CONDITIONAL, PARTIAL AND RANK CORRELATION FOR THE ELLIPTICAL COPULA; DEPENDENCE MODELLING IN UNCERTAINTY ANALYSIS

D. Kurowicka, R.M. Cooke
Delft University of Technology, Mekelweg 4, 2628CD Delft, Netherlands
D.Kurowicka@its.tudelft.nl; R.M. Cooke@its.tudelft.nl

Abstract: The copula-vine method of specifying dependence in high dimensional distributions has been developed in Cooke [1], Bedford and Cooke [6], Kurowicka and Cooke ([2], [4]), and Kurowicka et all [3]. According to this method, a high dimensional distribution is constructed from two dimensional and conditional two dimensional distributions of uniform variaties. When the (conditional) two dimensional distributions are specified via (conditional) rank correlations, the distribution can be sampled on the fly. When instead we use partial correlations, the specifications are algebraically independent and uniquely determine the (rank) correlation matrix. We prove that for the elliptical copulae ([3]), the conditional and partial correlations are equal. This enables on-the-fly simulation of a full correlation structure; something which here to fore was not possible.

1 Introduction

A unique joint distribution is specified by associating a copula (that is, a bivariate distribution on the unit square with uniform marginals) with each (conditional) bivariate distribution. An open question concerns the relation between conditional rank correlation and partial correlation. This is important for the following reason. Bedford and Cooke ([6]) show that a bijection exists between partial correlations on a regular vine and correlation matrices. Specifying a dependence structure in terms of partial correlations avoids all problems of positive definiteness and incomplete specification (unspecified partial correlation on the vine may be chosen arbitrarily, in particular, they may be set equal to zero).

Kurowicka and Cooke have shown that if (X, Y) and (X, Z) are copula distributions, and if (Y, Z) are independent given X, then linear regression of the copula is sufficient, and with extra conditions necessary, for zero partial correlation. It is conjectured that linear regression is sufficient for the equality of constant conditional correlation and partial correlation (this has been proved in some special cases including the elliptical copulae discussed below). The copulae implemented in current uncertainty analysis programs (PREP/SPOP and UNICORN) do not have the linear regression property. Kurowicka and Cooke ([2]) show that with the maximum entropy copula, and constant conditional rank correlation, the mean conditional correlation is close to the partial correlation. However, the approximate equality degrades as correlations become extreme, and the relation between conditional rank correlation and mean conditional correlation must be determined numerically. The problem of exactly sampling a distributions with given marginals and given rank correlation matrix has until now remained unsolved. We present a simulation using elliptical copula.

The elliptical copulae introduced in Kurowicka et al ([3]) are continuous, have linear regression and can realize all correlations in (-1,1). These copulae are obtained from two dimensional projections of linear transformations of the uniform distribution on the unit sphere in three dimensions.

In this paper we prove that when $(X,Y)$and (X,Z) are joined by elliptical copulae, and when the conditional copula for $(Y,Z|X)$ does not depend on X, then the conditional correlation of $(Y,Z|X)$ and the partial correlation are equal. We study the relation between conditional correlation and conditional rank correlation. The elliptical copulae thus provide a satisfactory solution to the problem of specifying dependence in high dimensional distributions. Given a fully specified correlation matrix, we can derive partial correlations on a regular vine, and convert this to an on-the-fly conditional sampling routine. The sample will exactly reproduce the specified correlations (modulo sampling errors).

In contrast, the standard methods for sampling high dimensional distributions involve transforming to joint normal and taking linear combinations to induce a target correlation matrix. This is, however, not the rank correlation matrix. Indeed, it is easy to show using a result of Pearson ([5]) that not every correlation matrix can be the rank correlation matrix of a joint normal distribution. We report simulation exercises indicating that the probability of sampling a normal rank correlation matrix from the set of correlation matrices of given dimension goes rapidly to zero as the dimension gets large.

2 Rank and product moment correlations

The obvious relationship between product moment and rank correlations follows directly from their definitions.

Definition 2.1 (Product moment correlation) *The product moment correlation $\rho(X_1, X_2)$ of two variables X_1 and X_2 is given by*

$$\rho(X_1, X_2) = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)\, Var(X_2)}}.$$

Definition 2.2 (Rank correlation) *The rank correlation $r(X_1, X_2)$ of two random variables X_1 and X_2 with probability distribution F_{12} and marginal probability distributions F_1, F_2 respectively, is given by*

$$r(X_1, X_2) = \rho(F_1(X_1), F_2(X_2)).$$

For uniform variables these two correlations are equal but in general they are different. The rank correlation has some important advantages over the product moment correlation. It always exists, can take any value in the interval [-1,1], and is independent of the marginal distributions.

Definition 2.3 (Conditionl correlation) *The conditional correlation of Y and Z given X, $\rho_{YZ|X}$, is the product moment correlation calculated with the conditional distributions Y, Z given X.*

Let us consider variables X_i with zero mean and standard deviations σ_i, $i = 1, \ldots, n$. Let the numbers $b_{12;3,\ldots,n}, \ldots, b_{1n;2,\ldots,n-1}$ minimize

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

Definition 2.4 (Partial correlation)

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

Partial correlations can be computed from correlations with the following recursive formula (Yule and Kendall [7]):

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

3 Rank and product moment correlations for joint normal

K. Pearson [5], proved that if vector (X_1, X_2) has a joint normal distribution then the relationship between rank and product moment correlation is given by following formula

$$\rho(X_1, X_2) \quad = \quad 2\sin\left(\frac{\pi}{6}r(X_1, X_2)\right). \tag{2}$$

The proof of this fact is based on the property that the derivative of the density function for bivariate normals with respect to correlation is equal to the second order derivative with respect to variables x_1 and x_2.

¿From the above we can conclude, however, that not every positive definite matrix with ones on the main diagonal is the rank correlation matrix of a joint normal distribution.

Let us consider following example:

Example 3.1 *Let us consider matrix A*

$$A \quad = \quad \begin{bmatrix} 1 & 0.7 & 0.7 \\ 0.7 & 1 & 0 \\ 0.7 & 0 & 1 \end{bmatrix}.$$

We can easily check that A is positive definite. However, the matrix B, such that

$$B(i,j) = 2\sin(\frac{\pi}{6}A(i,j)) \ for \ i,j = 1, \ldots, 3,$$

that is,

$$B \quad = \quad \begin{bmatrix} 1 & 0.7167 & 0.7167 \\ 0.7167 & 1 & 0 \\ 0.7167 & 0 & 1 \end{bmatrix}.$$

is not positive definite.

This should be taken into account in any procedure where the dependence structure for a set of variables with specified marginals is induced using the joint normal distribution. This procedure consists of transforming specified marginal distributions to standard normals and inducing a dependence structure by using the linear properties of the joint normal. From the above example we can see that is not always possible to find a correlation matrix inducing a given rank correlation matrix.
In fact we show in next section that the probability that randomly chosen matrix stays positive definite after transformation (2) goes rapidly to zero with matrix dimension.

4 Sampling a positive definite matrix with regular vine

A graphical model called vines was introduced in (Cooke [1]). A *vine* on N variables is a nested set of trees, where the edges of tree j are the nodes of tree $j+1$, 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. 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 partial correlation (for $j=1$ the conditions are vacuous) with values chosen arbitrarily in the interval $(-1,1)$. The partial 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 form 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 is always a doubleton. The Figure 1 below shows vine on four variables with assigned to the edges partial correlations. It can be shown that each such partial correlation regular vine specification

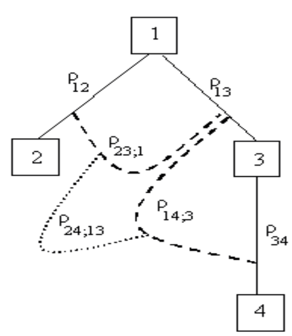

Figure 1: Partial correlation specification on regular vine on 4 variables.

uniquely determines the correlation matrix, and every full rank correlation matrix can be obtained in this way (Bedford and Cooke [6]). In other words, a regular vine provides a bijective mapping from $(-1,1)^{\binom{N}{2}}$ into the set of positive definite matrices with

1's on the diagonal.

We use the properties of a partial correlation specification on a regular vine to randomly sample positive definite matrices; we simply sample $\binom{N}{2}$ independent uniforms on (-1,1) and recalculate with formula (1) a positive definite matrix. We apply transformation (2) to the matrix obtained in this way and check if the transformed matrix stays positive definite. The table 2 shows results of simulations prepared in Matlab 5.3.

Dimension	3	4	5	6	7	8	9	10
Proportion	0.96	0.78	0.48	0.19	0.04	0.005	0.0004	0.0000

Table 1: Relationship between dimension of the matrix and proportions of the matrices retaining positive definite after transformation (2).

5 Conditional rank and product moment correlations for elliptical copulae

For copulae, that is bivariate distributions on the unit square with uniform marginals, rank and product moment correlations coincide. We are interested in the relation between conditional product moment and conditional rank correlations.

We consider the elliptical copula. This distribution has very similar properties to bivariate normal distributions. It is shown in (Kurowicka et al [3]) that it has linear regression and that partial and conditional correlations are equal if the conditional correlations are constant.

The density function of the elliptical copula with given correlation $\rho \in (-1,1)$ is

$$f_\rho(x,y) = \begin{cases} \frac{1}{\pi\sqrt{1-\rho^2}} \frac{1}{\sqrt{\frac{1}{4}-x^2-\left(\frac{y-\rho x}{\sqrt{1-\rho^2}}\right)^2}} & (x,y) \in B \\ 0 & (x,y) \notin B \end{cases}$$

where

$$B = \{(x,y)|\ x^2 + \left(\frac{y-\rho x}{\sqrt{1-\rho^2}}\right)^2 < \frac{1}{4}\}.$$

The Figure 2 shows graph of density function of the elliptical copula with correlation $\rho = 0.8$.

Some properties of elliptical copulae are shown in the theorem below (Kurowicka et al [3]).

Theorem 5.1 *If X, Y are joined by the elliptical copula with correlation ρ then*

(a) $E(Y|X) = \rho X$,

(b) $Var(Y|X) = \frac{1}{2}(1-\rho^2)\left(\frac{1}{4} - X^2\right)$,

(c) *for* $\rho X - \sqrt{1-\rho^2}\sqrt{\frac{1}{4}-X^2} < y < \rho X + \sqrt{1-\rho^2}\sqrt{\frac{1}{4}-X^2}$,

$F_{Y|X}(y) = \frac{1}{2} + \frac{1}{\pi}\arcsin\left(\frac{y-\rho X}{\sqrt{1-\rho^2}\sqrt{\frac{1}{4}-X^2}}\right)$,

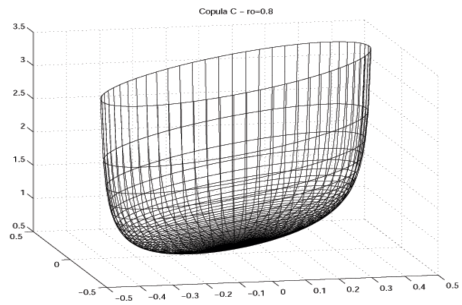

Figure 2: A density function of an elliptical copula with correlation 0.8.

(d) *for* $0 < t < 1$, $\quad F^{-1}_{Y|X}(t) = \sqrt{1-\rho^2}\sqrt{\frac{1}{4} - X^2}\sin(\pi(t-0.5)) + \rho X$.

Theorem 5.2 *Let X, Y and X, Z be joined by elliptical copula with correlations ρ_{XY} and ρ_{XZ} respectively and assume that the conditional copula for YZ given X does not depend on X; then the conditional correlation $\rho_{YZ|X}$ is constant in X.*

Proof. We calculate the conditional correlation for an arbitrary copula $f(u, v)$ using Theorem 5.1:

$$\rho_{YZ|X} = \frac{E(YZ|X) - E(Y|X)E(Z|X)}{\sigma_{Y|X}\sigma_{Z|X}}.$$

$$\begin{aligned}
E(YZ|X) &= \int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}} F^{-1}_{Y|X}(u)F^{-1}_{Z|X}(v)f(u,v)dudv = \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}} \rho_{XY}\rho_{XZ}X^2 f(u,v)dudv \\
&+ \int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}} \sqrt{1-\rho^2_{XY}}\rho_{XZ}\sqrt{\frac{1}{4} - X^2}X\sin(\pi u)f(u,v)dudv \\
&+ \int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}} \sqrt{1-\rho^2_{XZ}}\rho_{XY}\sqrt{\frac{1}{4} - X^2}X\sin(\pi v)f(u,v)dudv \\
&+ \int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}} \sqrt{1-\rho^2_{XY}}\sqrt{1-\rho^2_{XZ}}(\frac{1}{4} - X^2)\sin(\pi u)\sin(\pi v)f(u,v)dudv.
\end{aligned}$$

Since function f is a density function of the copulae and

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \sin(\pi u)du = 0$$

then we get

$$E(YZ|X) = \rho_{XY}\rho_{XZ}X^2 + \sqrt{1-\rho^2_{XY}}\sqrt{1-\rho^2_{XZ}}(\frac{1}{4} - X^2)I$$

where

$$I = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \sin(\pi u)\sin(\pi v) f(u,v) du dv.$$

¿From the above calculations and Theorem 5.1 we obtain

$$\rho_{YZ|X} = \frac{\rho_{XY}\rho_{XZ}X^2 + \sqrt{1-\rho_{XY}^2}\sqrt{1-\rho_{XZ}^2}(\frac{1}{4}-X^2)I - E(Y|X)E(Z|X)}{\sigma_{Y|X}\sigma_{Z|X}} = 2I.$$

Hence the conditional correlation $\rho_{YZ|X}$ doesn't depend on X, is constant. Moreover, this result doesn't depend on copula f. □

Now we take copula f as elliptical copula with rank correlation r. Calculating I we obtain relationship between r and $\rho_{YZ|X}$. This way the relationship between conditional product moment and conditional rank correlations will be found.

$$I = \int_{-\frac{1}{2}}^{\frac{1}{2}} du \int_{ru-\sqrt{1-r^2}\sqrt{\frac{1}{4}-u^2}}^{ru+\sqrt{1-r^2}\sqrt{\frac{1}{4}-u^2}} \frac{1}{\pi\sqrt{1-r^2}} \frac{sin(\pi u)\sin(\pi v)}{\sqrt{\frac{1}{4}-u^2-\left(\frac{v-ru}{\sqrt{1-r^2}}\right)^2}} dv.$$

Using transformation

$$u = a\cos(b); \qquad v = a\left(\sqrt{1-r^2}\sin(b) + r\cos(b)\right)$$

where

$$0 \le a \le \frac{1}{2}, \qquad 0 \le b \le 2\pi,$$

we reduce above integral to the following form

$$I = \frac{1}{\pi}\int_0^{2\pi}\int_0^{\frac{1}{2}} \frac{a}{\sqrt{\frac{1}{4}-a^2}} \sin(\pi a\cos(b)) \sin\left(\pi a\left[\sqrt{1-r^2}\sin(b) + r\cos(b)\right]\right) da db \quad (3)$$

Since the above integral is improper we first integrate by parts to get rid of the singularities and then calculate numerically. Table 2 below presents numerical results prepared in Maple.

$r_{YZ\|X}$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$\rho_{YZ\|X}$	0	0.096	0.193	0.290	0.388	0.487	0.586	0.687	0.790	0.894	1

Table 2: Relationship between conditional product moment and constant conditional rank correlation for variables joined by elliptical copula.

Acknowledgements The authors would like to thank M. de Bruin for help in finding efficient way of solving integral from this section and D. Lewandowski for preparing numerical results presented in Table 2.

6 Conclusions

These results mean that the problem of sampling exactly from a distribution with fixed marginals and given rank correlation matrix has been solved. In fact for Example 3.1 this could not be done with existing methods. The correlation specification on a regular vine with elliptical copula provides a very convenient way of sampling such a distribution. We get $\rho_{12} = \rho_{13} = 0.7$ and $\rho_{23} = 0$. The partial correlation $\rho_{23;1}$ can be calculated from (1) as -0.96. For the elliptical copula conditional product moment correlation is constant and equal to partial correlation. From (3) we compute that constant conditional rank correlation $r_{23|1} = -0.9635$ corresponds to (constant) conditional product moment correlation of -0.96. The sampling algorithm samples three independent uniform $(0,1)$ variables U_1, U_2, U_3. We assume that the variables X_1, X_2, X_3 are also uniform. Let $F_{r_{i,j|k};U_i}(X_j)$ denote the cumulative distribution function for X_j (a uniform variable, to be denoted U_j) under the conditional copula with rank correlation $r_{i,j|k}$, as a function of U_i. Then $X_j = F^{-1}_{r_{i,j|k};U_i}(U_j)$ expresses X_j as a function of U_j and U_i. The algorithm can now be stated as follows:

$$x_1 = u_1; \quad x_2 = F^{-1}_{r_{12};u_1}(u_2); \quad x_3 = F^{-1}_{r_{13};u_1} F^{-1}_{r_{23|1};u_2}(u_3).$$

Since for elliptical copula inverse cumulative distribution is given in functional form (5.1), this sampling procedure is very efficient and accurate.

References

[1] R.M. Cooke. Uncertainty modeling: examples and issues. *Safety Science*, 26, no 1/2:49–60, 1997.

[2] R.M. Cooke D. Kurowicka. Conditional and partial correlation for graphical uncertainty models. *in Recent Advances in Reldiability Theory, Birkhauser, Boston*, pages 259–276, 2000.

[3] J. Misiewicz D.Kurowicka and R.M Cooke. Elliptical copulae. *in appear*, 2000.

[4] D. Kurowicka and R.M. Cooke. A parametrization of positive definite matrices in terms of partial correlation vines. submitted to Lienear Algebra and its Applications, 2000.

[5] K. Pearson. Mathematical contributions to the theory of evolution. *Biometric*, Series. VI.Series, 1907.

[6] R.M. Cooke T.J. Bedford. Reliability methods as management tools: dependence modeling and partial mission success. In Chameleon Press, editor, *ESREL'95*, pages 153–160. London, 1995.

[7] G.U. Yule and M.G. Kendall. *An introduction to the theory of statistics*. Charles Griffin & Co. 14th edition, Belmont, California, 1965.