension: replace partial correlation on each edge with a bivariate copula (chosen based on tail analysis)
- Data example, and Diagnostics for tail dependence/asymmetry (tail property separated from dependence structure).

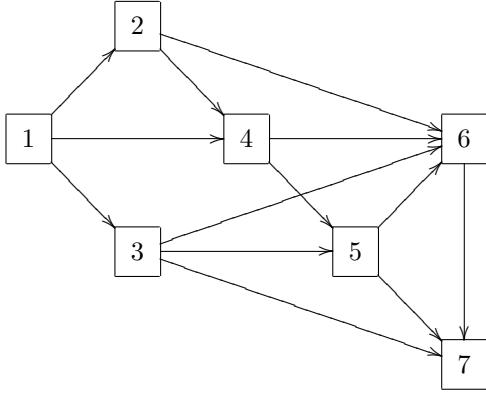


Markov tree with edges 12, 23, 24, 46, 56, 67. Connect variables with strongest dependence.

Conditional independence: $Y_1 \perp\!\!\!\perp Y_4 \perp\!\!\!\perp Y_3 | Y_2$, $Y_7 \perp\!\!\!\perp Y_5 \perp\!\!\!\perp Y_4 | Y_6$, $Y_2 \perp\!\!\!\perp Y_6 | Y_4$, $(Y_1, Y_3) \perp\!\!\!\perp (Y_5, Y_7) | (Y_2, Y_4, Y_6)$ etc.



2-truncated vine (Markov tree cannot explain all dependence): tree 2 with nodes 12, 23, 24, 46, 56, 67; edges 13|2, 34|2, 26|4, 45|6, 47|6; Conditional independence: $Y_4 \perp\!\!\!\perp Y_1 | (Y_2, Y_3)$, $Y_6 \perp\!\!\!\perp (Y_1, Y_3) | (Y_2, Y_4)$, $Y_5 \perp\!\!\!\perp (Y_1, Y_2, Y_3) | (Y_6, Y_4)$, $Y_7 \perp\!\!\!\perp (Y_1, Y_2, Y_3, Y_5) | (Y_6, Y_4)$.

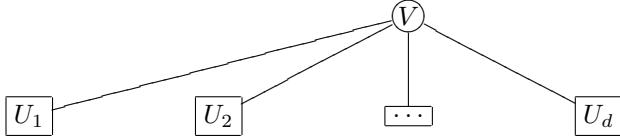


DAG=directed acyclic graph (Bauer 2013).

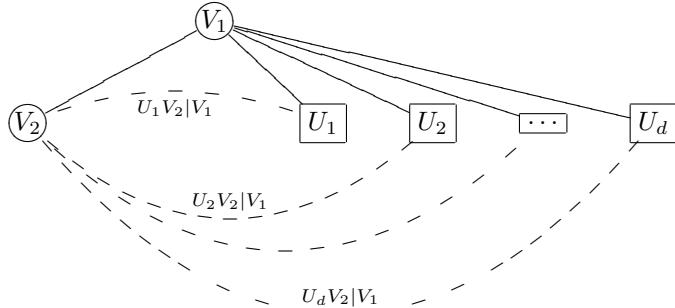
Joint density $f_1 f_{2|1} f_{3|1} f_{4|12} f_{5|34} f_{6|2345} f_{7|356}$. Edges labelled as:

12; 13; 24, 14|2; 35, 45|3; 56, 36|5, 46|35, 26|345; 67, 57|6, 37|56

$\rho_{jk;S}$ for Gaussian, $C_{jk;S}$ for pair-copulas where $F_{jk|S} = C_{jk;S}(F_{j|S}, F_{k|S})$. Conditional independence: $Y_3 \perp\!\!\!\perp Y_2|Y_1$, $Y_4 \perp\!\!\!\perp Y_3|(Y_1, Y_2)$, $Y_5 \perp\!\!\!\perp (Y_1, Y_2)|(Y_3, Y_4)$, $Y_6 \perp\!\!\!\perp Y_1|(Y_2, Y_3, Y_4, Y_5)$, $Y_7 \perp\!\!\!\perp (Y_1, Y_2, Y_4)|(Y_3, Y_5, Y_6)$. Expert opinion or learning algorithm (numbers assigned to variable after graph).



Path diagram for 1-factor. V is latent; 1-truncated vine rooted at latent (1-factor model) with edges $[VU_1], \dots, [VU_d]$; positive exchangeable/extendible is special case. Conditional independence given latent: $U_1 \perp\!\!\!\perp U_2 \perp\!\!\!\perp \dots \perp\!\!\!\perp U_d$.



Path diagram for 2-factor. V_1, V_2 latent; 2-truncated vine (2-factor) with edges $[V_1V_2], [V_1U_1], \dots, [V_1U_d]$. for tree \mathcal{T}_1 and $[V_2U_1|V_1], \dots, [V_2U_d|V_1]$ for tree \mathcal{T}_2 . There is a 1-1 map of loading parameters and correlations/partial correlations with latent variables. Conditional independence given latents: $U_1 \perp\!\!\!\perp U_2 \perp\!\!\!\perp \dots \perp\!\!\!\perp U_d|(V_1, V_2)$.

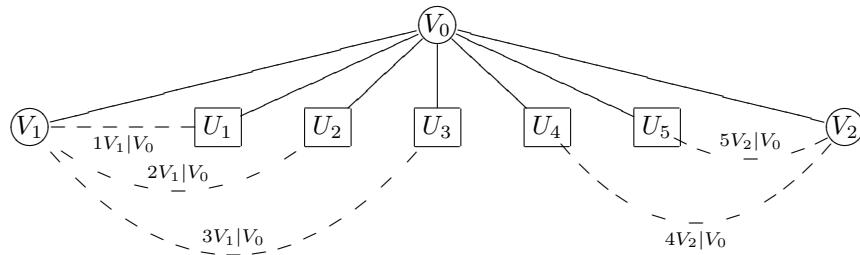
Bi-factor structural correlation matrix is a special case arising from the loading matrix \mathbf{A} (matrix of regression coefficients on latent variables $V_0, V_1, V_2, \dots, V_G$) with $p = G + 1$ based on G non-overlapping groups of variables.

\mathbf{A} can be written as

$$\begin{pmatrix} \phi_1 & \alpha_1 & 0 & \cdots & 0 \\ \phi_2 & 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_G & 0 & 0 & \cdots & \alpha_G \end{pmatrix},$$

where the g th group J_g (block of rows) has size d_g for $g = 1, \dots, G$.

For finance, the bi-factor model is plausible for many stock returns with the first column representing a global/common latent factor, and other columns representing latent factors for different sectors of the stock market. The structured zeros are based on additional information about the variables. The bi-factor model comes from psychometrics.



Bi-factor with 2 groups, 3 in group 1, 2 in group 2. 2-truncated vine, \mathcal{T}_1 with $[V_0V_1], [V_0V_2], [V_0U_1], [V_0U_2], [V_0U_3], [V_0U_4], [V_0U_5]$; \mathcal{T}_2 with $[U_1V_1|V_0], [U_2V_1|V_0], [U_3V_1|V_0], [U_4V_2|V_0], [U_5V_2|V_0]$. There is a 1-1 map of loading parameters and correlations/partial correlations with latent variables. V_0 is common latent; V_1, V_2 are group latent variables.

G groups: V_0 is common latent, V_g is latent for group g ($g = 1, \dots, G$). Conditional independence given latents: $U_j \perp\!\!\!\perp \dots \perp\!\!\!\perp U_k | (V_0, V_g)$, if variables j, \dots, k in group g . $U_p \perp\!\!\!\perp U_q | (V_0, V_g, V_h)$, if variable p in group g and variable q in group h .

Bi-factor copula density (Krupskii, 2014)

V_0, V_1, \dots, V_g are independent $U(0, 1)$.

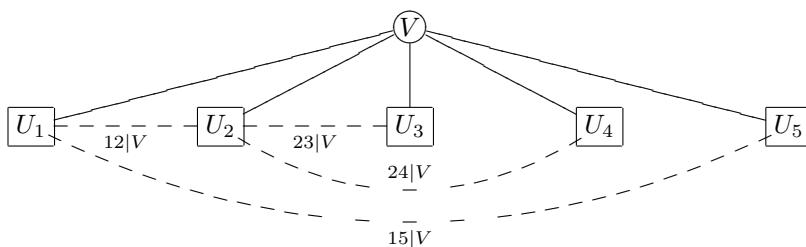
d_g variables in group g , $g = 1, \dots, G$. Use Bedford-Cooke vine density representation and integrate over latent variables.

$$c_{\mathbf{U}}(\mathbf{u}) = \int_0^1 \prod_{g=1}^G \left\{ \left[\prod_{i=1}^{d_g} c_{U_{ig}, V_0}(u_{ig}, v_0) \right] \int_0^1 \left[\prod_{i=1}^{d_g} c_{U_{ig}, V_g; V_0}(C_{U_{ig}|V_0}(u_{ig}|v_0), v_g) \right] dv_g \right\} dv_0.$$

This is a one-dimensional integral of a function which in turn is a product of G one-dimensional integrals. Numerical optimization (for MLE) with 2-dimensional Gaussian quadrature.

1-factor copula density

$$c_{\mathbf{U}}(\mathbf{u}) = \int_0^1 \prod_{j=1}^d c_{U_j, V}(u_j, v) dv$$



1-factor with residual dependence, example of SEM that is not a factor model (Brechmann and Joe, 2014).
1 latent variable; observed variables have Markov tree dependence given latent.
Conditional independence: $U_3 \perp\!\!\!\perp U_1 | (V, U_2)$, $U_4 \perp\!\!\!\perp (U_1, U_3) | (V, U_2)$, $U_5 \perp\!\!\!\perp (U_2, U_3, U_4) | (V, U_1)$.

Notation

$F = F_{1:d}$ is a d -variate distribution.

Marginal distributions are $F_S : S \subset \{1, \dots, d\}$.

Conditional distributions are $F_{S|T}$ if S, T are disjoint subsets of $\{1, \dots, d\}$.

$C_{jk;S}$ is the copula for $F_{jk|S}$ with conditional distributions $C_{j|k;S}, C_{k|j;S}$.

Φ is standard normal cumulative distribution function (cdf).

$\Phi_2(\cdot; \rho)$ is bivariate normal/Gaussian cdf.

$\Phi_d(\cdot; \Sigma)$ is d -variate normal/Gaussian cdf.

$\rho_{jk;S}$ is a partial correlation.

$N_d(\mu, \Sigma)$ for multivariate normal/Gaussian distribution.

Bayesian network example: compatibility of margins.

$$f_{1:7} = f_1 f_{2|1} f_{3|2} f_{4|23} f_{5|24} f_{6|15} f_{7|34} = \frac{f_{12} f_{234} f_{245}}{f_2 f_{24}} \frac{f_{156}}{f_{15}} \frac{f_{347}}{f_{34}}.$$

Can one avoid 7-dimensional model, by coming up with sub-models for $f_{12}, f_{24}, f_{34}, f_{15}, f_{234}, f_{245}, f_{156}, f_{347}$.

Cannot do independently because f_{12}, f_{15}, f_{245} have a constraint that the f_{25} margin is not algebraically independent of f_{12}, f_{15} . (Positive definite matrix constraint for $\rho_{12}, \rho_{15}, \rho_{25}$ for Gaussian.)

To get around these compatibility constraints, models use partial correlations for Gaussian and, more generally, bivariate copulas applied to conditional distributions — for example, $C_{jk;S}(F_{j|S}, F_{k|S})$ where S is a non-empty set of indices not including j or k .

Mixture of conditional distributions: Fréchet class $\mathcal{F}(F_{\{j\} \cup S}, F_{\{k\} \cup S})$ where $j \neq k$ and $j, k \neq S$ and S is a non-empty set of indices.

Different copula models are built from a sequence of mixtures.

Let $F_{\{j,k\} \cup S}$ be in the Fréchet class. Then line 2 has the application of Sklar's theorem, line 3 make use of the simplifying assumption to get a model in this class.

$$\begin{aligned} F_{\{j,k\} \cup S}(\mathbf{x}) &= \int_{\infty}^{\mathbf{x}_S} F_{jk|S}(x_j, x_k | \mathbf{y}_S) dF_S(\mathbf{y}_S) \\ &= \int_{\infty}^{\mathbf{x}_S} C_{jk;S}(F_{j|S}(x_j | \mathbf{y}_S), F_{j|S}(x_k | \mathbf{y}_S); \mathbf{y}_S) dF_S(\mathbf{y}_S) \\ &=_{\text{model}} \int_{\infty}^{\mathbf{x}_S} C_{jk;S}(F_{j|S}(x_j | \mathbf{y}_S), F_{j|S}(x_k | \mathbf{y}_S)) dF_S(\mathbf{y}_S) \end{aligned}$$

Assuming absolute continuity, one conditional distribution and the joint density are:

$$\begin{aligned} F_{j|\{k\} \cup S}(x_j | x_k, \mathbf{x}_S) &= \frac{\frac{\partial}{\partial x_k} C_{jk;S}(F_{j|S}(x_j | \mathbf{x}_S), F_{k|S}(x_k | \mathbf{x}_S)) f_S(\mathbf{x}_S)}{f_{\{k\} \cup S}(x_k, \mathbf{x}_S)} \\ &= \frac{C_{j|k;S}(F_{j|S}(x_j | \mathbf{x}_S) | F_{k|S}(x_k | \mathbf{x}_S)) f_{k|S}(x_k | \mathbf{x}_S) f_S(\mathbf{x}_S)}{f_{\{k\} \cup S}(x_k, \mathbf{x}_S)} \\ &= C_{j|k;S}(F_{j|S}(x_j | \mathbf{x}_S) | F_{k|S}(x_k | \mathbf{x}_S)) \\ f_{\{j,k\} \cup S}(x_j, x_k, \mathbf{x}_S) &= c_{jk;S}(F_{j|S}(x_j | \mathbf{x}_S), F_{k|S}(x_k | \mathbf{x}_S)) f_{j|S}(x_j | \mathbf{x}_S) f_{k|S}(x_k | \mathbf{x}_S) f_S(\mathbf{x}_S) \end{aligned}$$

Partial correlations for multivariate Gaussian

Let $(\mathbf{Y}_1^T, \mathbf{Y}_2^T)^T \sim N_d(\mathbf{0}, \Sigma)$ with subvectors of size $d-2$ and 2 , and let $\Sigma_{11}, \Sigma_{12}, \Sigma_{21}, \Sigma_{22}$ be blocks (submatrices) of Σ .

Let $\Sigma_{22;1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ be the covariance matrix of conditional distribution of $[\mathbf{Y}_2 | \mathbf{Y}_1 = \mathbf{y}_1]$.

Suppose $d \geq 3$ and Σ_{11} has dimension $(d-2) \times (d-2)$. From the 2×2 covariance matrix $\Sigma_{22;1}$, obtain the partial correlation $\rho_{d-1,d;1\dots d-2}$. With $\mathbf{a}_{d-1} = (\sigma_{1,d-1}, \dots, \sigma_{d-2,d-1})^\top$ and $\mathbf{a}_d = (\sigma_{1,d}, \dots, \sigma_{d-2,d})^\top$, then

$$\Sigma_{22;1} = \begin{pmatrix} \sigma_{d-1,d-1} - \mathbf{a}_{d-1}^\top \Sigma_{11}^{-1} \mathbf{a}_{d-1} & \sigma_{d-1,d} - \mathbf{a}_{d-1}^\top \Sigma_{11}^{-1} \mathbf{a}_d \\ \sigma_{d,d-1} - \mathbf{a}_d^\top \Sigma_{11}^{-1} \mathbf{a}_{d-1} & \sigma_{dd} - \mathbf{a}_d^\top \Sigma_{11}^{-1} \mathbf{a}_d \end{pmatrix}.$$

The conditional correlation is

$$\rho_{d-1,d;1\dots d-2} = \frac{\sigma_{d-1,d} - \mathbf{a}_{d-1}^\top \Sigma_{11}^{-1} \mathbf{a}_d}{[\sigma_{d-1,d-1} - \mathbf{a}_{d-1}^\top \Sigma_{11}^{-1} \mathbf{a}_{d-1}]^{1/2} [\sigma_{dd} - \mathbf{a}_d^\top \Sigma_{11}^{-1} \mathbf{a}_d]^{1/2}}.$$

If $\Sigma = \mathbf{R}$ is a correlation matrix, then in the above, $\sigma_{d-1,d} = \rho_{d-1,d}$ and $\sigma_{d-1,d-1} = \sigma_{dd} = 1$. Other partial correlations can be obtained by permutating indices and working with submatrices of Σ .

Partial correlation vine and extension to copula: d -variate, $(\rho_{jk})_{1 \leq j < k \leq d}$ be reparametrized in terms of $d-1$ correlations and $\binom{d-1}{2}$ partial correlations that are algebraically independent.

An example for $d = 5$. Consider a partial correlation vine with trees and edges as follows;

\mathcal{T}_1 : 12, 23, 24, 35; \mathcal{T}_2 : 13|2, 34|2, 25|3; \mathcal{T}_3 : 14|23, 45|23; \mathcal{T}_4 : 15|234.

Partial correlations are : $\rho_{12}, \rho_{13;2}, \rho_{14;23}, \rho_{15;234}$ etc. The vine array is

$$A = \begin{bmatrix} 1 & 1 & 2 & 2 & 3 \\ & 2 & 1 & 3 & 2 \\ & & 3 & 1 & 4 \\ & & & 4 & 1 \\ & & & & 5 \end{bmatrix} \text{ or } \begin{bmatrix} - & 12 & 23 & 24 & 35 \\ - & 13|2 & 34|2 & 25|3 \\ - & 14|23 & 45|32 \\ - & 15|324 \\ - \end{bmatrix}.$$

For vine copula with $C_{12}, C_{23}, C_{24}, C_{35}, C_{13;2}, C_{14;23}, C_{15;234}$ etc., the sequential mixing of conditional distributions is:

integrate $C_{13;2}(C_{1|2}, C_{3|2})$ to get C_{123} ,
integrate $C_{34;2}(C_{3|2}, C_{4|2})$ to get C_{234} ,
integrate $C_{25;3}(C_{2|3}, C_{5|3})$ to get C_{235} ,
integrate $C_{14;23}(C_{1|23}, C_{4|23})$ to get C_{1234} ,
integrate $C_{45;32}(C_{4|32}, C_{5|32})$ to get C_{2345} ,
integrate $C_{15;324}(C_{1|324}, C_{5|324})$ to get C_{12345} .

Example of non-vine: $A = \begin{bmatrix} 1 & 1 & 1 & 2 \\ & 2 & 2^\perp & 3 \\ & & 3 & 1^\perp \\ & & & 4 \end{bmatrix}$ is not a vine array, but when truncated, it could be an array

for a Bayesian network.

Sequential mixing of conditional distributions:

1. by trees to get (truncated) vine copula (rows of vine array);
2. by variable to get Bayesian network (columns of array).

The construction implies that the bivariate copulas assigned to the edges of the vine or Bayesian network are algebraically independent.

When all of the bivariate copulas on the edges are Gaussian, it implies that the corresponding partial correlations are algebraically independent.

Regular vine. [Cooke, Bedford, Kurowicka]. \mathcal{V} is a *regular vine on d elements*, with $\mathcal{E}(\mathcal{V}) = \mathcal{E}_1 \cup \dots \cup \mathcal{E}_{d-1}$ 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}_ℓ 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 $\{a, b\} \in \mathcal{E}_\ell$, $\#(a \Delta b) = 2$, where Δ 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;

...

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 variables at the two nodes is 2.

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

Comparison of graphical models:

- (a) Inverse correlation or precision matrix (Whittaker); parametrization does not extend from Gaussian to copula
- (b) Vine; extends from Gaussian to copula
- (c) Path diagram (structural equation model); extends to copula if SEM can be reparametrized to partial correlation vine; the class of all SEMs is very general.
- (d) Bayesian network (extends to pair-copula Bayesian network, integration needed when not special case of vine).

Property of multivariate Gaussian. Let $\mathbf{Z} \sim N_d(\mathbf{0}, \mathbf{R})$ where \mathbf{R} is nonsingular and let $\mathbf{R}^{-1} = (\rho^{jk})$. Then $\rho^{jk} \stackrel{\text{sign}}{=} -\rho_{jk;T(j,k)}$ where $T(j, k) = \{1, \dots, d\} \setminus \{j, k\}$. In particular, $\rho^{jk} = 0$ if $\rho_{jk;T(j,k)} = 0$ (variables j, k conditionally independent given the remaining variables). **Truncated vine implies many 0s in \mathbf{R}^{-1} .**

High-dimensional applications: want parsimonious (copula) model. For multivariate Gaussian; this means a parametrization of $\binom{d}{2}$ correlations with $O(d)$ parameters.

In classical multivariate statistics and psychometrics, examples are factor models.

A correlation matrix \mathbf{R} has the factor correlation structure if $\mathbf{R} = \mathbf{A}\mathbf{A}^\top + \mathbf{\Psi}^2$ where \mathbf{A} is $d \times p$ with $1 \leq p < d$ and $\mathbf{\Psi}^2$ is a diagonal matrix with elements $\psi_1^2, \dots, \psi_d^2$.

Truncated vines are parsimonious models as are Bayesian networks when the parent sets are small (say, all $\leq m$ in size). After ordering of indices of variables, $f_{\mathbf{X}_1, \dots, \mathbf{X}_d} = f_{\mathbf{X}_1} \prod_{j=2}^d f_{\mathbf{X}_j | PA(\mathbf{X}_j)}$, where $PA(X_j) \subset \{X_1, \dots, X_{j-1}\}$ is the set of parents of X_j .

Comparisons, properties and software

- Factor copula models are closed under margins, whereas vine copulas without latent variables are not closed under margins.
- Need factor copula models to cover positively dependent exchangeable distributions.
- Simple copula families from non-pair-copula constructions do not have flexible dependence.
- Multivariate distributions furthest from the simplifying assumption are additive factor models based on positive random variables (e.g. gamma convolution factor model).
- Truncated vines can approximate multivariate models including factor models and Bayesian networks; simplifying assumption can be extended by allowing parameters to be function of conditioning variables.

- Algorithms and software for truncated vines are available; continuing research on non-greedy algorithms for truncation.
 - Factor copula models: software for 1-factor, 2-factor, bi-factor, nested-factor (where 1- and 2-dimensional integration using Gauss-Legendre quadrature; log-likelihood, gradient and Hessian computed for modified Newton-Raphson for numerical maximum likelihood).
 - p -factor ($p \geq 3$) and tri-factor: software for multivariate t_ν with these structures (quasi-Newton optimization with analytic gradient).
-

Different parsimonious dependence models can be compared based on [fit statistics](#). Heuristic approach to first decide on dependence structure because space of copula models with pair-copulas applied to edges is too large.

A common discrepancy measure in the psychometrics and structural equation modeling literatures is:

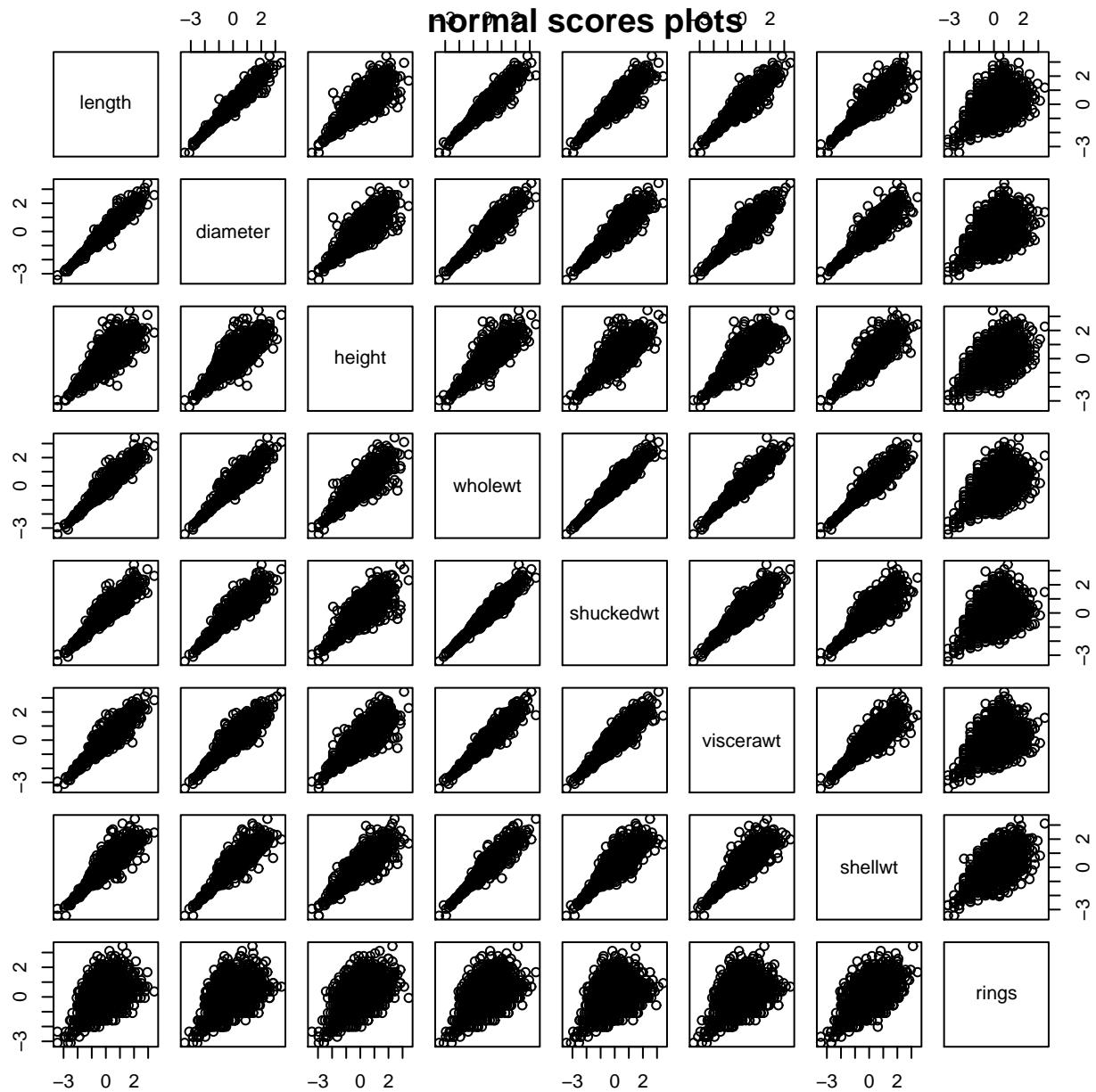
$$D_{\text{model}} = \log(\det[\mathbf{R}_{\text{model}}(\hat{\boldsymbol{\delta}})]) - \log(\det[\mathbf{R}_{\text{data}}]) + \text{tr}[\mathbf{R}_{\text{model}}^{-1}(\hat{\boldsymbol{\delta}}) \mathbf{R}_{\text{data}}] - d,$$

In the above, \mathbf{R}_{data} is the $d \times d$ matrix of normal scores (or polychoric/polyserial matrix), and $\mathbf{R}_{\text{model}}(\hat{\boldsymbol{\delta}})$ is the model-based correlation matrix based on the estimate of the parameter $\boldsymbol{\delta}$ in a truncated vine or Bayesian network. If either model has some conditional independence relations, then the dimension of $\boldsymbol{\delta}$ is less than $d(d-1)/2$. Other comparisons are the [AIC/BIC based on a Gaussian log-likelihood](#). Also useful are the average and max absolute deviations of the model-based correlation matrix from the empirical correlation matrix:

$\text{average}_{j < k} |\mathbf{R}_{\text{model},jk}(\hat{\boldsymbol{\delta}}) - \mathbf{R}_{\text{data},jk}|$ and $\max_{j < k} |\mathbf{R}_{\text{model},jk}(\hat{\boldsymbol{\delta}}) - \mathbf{R}_{\text{data},jk}|$.

[Data example: abalone males](#), $d = 8$ variables, $n = 1528$; from archive.ics.uci.edu/ml/datasets.html [sorte de gros escargot, Abalone Seeohr, oreja marina, orecchia di mare, zeeoor, baoyu]

1. Length continuous mm Longest shell measurement
2. Diameter continuous mm perpendicular to length
3. Height continuous mm with meat in shell
4. Whole weight continuous grams whole abalone
5. Shucked weight continuous grams weight of meat
6. Viscera weight continuous grams gut weight (after bleeding)
7. Shell weight continuous grams after being dried
8. Rings integer +1.5 gives the age in years



Denote as V_1, \dots, V_8 . V_8 has values from 1 to 28; treat as continuous for analysis of correlation matrix of normal scores.

`bnlearn` leads to the following matrix of sequential regressions:

$$\begin{array}{c|ccccc|c}
 V_1 & & & & & & \epsilon_1 \\
 V_4 & 0.957V_1 & & & & & \psi_4\epsilon_4 \\
 V_5 & 0.142V_1 & 0.831V_4 & & & & \psi_5\epsilon_5 \\
 V_6 & 0.153V_1 & 0.914V_4 & -0.105V_5 & & & \psi_6\epsilon_6 \\
 V_7 & 0.173V_1 & 1.852V_4 & -0.817V_5 & -0.293V_6 & & \psi_7\epsilon_7 \\
 V_2 & 0.724V_1 & & 0.079V_5 & & 0.193V_7 & \psi_2\epsilon_2 \\
 V_3 & & & & 0.307V_6 & 0.606V_7 & \psi_3\epsilon_3 \\
 V_8 & & 0.750V_4 & -1.131V_5 & & 0.715V_7 & 0.136V_3 \\
 & & & & & & \psi_8\epsilon_8
 \end{array}$$

The model can be written as

$$f_{14567}f_{2|157}f_{3|67}f_{8|3457}.$$

The vine array for the best 3-truncated is:

$$A = \begin{bmatrix} 5 & 5 & 5 & 7 & 7 & 5 & 1 & 7 \\ 7 & 7 & 5 & 5 & 6 & 7 & 5 \\ 6 & 6 & 6 & 7 & 5 & 4 \\ 1 & 1 & 1 & 6 & 6 \\ 4 & 4 & 4 & 1 \\ 3 & 3 & 3 \\ 2 & 2 \\ 8 \end{bmatrix}$$

The 3-truncated model can be written as

$$f_{5671}f_{4|756}f_{3|567}f_{2|175}f_{8|754}.$$

This is quite close to the Bayesian network: $f_{14567}f_{2|157}f_{3|67}f_{8|3457}$. From the correlation matrix of normal scores, $\rho_{14;567} = 0.071$ also $\rho_{38;457} = 0.085$ (note that the regression coefficient for V_3 is smallest in the regression of V_8 on V_3, V_4, V_5, V_7) and $\rho_{35;67} = 0.305$.

Hence the Bayesian network and 3-truncated vine are very similar. The 3-truncated model is

$$f_{5671}f_{4|756}f_{3|567}f_{2|175}f_{8|754}.$$

With pair-copulas, it can be written as:

$$f_5f_7f_6f_1f_4f_3f_2f_8c_{57}c_{56}c_{76;5}c_{71}c_{51;7}c_{61;75}c_{74}c_{54;7}c_{64;75}c_{53}c_{63;5}c_{73;56}c_{12}c_{72;1}c_{52;17}c_{78}c_{58;7}c_{48;75}$$

The BN array is:

$$\begin{bmatrix} 6 & 6 & 7 & 7 & 7 & 7 & 1 & 7 \\ 7 & 6 & 5 & 5 & 6 & 7 & 5 \\ 5 & 6 & 6 & 5\perp & 5 & 4 \\ 1 & 1 & 4\perp & - & 3 \\ 4 & - & - & - & - \\ 3 & - & - & - & - \\ 2 & - & - & - & - \\ 8 \end{bmatrix}$$

Pair-copulas include $C_{73}, C_{63;7}, C_{53;76}^\perp, C_{43;765}^\perp$ and c_{76543} must be integrated to get $C_{3|457}$ as an argument of $C_{38;754}$.

Step 0: check that variables are monotonically related. Step 1: Compare dependence structure (based on normal scores).

model	Dfit	maxabs	AIC	BIC	#par		model	Dfit	maxabs	AIC	BIC	#par
1fact	1.859	0.143	13238	13281	8		1tr	1.104	0.196	12082	12120	7
2fact	0.871	0.049	11742	11821	15		2tr	0.207	0.030	10724	10793	13
3fact	0.245	0.015	10797	10909	21		3tr	0.027	0.028	10460	10556	18
BN	0.020	0.011	10450	10552	19		4tr	0.008	0.003	10437	10555	22

Step 2. Diagnostics for tail asymmetry. The semi-correlations (correlations of normal scores in lower/upper quadrants) are included below.

j1	j2	ncor	lcor	ucor	bvnseemic	tree1	j1	j2	ncor	lcor	ucor	bvnseemic	tree1
1	2	0.973	0.974	0.891	0.930	tree1	1	7	0.907	0.926	0.687	0.782	tree1
1	3	0.832	0.832	0.515	0.646		2	7	0.918	0.933	0.728	0.805	
2	3	0.844	0.845	0.550	0.666		3	7	0.881	0.873	0.625	0.732	
1	4	0.957	0.960	0.834	0.892		4	7	0.946	0.950	0.803	0.865	tree1
2	4	0.953	0.957	0.824	0.881		5	7	0.865	0.885	0.591	0.703	tree1
3	4	0.872	0.861	0.633	0.716		6	7	0.889	0.907	0.635	0.748	
1	5	0.938	0.940	0.762	0.847		1	8	0.419	0.480	-0.084	0.209	
2	5	0.925	0.930	0.726	0.819		2	8	0.444	0.494	-0.016	0.227	
3	5	0.818	0.814	0.527	0.625	tree1	3	8	0.500	0.478	0.096	0.269	
4	5	0.967	0.956	0.877	0.915		4	8	0.457	0.481	-0.046	0.236	
1	6	0.929	0.934	0.762	0.829		5	8	0.328	0.423	-0.158	0.151	
2	6	0.918	0.923	0.734	0.806		6	8	0.420	0.489	-0.073	0.210	
3	6	0.847	0.839	0.574	0.672		7	8	0.573	0.527	0.166	0.331	tree1
4	6	0.959	0.953	0.850	0.896								
5	6	0.922	0.911	0.730	0.814	tree1							

Steps for fitting truncated vine

1. 2 and 3-parameter pair-copulas for tree 1 because $n > 1000$. BB1 (Archimedean copula based on Mittag-Leffler LT with asymmetric lower/upper tail dependence) is overall best fit (among copulas skewed to joint lower tail). In one or two edges, a 3-parameter copula could be used.
2. Get pseudo-observations from the conditional distributions after tree 1.
3. Conditional Spearman's rho as a function of conditioning variable (method of Gijbels et al): no patterns
4. Normal scores plots and semi-correlations of pseudo-observations, tail asymmetry is mostly handled with tree 1. Gaussian copula OK for trees 2 and 3, but survival Gumbel would be a small improvement for one tree 2 edge.

Theorem: To get a vine copula model with bivariate lower (upper) tail dependence for all pairs, a sufficient condition is the pair-copulas in tree 1 all have lower (upper) tail dependence.

Take home messages

- Parametrize Gaussian dependence with partial correlations $\rho_e : e \in \mathcal{E}$ that are algebraically independent and can be obtained from sequentially mixing conditional distributions.
 - Any such dependence structure can be converted to a copula model for non-Gaussian (tail) dependence.
-

References:

- Joe, H (July2014). *Dependence Modeling with Copulas*. Chapman & Hall/CRC. Software at <http://copula.stat.ubc.ca>
- Bauer, A X (2013). Pair-copula constructions for non-Gaussian Bayesian networks. PhD thesis, Technische Universität München.
- Krupskii, P (2014). Structured factor copulas and tail inference. PhD thesis, University of British Columbia. <http://vine-copula.org>