

## Comparison of dependence structures

### Part 1

- Discrepancy measures of fit to compare dependence structures
- Data example to compare factor structures, truncated vines and Bayesian network, plus mention of R code/packages
- Brief explanations of algorithms

---

### Part 2 (some theory that will be covered at a later date)

- Explanation of vine through sequential mixture of conditional distributions
- Brief mention of Sklar's theorem for copulas (probability integral transform)
- Pair-copula construction has algebraically independent bivariate copulas on edges of a vine. When all copulas on edges are Gaussian, a partial correlation vine  $\{\rho_e : e \in \mathcal{E}(\mathcal{V})\}$  is obtained and the  $\rho_e$ 's are algebraically independent. This is the Gaussian re-parametrization of the correlation matrix that extends to copulas.

---

Different parsimonious dependence models can be compared based on [fit statistics](#).

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

$$\mathbf{D}_{\text{model}} = \log(\det[\mathbf{R}_{\text{model}}(\hat{\boldsymbol{\delta}})]) - \log(\det[\mathbf{R}_{\text{data}}]) + \text{tr}[\mathbf{R}_{\text{model}}^{-1}(\hat{\boldsymbol{\delta}})\mathbf{R}_{\text{data}}] - \mathbf{d} \geq 0, \quad (1)$$

$\mathbf{R}_{\text{data}}$  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 vector  $\boldsymbol{\delta}$  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**, because (1) is larger for a model that nests in a less parsimonious model. Also useful are the average and max absolute deviations of the model-based correlation matrix from the empirical correlation matrix:

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](http://archive.ics.uci.edu/ml/datasets.html)

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:

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

(Note that  $V_2$  row can appear also as last or second to last).

The model can be written as

$$\mathbf{f}_{1:8} = \mathbf{f}_{4|1}\mathbf{f}_{5|14}\mathbf{f}_{6|145}\mathbf{f}_{7|1456}\mathbf{f}_{2|157}\mathbf{f}_{3|67}\mathbf{f}_{8|3457} = \mathbf{f}_{14567}\mathbf{f}_{2|157}\mathbf{f}_{3|67}\mathbf{f}_{8|3457}$$

The vine array for the best 3-truncated, based on (1), is:

$$A = (a_{kj})_{k \leq j} = \begin{bmatrix} 5 & 5 & 5 & 7 & 7 & 5 & 1 & 7 \\ & 7 & 7 & 5 & 5 & 6 & 7 & 5 \\ & & 6 & \mathbf{6} & \mathbf{6} & \mathbf{7} & \mathbf{5} & \mathbf{4} \\ & & & 1 & - & - & - & - \\ & & & & 4 & - & - & - \\ & & & & & 3 & - & - \\ & & & & & & 2 & - \\ & & & & & & & 8 \end{bmatrix}$$

The vine array for a vine is not unique, but given the vine array which has an ordering, there is a derived Bayesian network. The Bayesian network from a truncated vine has directed edges from variable  $a_{kj}$  to  $a_{jj}$  (down each column) for  $k = 1, \dots, \min\{3, j-1\}$ ; for example, in column 5, variables 7,5,6 are each directed into variable 4. The 3-truncated model can be written as

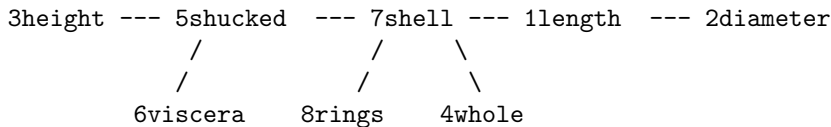
$$f_{1:8} = \mathbf{f}_5\mathbf{f}_7|_5\mathbf{f}_6|_57\mathbf{f}_1|_756\mathbf{f}_4|_756\mathbf{f}_3|_567\mathbf{f}_2|_175\mathbf{f}_8|_754 = \mathbf{f}_{5671}\mathbf{f}_4|_756\mathbf{f}_3|_567\mathbf{f}_2|_175\mathbf{f}_8|_754.$$

This is quite close to the Bayesian network:  $\mathbf{f}_{14567}\mathbf{f}_{2|157}\mathbf{f}_{3|67}\mathbf{f}_{8|3457}$ . 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 similar.

It was shown through normal scores plots that bivariate margins are not Gaussian. The advantage of the 3-truncated vine over the Bayesian network that comes from **bnlearn** is that the non-Gaussian copula version of the graph is computationally more tractable. [Explanations to come later.]

For the optimal 3-truncated vine, the edges of  $\mathcal{T}_1$  are 57, 56, 71, 74, 53, 12, 78.  $\mathcal{T}_2$  has 76|5, 51|7, 54|7, 63|5, 72|1, 58|7;  $\mathcal{T}_2$  has 61|75, 64|75, 73|56, 52|17, 48|75.



1. Length, 2. Diameter, 3. Height,
4. Whole weight, 5. Shucked weight, 6. Viscera weight, 7. Shell weight,
8. Rings.

---

Compare dependence structure (based on correlation matrix of 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

For a parsimonious dependence structure  $\mathbf{R}(\boldsymbol{\delta})$ , the optimal parameter vector  $\boldsymbol{\delta}$  minimizes (1) or equivalently maximizes the Gaussian log-likelihood with this dependence structure. For a truncated vine, the optimal parameters given the vine and truncation level are the sample partial correlations on the edges of the vine (Brechmann and Joe, 2014). But then one needs to optimize over the set of truncated vines.

Linear Markov tree:  $\boldsymbol{\delta} = (\rho_{j-1,j} : j = 2, \dots, d)$ .

Markov tree:  $\boldsymbol{\delta} = (\rho_e : e \in \mathcal{T}_1)$

2-truncated vine :  $\boldsymbol{\delta} = (\rho_e : e \in \mathcal{T}_1 \cup \mathcal{T}_2)$

$p$ -factor:  $\boldsymbol{\delta}$  are the loadings or a transformed vector.

There are mapping between regression parameters (loadings) and a truncated partial correlation vine.

---

- factor structures: **factanal** in R for factor analysis
- Bayesian network: **bn.fit** (and **hc [hillclimbing]**) in **bnlearn** R package
- truncated vine: functions in **CopulaModel** R package  
(at <http://copula.stat.ubc.ca>, compile from source); see wrapper functions on the course web site: (i) enumeration through truncated vines for  $d \leq 8$ , (ii) sequential (greedy) minimum spanning trees by tree level of the vine; (iii) genetic algorithm based on “best” spanning trees at each tree level.

For results, checking loadings of factor structure and edges of truncated vine/Bayesian network for interpretations.

---

#### Number of parameters for factor structure with $p$ latent variables

For multivariate Gaussian with standardization to  $N(0, 1)$  margins, the stochastic representation for observed variables  $Z_1, \dots, Z_d$ , latent variables  $W_1, \dots, W_p$  ( $p \ll d$ ) and disturbance variables  $\epsilon_1, \dots, \epsilon_d$  is

$$Z_j = \alpha_{j1}W_1 + \dots + \alpha_{jp}W_p + \psi_j\epsilon_j, \quad j = 1, \dots, d.$$

$$\begin{pmatrix} Z_1 \\ \vdots \\ Z_d \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{d1} & \cdots & a_{dp} \end{pmatrix} \begin{pmatrix} W_1 \\ \vdots \\ W_p \end{pmatrix} + \begin{pmatrix} \psi_1\epsilon_1 \\ \vdots \\ \psi_d\epsilon_d \end{pmatrix}.$$

The matrix  $\mathbf{A} = (\alpha_{jk})$  is called the **matrix of loadings**. The covariance matrix is  $\Sigma_Z = \mathbf{A}\mathbf{A}^T + \Psi^2$  if  $\mathbf{W}$ 's and  $\epsilon$ 's are independent  $N(0,1)$  and  $\Psi^2$  is a diagonal matrix of the  $\psi_j^2$ .  $\mathbf{A}$  is not unique because  $(\mathbf{A}\mathbf{P})(\mathbf{P}^T\mathbf{A}^T) = \mathbf{A}\mathbf{A}^T$  if  $\mathbf{P}$  is an orthogonal  $p \times p$  matrix. Model is **structured** if there are many structured zeroes in  $\mathbf{A}$ . In general, interpret factor based on larger loadings by columns.

It is possible to choose  $\mathbf{P}$  as a product of simple rotations so that (a)  $\mathbf{A}^* = \mathbf{A}\mathbf{P}$  has one zero in column 2, two zeros in column 3,  $\dots$ ,  $p-1$  zeros in column  $p$ ; (b) given the zero positions, each column of  $\mathbf{A}^*$  is unique up to sign. Hence the number of parameters for the  $p$ -factor structure is counted as  $dp - p(p-1)/2$ .

---

Bayesian network for  $d$  variables: possible idea of a greedy algorithm is below.

Want to get a permutation order  $j_1, \dots, j_d$  of variables  $1, \dots, d$  that lead to parsimonious conditional distributions.

Step 0:  $m \leftarrow d$ .

Step 1: Regress each of the  $m$  variables on the remainder, set to 0 for any  $\hat{\beta}$  that is not significantly different from 0.

Step 2:  $j_m$  is the variable with the fewest significant  $\beta$ 's on the right-hand side.

Step 3: eliminate variable  $j_m$  from further regressions;  $m \leftarrow m - 1$ ; go to Step 1.

---

Inefficient for  
large dimensions

Algorithm for abalone male normal scores data

Table of t-values for first iteration,  $i_8 = 2$ ,  $y_2 \sim y_1, y_3, y_7$  (close to `bnlearn`)

**Table is symmetric in t-values, but not  $\hat{\beta}$ 's;**

this table can also be compared with the precision matrix.

response\expl.	#sig	y1	y2	y3	y4	y5	y6	y7	y8
y1	4	-	37*	-3.7*	1.3	5.8*	6.6*	1.5	-0.7
y2	3	37*	-	2.6*	1.0	1.8	-0.7	6.2*	0.6
y3	6	-3.7*	2.6*	-	-1.6	3.6*	7.3*	12*	3.6*
y4	4	1.3	1.0	-1.6	-	51*	27*	37*	5.7*
y5	6	5.8*	1.8	3.6*	51*	-	-12*	-24*	-12*
y6	5	6.6*	-0.7	7.3*	27*	-12*	-	-14*	-2.3
y7	6	1.5	6.2*	12*	37*	-24*	-14*	-	7.9*
y8	4	-0.7	0.6	3.6*	5.7*	-12*	-2.3	7.9*	-

Table of t-values for second iteration,  $i_7 = 8$ ,  $y_8 \sim y_3, y_4, y_5, y_7$  (matches `bnlearn`)

response\expl.	#sig	y1	y3	y4	y5	y6	y7	y8
y1	5	-	-2.6*	2.7*	9.8*	8.4*	8.1*	-0.4
y3	5	-2.6*	-	-1.5	3.8*	7.3*	13*	3.6*
y4	5	2.7*	-1.5	-	52*	27*	38*	5.8*
y5	6	9.8*	3.8*	52*	-	-12*	-24*	-12*
y6	5	8.5*	-7.3*	27*	-12*	-	-14*	-2.3
y7	6	8.1*	13*	38*	-24*	-14*	-	8.1*
y8	4	-0.4	3.7*	5.8*	-12*	-2.4	8.1*	-

third iteration:  $i_6 = 3$ :  $y_3 \sim y_6, y_7, y_1, y_5$  or  $y_3 \sim y_6, y_7$  depending on sig. level (0.01 or 0.005).

fourth iteration: saturated after this for  $y_1, y_4, y_5, y_6, y_7$ ; all significant when regressing one variable on the rest.

---

Part 2: towards algebraic independence for partial correlation vine

A brief introduction to copulas and sequential mixing of conditional distributions, as this provides the proof of algebraic independence for the partial correlation vine, as well as a more general result.

**Sklar's theorem:** For a  $d$ -variate distribution  $F \in \mathcal{F}(F_1, \dots, F_d)$ , with  $j$ th univariate margin  $F_j$ , the copula associated with  $F$  is a distribution function  $C : [0, 1]^d \rightarrow [0, 1]$  that satisfies

$$F(\mathbf{y}) = C(F_1(y_1), \dots, F_d(y_d)), \quad \mathbf{y} \in \mathbb{R}^d. \quad (2)$$

If  $F$  is a continuous  $d$ -variate distribution function with univariate margins  $F_1, \dots, F_d$ , and quantile functions  $F_1^{-1}, \dots, F_d^{-1}$ , then

$$C(\mathbf{u}) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)) \quad (3)$$

is the unique choice.

Proof. If  $\mathbf{Y} \sim F$  and  $F$  is continuous, then  $(F_1(Y_1), \dots, F_d(Y_d)) \sim C$ .

---

If  $F(x_1, x_2, x_3)$  is continuous and has univariate margins  $F_1, F_2, F_3$ .

Note that there is a representation

$$F_{123}(x_1, x_2, x_3) = \int_{-\infty}^{x_2} F_{13|2}(x_1, x_3|y) dF_2(y).$$

By Sklar's theorem, there is a copula  $C_{13;2}(\cdot; x_2)$  such that

$$F_{13|2}(x_1, x_3|y) = C_{13;2}(F_{1|2}(x_1|y), F_{3|2}(x_3|y); y).$$

since  $F_{13|2}(x_1, x_3|y)$  is bivariate with univariate margins  $F_{1|2}(x_1|y)$  and  $F_{3|2}(x_3|y)$ . [Notation:  $F_{13|2}$  is a conditional distribution,  $C_{13;2}$  is not conditional]

[Simplest vine copula model by specifying  \$C\_{13;2}, C\_{12}, C\_{23}\$](#)

With conditional dependence and  $(X_1, X_2, X_3) \sim F_{1;3}$ ,  $F_{12} = C_{12}(F_1, F_2)$ ,  $F_{23} = C_{23}(F_2, F_3)$ ,

$$\mathbf{F}_{123}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = \int_{-\infty}^{\mathbf{x}_2} \mathbf{F}_{13|2}(\mathbf{x}_1, \mathbf{x}_3|\mathbf{y}) d\mathbf{F}_2(\mathbf{y}) = \int_{-\infty}^{\mathbf{x}_2} \mathbf{C}_{13;2}(\mathbf{F}_{1|2}(\mathbf{x}_1|\mathbf{y}), \mathbf{F}_{3|2}(\mathbf{x}_3|\mathbf{y}); \mathbf{y}) d\mathbf{F}_2(\mathbf{y})$$

$F_{1|2} = C_{1|2}(F_1|F_2)$ ,  $F_{3|2} = C_{3|2}(F_3|F_2)$ ,  $C_{13;2}, C_{12}, C_{23}$  algebraically independent.

---

If  $F_{1;3}$  is trivariate normal with correlations  $\rho_{12}, \rho_{13}, \rho_{23}$ , then

$$\mathbf{C}_{13;2}(\mathbf{u}, \mathbf{v}; \mathbf{y}) = \Phi_2(\Phi^{-1}(\mathbf{u}), \Phi^{-1}(\mathbf{v}); \rho_{13;2})$$

for any  $y$ , where  $\rho_{13;2}$  is the partial correlation of  $[X_1, X_3|X_2 = x_2]$ .

From construction point of view,  $C_{13;2}(u, v; y)$  are arbitrary; if for all  $y$ , these are taken as  $\Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \theta)$  then  $\rho_{12}, \rho_{23}, \rho_{13;2}$  **are algebraically independent in  $(-1, 1)$** .

With the **simplifying assumption** that  $C_{13;2}(u, v; y)$  doesn't depend on  $y$ , the [simplest vine](#) is:

$$\mathbf{F}_{1;3}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = \int_{-\infty}^{\mathbf{x}_2} \mathbf{C}_{13;2}(\mathbf{F}_{1|2}(\mathbf{x}_1|\mathbf{y}) \mathbf{F}_{3|2}(\mathbf{x}_3|\mathbf{y})) d\mathbf{F}_2(\mathbf{y}),$$

where  $F_{12} = C_{12}(F_1, F_2)$ ,  $F_{23} = C_{23}(F_2, F_3)$  (via Sklar's theorem) and  $C_{12}, C_{23}, C_{13;2}$  **are algebraically independent**. (These replace  $\rho_{12}, \rho_{23}, \rho_{13;2}$ .)

---

[Recursion](#) used when the vine copula is extended to more than 3 variables, when assuming that all distributions are absolutely continuous.

$$F_{123}(x_1, x_2, x_3) = \int_{-\infty}^{x_2} C_{13;2}(F_{1|2}(x_1|y), F_{3|2}(x_3|y)) dF_2(y)$$

For the next step, something like  $F_{3|12}$  might be needed.

$$F_{3|12}(x_3|x_1, x_2) = \frac{\partial^2 F_{123}(x_1, x_2, x_3)}{\partial x_2 \partial x_1} \Big/ f_{12}(x_1, x_2) \quad (\text{from bottom of this slide})$$

$$\frac{\partial F_{123}(x_1, x_2, x_3)}{\partial x_2} = C_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) f_2(x_2)$$

$$\frac{\partial^2 F_{123}(x_1, x_2, x_3)}{\partial x_2 \partial x_1} = C_{3|1;2}(F_{3|2}(x_3|x_2)) |F_{1|2}(x_1|x_2)| f_{1|2}(x_1|x_2) f_2(x_2)$$

$$= C_{3|1;2}(F_{3|2}(x_3|x_2))|F_{1|2}(x_1|x_2))f_{12}(x_1, x_2),$$

where

$$C_{3|1;2}(u_3|u_1) = \frac{\partial C_{13;2}(u_1, u_3)}{\partial u_1}.$$

A similar expression exists for  $F_{1|32}$ .

Note that

$$\begin{aligned} F_{3|12}(x_3|x_1, x_2) &= \lim_{\epsilon \rightarrow 0^+} \Pr(X_3 \leq x_3 | x_1 \leq X_1 < x_1 + \epsilon, x_2 \leq X_2 < x_2 + \epsilon) \\ &= \lim_{\epsilon \rightarrow 0^+} \frac{\Pr(x_1 \leq X_1 < x_1 + \epsilon, x_2 \leq X_2 < x_2 + \epsilon, X_3 \leq x_3)/\epsilon^2}{\Pr(x_1 \leq X_1 < x_1 + \epsilon, x_2 \leq X_2 < x_2 + \epsilon)/\epsilon^2} \end{aligned}$$

leading to the above.

**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 (4) has the application of Sklar's theorem, (5) make use of the simplifying assumption to get a model in this class.

$$\begin{aligned} \mathbf{F}_{\{j,k\} \cup S}(\mathbf{x}) &= \int_{-\infty}^{\mathbf{x}_S} \mathbf{F}_{jk|S}(\mathbf{x}_j, \mathbf{x}_k | \mathbf{y}_S) d\mathbf{F}_S(\mathbf{y}_S) \\ &= \int_{-\infty}^{\mathbf{x}_S} \mathbf{C}_{jk;S}(\mathbf{F}_{j|S}(\mathbf{x}_j | \mathbf{y}_S), \mathbf{F}_{k|S}(\mathbf{x}_k | \mathbf{y}_S); \mathbf{y}_S) d\mathbf{F}_S(\mathbf{y}_S) \end{aligned} \quad (4)$$

$$=^{model} \int_{-\infty}^{\mathbf{x}_S} \mathbf{C}_{jk;S}(\mathbf{F}_{j|S}(\mathbf{x}_j | \mathbf{y}_S), \mathbf{F}_{k|S}(\mathbf{x}_k | \mathbf{y}_S)) d\mathbf{F}_S(\mathbf{y}_S) \quad (5)$$

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

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

**An example of sequential mixing 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 :  $\rho_{12}, \rho_{23}, \rho_{24}, \rho_{35}, \rho_{13;2}, \rho_{34;2}, \rho_{25;3}, \rho_{14;23}, \rho_{45;23}, \rho_{15;234}$ . **Vine array is**

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

For vine copula with  $C_{12}, C_{23}, C_{24}, C_{35}, C_{13;2}, C_{34;2}, C_{25;3}, C_{14;23}, C_{45;32}, C_{15;234}$  etc., the sequential mixing of conditional distributions is:

with uniform margins for  $F_1, \dots, F_5$ ,  $F_{12} = C_{12}$ ,  $F_{23} = C_{23}$ , etc.,

integrate  $C_{13;2}(C_{1|2}, C_{3|2})$  to get  $C_{123}$ , [ $C_{1|2}$  comes from derivative of  $C_{12}$  etc.]

integrate  $C_{34;2}(C_{3|2}, C_{4|2})$  to get  $C_{234}$ , [ $C_{3|2}, C_{2|3}$  come from  $C_{23}$ ]

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  $C_{1234}$ , [ $C_{1|23}$  comes from partial derivatives of  $C_{123}, C_{23}$ ]  
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  $C_{12345}$  [ $C_{1|324}$  comes from partial derivatives of  $C_{1234}, C_{234}$ ]

---

**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, with the extra constraints for vines).

**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.**

---

Sequential mixture of conditional distributions for Bayesian network.

Before for abalone data:  $\mathbf{f}_{14567}\mathbf{f}_{2|157}\mathbf{f}_{3|67}\mathbf{f}_{8|3457}$ . The BN array in contrast to vine array:

$$\begin{bmatrix} 6 & 6 & 7 & 7 & 7 & 7 & 1 & 7 \\ & 7 & 6 & 5 & 5 & 6 & 7 & 5 \\ & & 5 & 6 & 6 & 5^\perp & 5 & 4 \\ & & & 1 & 1 & 4^\perp & - & 3 \\ & & & & 4 & - & - & - \\ & & & & & 3 & - & - \\ & & & & & & 2 & - \\ & & & & & & & 8 \end{bmatrix}, \quad A = \begin{bmatrix} 5 & 5 & 5 & 7 & 7 & 5 & 1 & 7 \\ & 7 & 7 & 5 & 5 & 6 & 7 & 5 \\ & & 6 & 6 & 6 & 7 & 5 & 4 \\ & & & 1 & - & - & - & - \\ & & & & 4 & - & - & - \\ & & & & & 3 & - & - \\ & & & & & & 2 & - \\ & & & & & & & 8 \end{bmatrix}$$

Pair-copulas (in column 6) for BN 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  $C_{38;754}$ . [Column 8 doesn't satisfy condition of vine array for 7543]

Otherwise,  $C_{67}, C_{75}, C_{71}, C_{74}, C_{73}, C_{12}, C_{78}$  are specified,

$C_{751}$  can be obtained via  $C_{51;7}, C_{75}, C_{71}$ ,

$C_{754}, C_{764}, C_{172}, C_{758}$  can be obtained with specified  $C_{54;7}, C_{63;7}, C_{72;1}, C_{58;7}$

---

**General comments about extending multivariate Gaussian to multivariate non-Gaussian.**

1. Parametrization of correlation matrix with positive definite constraint: nonlinear constraints for general bivariate distributions are much more complicated. There are no general constructions for multivariate non-Gaussian that gets parameters from bivariate margins (unless the construction is derives from multivariate Gaussian in some way such that elliptical distributions and Hüsler-Reiss).

2. Partial correlation vine with algebraically independent parameters: extends to vine pair-copula construction, the most flexible methodology for multivariate non-Gaussian. The power of this method has yet to be fully used in machine learning, psychometrics, etc.

3. Partial correlation parametrization of Bayesian network: the extension is called the *pair-copula Bayesian network* — because of numerical integration, it is harder to implement compared with the vine pair-copula construction.