1

Conditional and Partial Correlation For Graphical Uncertainty Models

Dorota Kurowicka[1], **Roger Cooke**

Delft University of Technology, Mekelweg 4, Delft 2628 CD, The Netherlands

Abstract: We study the relationship between partial correlation and constant conditional correlation with particular attention to copulae used in high dimensional graphical models. Sufficient and, in some cases, necessary conditions for equality are obtained. Numerical results show that the difference between partial and conditional correlation is small for the minimum information copula. When approximate equality holds, regular vines enable us to specify a correlation structure without algebraic constraints (e.g. positive definiteness) and to translate this structure into an on-the-fly sampling algorithm.

Keywords and phrases: Partial correlation, conditional correlation, conditional independence, Markov tree, copula, entropy, information, reliability model

1.1 Introduction

Mathematical models such as reliability diagrams, fault trees, accelerated life testing models, etc, rely on parameters whose values cannot always be perfectly measured. Nowadays, even elementary texts in risk and reliability prescribe uncertainty analysis for such models and present elementary methods (see eg Andrews and Moss 1993). Elementary methods inevitably assume that the uncertainties over different parameters are independent. This is often unrealistic. Methods for uncertainty analysis with dependence are currently an active research topic. This article develops tools for representing dependence in high dimensional distributions, such as those arising in uncertainty analysis of large fault trees.

[1] On leave from Gdynia Technical University

The Markov tree method for specifying dependence in high dimensional distributions permits on the fly sampling and has attractive theoretical features (see section 1.3). However, it is limited by the fact that only a 'treefull' of constraints can be specified. Another popular approach (Iman and Conover 1982) abandons on the fly sampling. A large sample matrix is held in memory and transformed to realize a given (rank) correlation matrix. For large problems, many cells of the correlation matrix will typically be unspecified, and this approach encounters the so called matrix completion problem (Laurent 1999): can a partially specified matrix be extended to a positive (semi) definite matrix? If an extension is possible, which extenions should be used? Furthermore, for large problems, holding a sample matrix in memory imposes unwelcome tradeoffs between speed and accuracy. Vines promise to combine the advantages of both approaches while avoiding the matrix completion problem. The key element is this: when conditional rank correlation is held constant, the partial correlation and mean conditional product moment correlation are approximately equal.

We first discuss partial and conditional correlation, and the graphical models in which these are used. We then study conditions under which these two correlations are identical. After introducing the Fréchet, the diagonal band and the minimum information copulae, we present numerical results.

1.2 Partial and conditional correlation

For variables X_1 and X_2 with zero mean and standard deviations σ_1 and σ_2, let b_{12} be the number which minimizes

$$E(X_1 - b_{12}X_2)^2.$$

The product moment or Pearson correlation $\rho(X_1, X_2)$ between X_1 and X_2 is defined as

$$\rho(X_1, X_2) = (b_{12}b_{21})^{\frac{1}{2}}.$$

It is easy to show that $\rho(X_1, X_2) = \frac{\mathrm{Cov}(X_1,X_2)}{\sigma_1\sigma_2}$. Consider variables X_i with zero mean and standard deviations $\sigma_i, i = 1, \ldots, n$.
Let the numbers $b_{12;3,\ldots,n}, \ldots, b_{1n;3,\ldots,n-1}$ minimize

$$E(X_1 - b_{12;3,\ldots,n}X_2 - \ldots - b_{1n;3,\ldots,n-1}X_n)^2;$$

then the partial correlations are defined as (Yule and Kendall 1965):

$$\rho_{12;3,\ldots,n} = (b_{12;3,\ldots,n} b_{21;3,\ldots,n})^{\frac{1}{2}}, \text{ etc.}$$

Partial correlations can be computed from correlations with the following recursive formula:

$$\rho_{12;3,\ldots,n} = \frac{\rho_{12;3,\ldots,n-1} - \rho_{1n;3,\ldots,n-1} \cdot \rho_{2n;3,\ldots,n-1}}{\sqrt{1-\rho^2_{1n;3,\ldots,n-1}}\sqrt{1-\rho^2_{2n;3,\ldots,n-1}}}. \tag{1.1}$$

The conditional correlation of Z and Y given X;

$$\rho_{YZ|X} = \rho(Y|X, Z|X).$$

is the product moment correlation computed with the conditional distribution given X. In general this depends on the value of X, but it may be constant. Letting F_X, F_Y denote the cumulative distribution functions of X and Y; the rank correlation between X and Y is:

$$r(X,Y) = \rho(F_X(X), F_Y(Y)).$$

For the joint normal distribution, partial and conditional correlations coincide.

We define the mean absolute difference between partial and conditional correlation or conditional rank correlation as

$$\begin{aligned} \Delta(YZ|X) &= E(|\rho_{YZ;X} - \rho_{YZ|X}|), \\ \Delta_r(YZ|X) &= E(|\rho_{YZ;X} - r_{YZ|X}|). \end{aligned}$$

If Y and Z are independent conditional on X, then of course $r_{YZ|X} = \rho_{YZ|X} = 0$ and we write

$$\Delta(YZ|X) = \Delta.$$

We shall see in section 1.4 that Δ may be quite large, though a sharp upper bound is not known at present.

1.3 Trees, Vines and Copula's

Trees and vines are graphical modelling tools for specifying dependence structures in high dimensional distributions. We restrict attention to variables with a uniform distribution on $[0,1]$ and present the main concepts informally. A tree on N variables specifies at most $N-1$ edges between the variables. Each edge may be associated with a copula, that is a distribution on $[0,1]^2$ with uniform marginals. Popular copulae in this context are the diagonal band (Cooke and Waij 1986) and the minimum information copulae (Meeuwissen and Bedford 1997); these copulae are continuous and can realize any correlation value in $[-1,1]$ (for the other copulae see (Dall'Aglio et al. 1991) and (Nelsen,1999)). Given any tree on N variables with copulae on the edges, a joint distribution can always be constructed satisfying the tree-copulae specification. Moreover, it can be shown (Cooke 1997) that there is a unique minimum information joint distribution satisfying the tree-copulae specification and under this distribution the tree becomes a Markov tree. Distributions specified in this way can be sampled on the fly. The tree-copulae method of specifying a joint distribution is limited by the fact that there can be at most $N-1$ edges on the tree.

A vine on N variables is a nested set of trees, where the edges of tree j are the nodes of tree $j+1$; $j=1,\ldots,N-2$, and each tree has the maximum number of edges. A regular vine on N variables is a vine in which two edges in tree j are joined by an edge in tree $j+1$ only if these edges share a common node, $j=1,\ldots,N-2$. There are $(N-1)+(N-2)+\ldots+1=\frac{N(N-1)}{2}$ edges in a regular vine on N variables. Each edge in a regular vine may be associated with a constant conditional rank (Conditional rank correlations are implemented in the sampling algorithms; however, as we know the conditional copula distributions and the relation between rank and mean product moment correlations for these distributions, we could just as well associate mean conditional product moment correlations) correlation (for $j=1$ the conditions are vacuous) and, using the diagonal band or minimum information copulae, a unique joint distribution satisfying the vine-copulae specification with minimum information can be constructed and sampled on the fly (Cooke 1997). Moreover, the (constant conditional) rank correlations may be chosen arbitrarily in the interval $[-1,1]$.

Figure 1 shows a regular vine on 5 variables. The four nested trees are distinguished by the line style of the edges; tree 1 has solid lines, tree 2 has dashed lines, etc.The conditional rank correlations associated with each edge are determined as follows: the variables reachable from a given edge are called the constraint set of that edge. When two edges are joined by an edge of the next tree, the intersection of the respective constraint sets are the conditioning variables, and the symmetric difference of the constraint sets are the conditioned variables. The regularity condition insures that the symmetric difference of the constraint sets always contains two variables. Note that each pair of variables occurs once as conditioned variables.

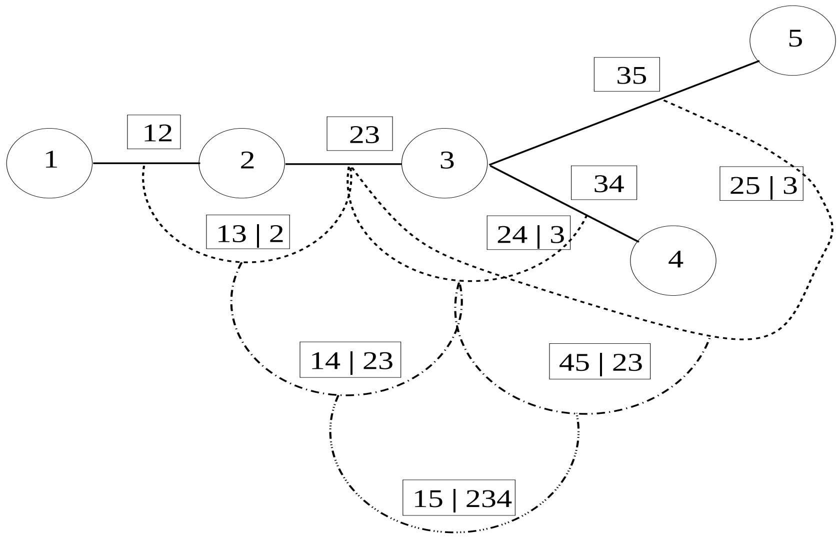

Figure 1: A regular vine on 5 variables

The edges of a regular vine may also be associated with partial correlations, with values chosen arbitrarily in the interval $(-1, 1)$. Using the recursive formulae (1.1) it can be shown that each such partial correlation regular vine uniquely determines the correlation matrix, and every full rank correlation matrix can be obtained in this way (Bedford and Cooke 1999). In other words, a regular vine provides a bijective mapping from $(-1, 1)^{N(N-1)/2}$ into the set of positive definite matrices with 1's on the diagonal. One verifies that ρ_{ij} can be computed from the sub-vine generated by the constraint set of the edge whose conditioned set is $\{i, j\}$ using recursive the formulae (1.1).We can determine numerically the mean conditional product moment correlation for a given constant conditional rank correlation. If this mean product moment correlation were (approximately) equal to the partial correlations, then the recursive formulae (1.1) could be applied to (approximately) compute the entire correlation matrix of the joint distribution constructed from the regular vine-copula specification. Alternatively, an arbitrary correlation matrix could be used to compute the partial correlations on a regular vine, and these in turn used to determine the constant conditional rank correlations, and to (approximately) sample the distribution on the fly. The degree to which partial correlations and mean constant conditional product moment correlations agree is a property of the copulae used, and the correlation values themselves.

1.4 Conditions for $\Delta = 0$

The following example shows that Δ may be large.

Proposition 1 Let

(a) X is distributed uniformly on an interval $[0,1]$,

(b) Y, Z are independent given X,

(c) $Y|X$ and $Z|X$ are distributed uniformly on $[0, X^k], k > 0$,

then

$$\Delta \;=\; \frac{3k^2(k-1)^2}{4(k^4+4k^2+3k+1)} \to \frac{3}{4} \tag{1.2}$$

The case $k = 2$ was proposed by P. Groeneboom.

Proof: We get
$E(Y) = E(Z) = E(E(Y|X)) = E(\frac{X^k}{2}) = \frac{1}{2(k+1)}$,
$E(Y^2) = E(Z^2) = E(E^2(Y|X)) = E(\frac{X^{2k}}{3}) = \frac{1}{3(2k+1)}$,
$\mathrm{Var}(Y) = \mathrm{Var}(Z) = \frac{1}{3(2k+1)} - (\frac{1}{2(k+1)})^2$,
$E(XY) = E(XZ) = E(E(XY|X)) = E(X(E(Y|X)) = E(\frac{X^{k+1}}{2}) = \frac{1}{2(k+2)}$,
$\mathrm{Cov}(X,Y) = \mathrm{Cov}(X,Z) = E(XY) - E(X)E(Y) = \frac{1}{2(k+2)} - \frac{1}{2}\frac{1}{2(k+1)}$,
$E(YZ) = E(E(YZ|X)) = E(E(Y|X)E(Z|X)) = E(\frac{X^{2k}}{4}) = \frac{1}{4(2k+1)}$,
$\mathrm{Cov}(Y,Z) = E(YZ) - E(Y)(E(Z) = \frac{1}{4(2k+1)} - \frac{1}{4(k+1)^2}$.

From the above calculations we obtain

$$\rho_{YZ} = \frac{\mathrm{Cov}(Y,Z)}{\sigma_Y \sigma_Z} \;=\; \frac{3k^2}{4k^2+2k+1}$$

and

$$\rho_{XY}\rho_{XZ} = \frac{\mathrm{Cov}^2(X,Y)}{\mathrm{Var}X\,\mathrm{Var}Y} \;=\; \frac{9k^2(2k+1)}{(k+1)^2(4k^2+2k+1)}$$

$$\Delta = \frac{\rho_{YZ} - \rho_{XY}\rho_{XZ}}{\sqrt{1-\rho_{XY}^2}\sqrt{1-\rho_{XZ}^2}} \;=\; \frac{3k^2(k-1)^2}{4(k^4+2k^2+k+1)} \to \frac{3}{4} \text{ as } k \to \infty \quad \Box$$

$Y\|X$ and $Z\|X$	Δ
$[0, x]$	0.0000
$[0, x^2]$	0.0769
$[0, x^3]$	0.2126
$[0, x^4]$	0.3243
$[0, x^5]$	0.4049
$[0, x^{10}]$	0.5824
$[0, x^{100}]$	0.7348
$[0, x^{1000000}]$	0.7500

Table 1.1: Numerical results for Proposition 1.1

Table 1.1 shows some numerical results. We note that unconditional distributions of Y and Z are not uniform.

Theorem 1
Let

(a) X, Y, Z have mean 0,

(b) Y and Z be independent given X,

(c) $E(Y|X) = AX,\ E(Z|X) = BX,$ where $A, B \neq 0$

then

$$\Delta = 0. \tag{1.3}$$

Proof. Since (1.3) is equivalent to

$$\frac{\text{Cov}(Y,Z)}{\sigma_Y \sigma_Z} = \frac{\text{Cov}(Y,X)}{\sigma_Y \sigma_X} \frac{\text{Cov}(Z,X)}{\sigma_Z \sigma_X}$$

it suffices to show that

$$\sigma_X^2 \text{Cov}(Y,Z) = \text{Cov}(Y,X)\text{Cov}(Z,X) \tag{1.4}$$

We get

$$E(ZY) = E(E(YZ|X)) = E(E(Y|X)(E(Z|X)) = AB\sigma_X^2,$$
$$E(XY) = E(XE(Y|X)) = A\sigma_X^2,$$
$$E(XZ) = B\sigma_X^2.$$

Hence (1.3) holds. □
Theorem 2
Suppose (1.3) holds and that

(a) X, Y, Z have mean 0,

(b) Y and Z are independent given X,

(c) $E(Y|X) = E(Z|X)$;

then

$$E(Y|X) = AX, \text{ where } A \neq 0.$$

Proof. If (1.3) holds then (1.4) is satisfied as well. From (b) and (c) we get

$$\sigma_X^2 E([E(Y|X)]^2) = [E(XE(Y|X))]^2.$$

Applying Cauchy-Schwarz inequality in the case of equality we obtain

$$E(Y|X) = AX. \square$$

1.5 Copulae

Definition 1
A copula Ξ is a class of joint distributions on the unit square with uniform marginals.
In the following we transform the unit square to $[-\frac{1}{2}, \frac{1}{2}]^2$ to simplify the calculations.
Definition 2
Random variables X and Y are joined by copula f if and only if their joint distribution has density

$$f(F_X^{-1}(u_1), F_Y^{-1}(u_2)).$$

The simplest copula is the Fréchet copula, $\Xi = \{f_1, f_{-1}\}$ where

$$f_1(u_1, u_2) = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } u_1 = u_2 \\ 0 & \text{otherwise} \end{cases}$$

and

$$f_{-1}(u_1, u_2) = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } u_1 = -u_2 \\ 0 & \text{otherwise} \end{cases}$$

where $u_1, u_2 \in [-\frac{1}{2}, \frac{1}{2}]$.
Let $\Phi(A)$ denote the class of mixtures of f_1 and f_{-1} with parameter $A \in [0,1]$ on the unit square $[-\frac{1}{2}, \frac{1}{2}]^2$, that is $f \in \Phi(A)$ if

$$f(u_1, u_2) \quad = \quad Af_1(u_1, u_2) + (1 - A)f_{-1}(u_1, u_2).$$

It is easy to see that this mixture of the Fréchet copulas has linear regression. For the variables X, Y, Z joined by the mixtures of the Fréchet copulas the assumptions of Theorem 1 are fulfilled so in this case (1.1) holds.
The *diagonal band* distribution on the unit square $[-\frac{1}{2}, \frac{1}{2}]^2$ is given below. For the positive correlation the mass is concentrated on the diagonal band with vertical bandwidth $\beta = 1 - \alpha$. Mass is distributed uniformly on the inscribed rectangle and is uniform but is "twice as thick" in the triangular corners. We can easily verify that the mass on the rectangle is equal to $\frac{1}{2\beta}$ and on the triangles $\frac{1}{\beta}$. For negative correlation the band is drawn between the other corners.

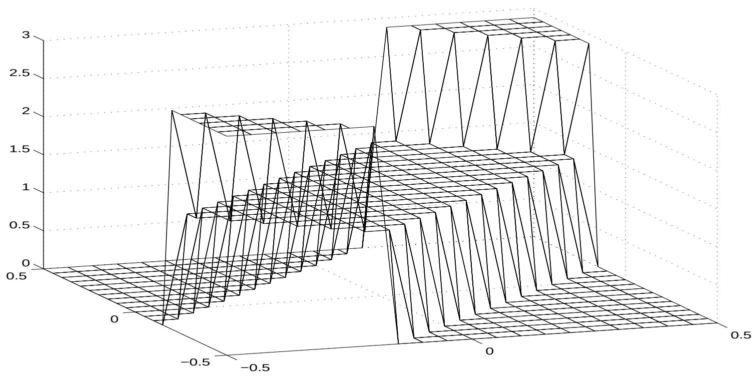

Figure 2: The diagonal band distribution with correlation 0.8

The regression for the diagonal band distribution are given by the following formulae:

$0 \leq \alpha \leq 0.5$

$$E(Y|X,\alpha) = \begin{cases} \frac{1}{2(1-\alpha)}[X^2 + X + (\alpha - 0.5)^2], & -0.5 \leq X \leq -0.5 + \alpha \\ \frac{\alpha}{1-\alpha}X, & -0.5 + \alpha \leq X \leq 0.5 - \alpha \\ -\frac{1}{2(1-\alpha)}[X^2 - X + (\alpha - 0.5)^2], & 0.5 - \alpha \leq X \leq 0.5 \end{cases}$$

$0.5 \leq \alpha \leq 1$

$$E(Y|X,\alpha) = \begin{cases} \frac{1}{2(1-\alpha)}[X^2 + X + (\alpha - 0.5)^2], & -0.5 \leq X \leq 0.5 - \alpha \\ X, & 0.5 - \alpha \leq X \leq -0.5 + \alpha \\ -\frac{1}{2(1-\alpha)}[X^2 - X + (\alpha - 0.5)^2], & -0.5 + \alpha \leq X \leq 0.5 \end{cases}$$

For the negative α we can find the regression as follows

$$E(Y|X = x, \alpha) = E(Y|X = -x, -\alpha).$$

The correlation coefficient can be calculated from the formula (Cooke and Waij 1986)

$$\rho = \operatorname{sgn}(\alpha)((1-|\alpha|)^3 - 2(1-|\alpha|)^2 + 1).$$

The *minimum information* copula is the joint density function $g(x,y)$ with minimal relative information with respect to the uniform density given uniform marginals and a given correlation. The density $g(x,y)$ has functional form (Bedford and Meeuwissen 1997)

$$g(x,y) = \kappa(x)\kappa(y)e^{\theta xy}$$

for (x,y) in unit square $[-\frac{1}{2},\frac{1}{2}]^2$. Function $\kappa(x)$ is even around $x=0$.

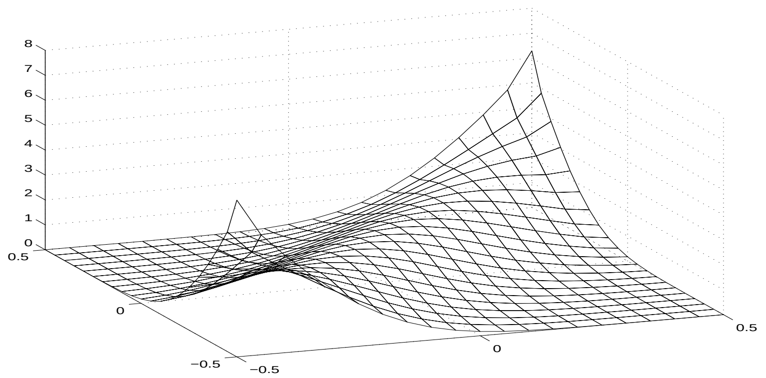

Figure 3: The minimum information distribution with correlation 0.8

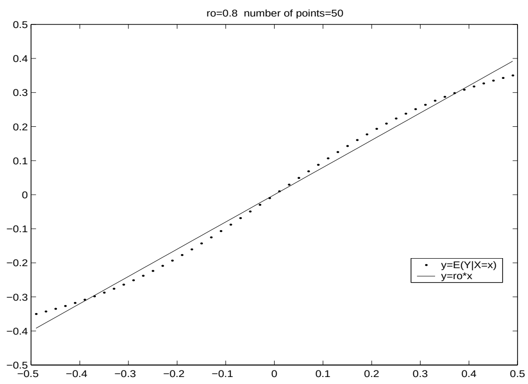

Figure 4: The conditional expectation for minimum information distribution with correlation 0.8.

1.6 Computing Δ

Let us consider variables X, Y uniform on $[-\frac{1}{2}, \frac{1}{2}]$. We may write

$$E(Y|X) = kX + \epsilon(X).$$

We want to find the coefficient k which minimizes the square error given by

$$\int (E(Y|X) - kX)^2 \, dx.$$

Setting the derivative with respect to k equal to 0:

$$\begin{aligned} \frac{d}{dx} \int (E(Y|X) - kX)^2 \, dx &= 0, \\ \int X(E(Y|X) - kX) \, dx &= 0, \\ \int XE(Y|X) \, dx &= \int kX^2 \, dx, \\ \mathrm{Cov}(X, Y) &= k\sigma_X^2 \end{aligned}$$

and finally since $\sigma_X = \sigma_Y$

$$k = \frac{\mathrm{Cov}(X,Y)}{\sigma_X^2} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \rho.$$

Thus, the best approximation of the regression is the line with coefficient equal to ρ.
The mean square difference with linear regression is equal to the variance of the conditional expectation minus a correction term $\rho^2\sigma_X^2$.
Proposition 2

$$\int (E(Y|X) - \rho X)^2 \, dx = \mathrm{Var}(E(Y|X)) - \rho^2\sigma_X^2.$$

Proof:

$$\begin{aligned} \int (E(Y|X) - \rho X)^2 \, dx &= \int E(Y|X)^2 \, dx - 2\rho^2 \mathrm{Cov}(X,Y) + \rho^2\sigma_X^2 \\ &= \int E(Y|X)^2 \, dx - \rho^2\sigma_X^2 \\ &= \mathrm{Var}(E(Y|X)) - \rho^2\sigma_X^2. \ \Box \end{aligned}$$

Theorem 3 Let us consider variables X, Y, Z uniform on $[-\frac{1}{2}, \frac{1}{2}]$ and such that Y and Z independent given X then

$$\Delta = \frac{\rho_{YZ} - \rho_{XY}\rho_{XZ}}{\sqrt{1-\rho_{XY}^2}\sqrt{1-\rho_{XZ}^2}} = \frac{\frac{\int \epsilon_Y(X)\epsilon_Z(X)\,dx}{\sigma_Y\sigma_Z}}{\sqrt{1-\rho_{XY}^2}\sqrt{1-\rho_{XZ}^2}},$$

where

$$\begin{aligned}\epsilon_Y(X) &= E(Y|X) - \rho_{XY}X \\ \epsilon_Z(X) &= E(Z|X) - \rho_{ZY}X.\end{aligned}$$

Proof. We get

$$E(YZ) = E(E(YZ|X)) = E(E(Y|X)E(Z|X)) =$$
$$\rho_{XY}\rho_{XZ}\sigma_X^2 + \rho_{XY}\int X\epsilon_Z(X)\,dx + \rho_{XZ}\int X\epsilon_Y(X)\,dx + \int \epsilon_Y(X)\epsilon_Z(X)\,dx.$$

Since

$$\begin{aligned}\int X\epsilon_Z(X) &= \int X\epsilon_Y(X) \\ &= \int (XE(Y|X) - \rho_{XY}X^2)\,dx \\ &= \mathrm{Cov}(X,Y) - \rho_{XY}\sigma_X^2 = \mathrm{Cov}(X,Y) - \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}\sigma_X^2 = 0\end{aligned}$$

then we get

$$E(YZ) = \rho_{XY}\rho_{XZ}\sigma_X^2 + \int \epsilon_Y(X)\epsilon_Z(X)\,dx.$$

We can easily calculate that

$$E(XY) = \rho_{XY}\sigma_X^2 \text{ and } E(XZ) = \rho_{XZ}\sigma_X^2.$$

Hence we get

$$\begin{aligned}|\rho_{YZ} - \rho_{XY}\rho_{XZ}| &= \left|\frac{E(YZ)}{\sigma_Y\sigma_Z} - \frac{E(XY)E(XZ)}{\sigma_Y\sigma_Z\sigma_X^2}\right| \\ &= \frac{1}{\sigma_Y\sigma_Z}\left|\int \epsilon_Y(X)\epsilon_Z(X)\,dx\right|\end{aligned}$$

which concludes the proof. □

Remark

In the case when $E(Y|X) = E(Z|X)$ we get

$$\Delta = \frac{|\rho_{YZ} - \rho^2|}{1-\rho^2} = \frac{\mathrm{Var}(E(Y|X))}{\sigma_X^2(1-\rho^2)} - \frac{\rho^2}{1-\rho^2}. \tag{1.5}$$

Next we examine the difference between conditional and partial correlation when we assume that conditional correlation is constant.

Theorem 4

Let

(a) X, Y, Z be uniform on$[-\frac{1}{2}, \frac{1}{2}]$,

(b) $E(Y|X) = E(Z|X) = AX$, where $A \neq 0$,

(c) $\sigma_{Y|X} = \sigma_{Z|X}$,

(d) $\rho_{YZ|X} = r$,

then

$$\rho_{YZ;X} = r.$$

Proof. It is easy to see that

$$\rho_{XZ} = \rho_{XY} = A.$$

Since

$$r = \rho_{YZ|X} = \frac{\mathrm{Cov}(Y|X, Z|X)}{\sigma_{Y|X}\sigma_{Z|X}} = \frac{E(YZ|X) - A^2X^2}{\mathrm{Var}(Y|X)}$$

then

$$E(YZ|X) = r\mathrm{Var}(Y|X) - A^2X^2.$$

From the above we get

$$\rho_{YZ} = \frac{E(E(YZ|X))}{\sigma_X^2} = \frac{E(r\mathrm{Var}(Y|X) - A^2X^2)}{\sigma_X^2}.$$

Since

$$E(\mathrm{Var}(Y|X)) = \mathrm{Var}(Y) - \mathrm{Var}(E(Y|X)) = \sigma_X^2 - A^2\sigma_X^2 = \sigma_X^2(1 - A^2)$$

then

$$\rho_{YZ} = \frac{r\sigma_X^2(1 - A^2) + A^2\sigma_X^2}{\sigma_X^2} = r(1 - A^2) + A^2.$$

Finally we can calculate the partial correlation Y, Z given X

$$\rho_{YZ;X} = \frac{\rho_{YZ} - \rho_{XY}\rho_{XZ}}{\sqrt{1 - \rho_{XY}^2}\sqrt{1 - \rho_{XZ}^2}} = \frac{r(1 - A^2) + A^2 - A^2}{1 - A^2} = r. \quad \square$$

Theorem 5
Let

(a) X, Y, Z be uniform on$[-\frac{1}{2}, \frac{1}{2}]$

(b) Y, X and Z, X be joined by the mixture of the Fréchet copulas with parameters A_Y, A_Z respectively,

(c) $\rho_{YZ|X} = r$,

then

$$\rho_{YZ;X} = r.$$

Proof. We get

$$\begin{aligned}
E(Y|X) &= X(2A_Y - 1),\\
E(Z|X) &= X(2A_Z - 1),\\
E((Y|X)^2) &= X^2(1 - A_Y) + X^2 A_Y = X^2,\\
\mathrm{Var}(Y|X) &= X^2 - X^2 A_Y^2 = X^2(1 - A_Y^2).
\end{aligned}$$

From the above we obtain

$$\sigma_{Y|X} = X\sqrt{1 - A_Y^2} \text{ and } \sigma_{Z|X} = X\sqrt{1 - A_Z^2}.$$

We also get

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sigma_Y \sigma_X} = \frac{E(xE(Y|X))}{\sigma_Y \sigma_X} = \frac{A_Y \sigma_X^2}{\sigma_X^2} = A_Y.$$

Analogously we obtain

$$\rho_{XZ} = A_Z.$$

From the above calculations and by (c)

$$r = \rho_{YZ|X} = \frac{\mathrm{Cov}(Y|X, Z|X)}{\sigma_{Y|X}\sigma_{Z|X}} = \frac{E(YZ|X) - A_Y A_Z X^2}{X^2\sqrt{1 - A_Y^2}\sqrt{1 - A_Z^2}}.$$

Hence we calculate that

$$\begin{aligned}
E(YZ|X) &= r\sqrt{1 - A_Y^2}\sqrt{1 - A_Z^2}X^2 + A_Y A_Z X^2\\
&= X^2(r\sqrt{1 - A_Y^2}\sqrt{1 - A_Z^2} + A_Y A_Z)
\end{aligned}$$

and find

$$\begin{aligned}
\rho_{YZ} &= \frac{E(E(YZ|X))}{\sigma_Y \sigma_Z}\\
&= \frac{E(X^2(r\sqrt{1 - A_Y^2}\sqrt{1 - A_Z^2} + A_Y A_Z))}{\sigma_X^2}\\
&= \frac{\sigma_X^2(r\sqrt{1 - A_Y^2}\sqrt{1 - A_Z^2} + A_Y A_Z)}{\sigma_X^2}\\
&= r\sqrt{1 - A_Y^2}\sqrt{1 - A_Z^2} + A_Y A_Z.
\end{aligned}$$

Hence partial correlation Y, Z given X is as follows

$$\rho_{YZ;X} = \frac{\rho_{YZ} - \rho_{XY}\rho_{XZ}}{\sqrt{1-\rho_{XY}^2}\sqrt{1-\rho_{XZ}^2}} \qquad = \frac{r\sqrt{1-A_Y^2}\sqrt{1-A_Z^2} + A_Y A_Z - A_Y A_Z}{\sqrt{1-A_Y^2}\sqrt{1-A_Z^2}} = r$$

which concludes the proof. □

1.7 Numerical results

We calculate Δ for several values of $\rho = \rho_{XY} = \rho_{XZ}$ using (1.5). The results are prepared in Matlab 5.3 and presented in Table 1.2. A discrete version of the minimum information distribution was obtained in Matlab 5.3 as a solution of the optimization problem. Table 1.2 contains the results for discrete minimum information distribution where the unit interval was divided uniformly into 50 equal segments. If we take a better approximation of this distribution we find that Δ becomes smaller.
The above results show that in the case of conditional independence for diagonal band and minimum information distributions the correlation between conditionally independent variables is almost equal to the product of correlation between them and the variable on which we conditionalize. In all cases Δ is lower for the minimum information copula.

ρ	Δ	
	DIAGONAL BAND	MINIMUM INFORMATION
0.1	7.08594e-6	2.0455e-6
0.2	1.54056e-4	1.0002e-5
0.3	8.64019e-4	4.0192e-5
0.4	2.82514e-3	1.6511e-4
0.5	6.84896e-3	6.1690e-4
0.6	1.34098e-2	2.0390e-3
0.7	2.16325e-2	5.8727e-3
0.8	2.92942e-2	1.3867e-2
0.9	3.27922e-2	2.2426e-2

Table 1.2: The comparison of Δ for diagonal band and minimum information distributions for $\rho = \rho_{XY} = \rho_{XZ}$.

Suppose in Table 1.3 we fix ρ_{XY}, ρ_{XZ} and the conditional rank correlation and sample using the minimum information copula. Table 1.3 compares partial correlation, and the mean conditional product moment correlation for some illustrative cases. We see that the difference between them can be in order of 4%, and thus is larger than in Table 1.2 where Y and Z are conditionally independent given X.

Stipulated			Computed				
ρ_{XY}	ρ_{XZ}	$r_{YZ\|X}$	ρ_{YZ}	$\rho_{YZ\,X}$	$E\rho_{YZ\|X}$	Δ_r	Δ
0.1	0.7	0.0	0.0702	0.0002	-1.347e-14	2.5e-4	2.5e-4
0.9	0.9	0.9	0.9795	0.8921	0.9004	0.0079	0.0083
-0.9	0.9	-0.9	-0.9739	-0.8626	-0.8469	0.0374	0.0157
-0.9	0.9	0.9	-0.6631	0.7729	0.8098	0.1271	0.0369
-0.5	0.4	0.7	0.3267	0.6635	0.6525	0.0365	0.0110
0.3	0.9	-0.2	0.1906	-0.1909	-0.1896	0.0091	0.0013
-0.1	-0.3	-0.8	-0.7173	-0.7873	-0.7719	0.0127	0.0154
0.8	0.8	0.8	0.9143	0.7619	0.7540	0.0381	0.0079

Table 1.3: The results of the simulations for minimum information distribution.

In Table 1.2 the conditional rank correlation and the conditional product moment correlation are equal and equal to zero. In Table 1.3 the conditional rank correlation is constant, and not equal to the (non constant) conditional product moment correlation. We believe that improved numerical routines will give better approximations as the stipulated correlations become more extreme.

1.8 Conclusions

We have seen that mean conditional product moment correlation under constant conditional rank correlation with the minimum information copula provides a good approximation to the partial correlation, particularly if the stipulated correlation values are less than 0.9 in absolute value. As explained in section 1.3, this means that we can (approximately) specify a correlation structure by giving the partial correlation values on a regular vine. The advantage of this is that these values are algebraically independent; they need satisfy no condition like positiveness definiteness, and the matrix completion problem does not arise. Alternatively, We can start with an arbitrary correlation matrix, and compute the partial correlations on a regular vine. Setting these equal to mean conditional product moment correlations under constant conditional rank correlations, we can retrieve the conditional rank correlations and thus, combined with the minimum information copula, determine a sampling routine.

This sampling routine works on the fly: We draw one sample vector at a time, we need not retain large numbers of sample vectors in memory.

1.9 References

1. Andrews, J.D. and Moss, T.R.(1993) *Reliability and Risk Assessment*, Wiley, New York.
2. Bedford, T.J. and Cooke R.M. (1999). Vines -a new graphical model for dependent random variables. Dept. Mathematics, T.U.Delft.
3. Cooke, R.M. (1977). Markov and entropy properties of tree- and vine-dependent variables. *Proceedings of the ASA Section on Bayesian Statistical Science.* Alexandria VA.
4. Cooke, R.M. and Waij, R. (1986). Monte Carlo sampling for generalized knowledge dependence with application to human reliability. *Risk Analysis*.pp. 335-343, 6, nr. 3
5. Dall'Aglio, G. and Kotz, S. and Salinetti, G. (1991)*Advances in Probability Distributions with Given Marginals; Beyond the Copulas* Kulwer Academic Publishers.
6. Iman, R. and Conover, W. (1982). A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics - Simulation and Computation*pp. 311-334, 11 (3)
7. Kendall and Stewart (1967) *The Advanced Theory of Statistics vol 2 Inference and Relationship*, 2end edition chapt 27 pp. 330-390,
8. Laurent, M. (1997). Cuts, matrix completions and graph rigidity. *Mathematical Programming* pp.255-283, 79;
9. Meeuwissen A.M.H. and Bedford T.J. (1997). Minimal information distributions with given rank correlation for use in uncertainty analysis. *J. Stat. Comp. And Simul.* pp. 143-175, 57, nos. 1-4,
10. Nelsen, R.B. (1999) *An Introduction to Copulas* Springer, New York.
11. Yule, G.U. and Kendall M.G. (1965)*An Introduction to the Theory of Statistics*,pp. 281-309 Charles Griffin & Co., 14th edition, London, chapt. 12, .