

---

# Copula Bayesian Networks: Supplementary Material

---

**Gal Elidan**  
Department of Statistics  
Hebrew University  
Jerusalem, 91905, Israel  
galel@huji.ac.il

## 1 Kernel Distribution and Density Estimation

We briefly describe the standard non-parametric kernel density or Parzen window (Parzen, 1962) method we use for estimating the univariate densities in our Copula Network construction. Let  $x[1], \dots, x[M]$  be i.i.d. samples of a random variable  $X$ . The kernel density approximation of its probability density function is

$$\hat{f}_h(x) = \frac{1}{Mh} \sum_{i=1}^M K\left(\frac{x - x_i}{h}\right),$$

where  $K$  is some kernel function and  $h$  is a smoothing parameter called the bandwidth. Qualitatively, the method approximates the distribution by placing small “bumps” (determined by the kernel) at each data point. Thus, higher density values will result in regions where there is a concentration of data samples. A histogram representation of a density, for example, can be thought of as placing a uniform box at each data point whose width equals to the width of the histogram bin.

A common, mathematically convenient and smoother, choice for  $K$  is the standard Gaussian density function

$$K\left(\frac{x - x_i}{h}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-x_i)^2}{2h^2}}.$$

In this case, the variance of the Kernel is controlled indirectly by the choice of the bandwidth parameter  $h$ . A common choice for  $h$  that works well in practice (and is optimal when  $X$  is known to be normally distributed) is

$$h = \hat{\sigma} \times \left[ \frac{4}{3M} \right]^{\frac{1}{5}}$$

where  $\hat{\sigma}$  is the robust estimator for the standard deviation of  $X$ :

$$\hat{\sigma} = \text{median}(|x_i - \text{median}(\{x_i\})|) / 0.6745.$$

For the distribution function, the procedure is similar with the standard normal cumulative distribution as the Kernel function.

In all the experiments considered in this work we use this simplest variant of a Gaussian Kernel estimator for each univariate density and distribution. See, for example, Bowman & Azzalini (1997) for further details and for more elaborate kernel based density estimation approaches.

## 2 Multivariate normal copula

The multivariate Gaussian copula is constructed directly via an inversion of the parameters on both sides of Sklar’s theorem (Sklar, 1959):

$$C(u_1, \dots, u_N) = \Phi_{\Sigma}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n))$$

where  $\Phi$  is the standard normal cumulative distribution function and  $\Phi_{\Sigma}$  is the standard normal cumulative distribution function with correlation matrix  $\Sigma$ .

## 2.1 Multivariate Normal Copula Density

Using  $x_i$  to denote  $\Phi^{-1}(u_i)$  so that  $u_i = \Phi(x_i)$ , we have that  $\partial u_i / \partial x_i = \varphi(x_i) = \varphi(\Phi^{-1}(u_i))$ , with  $\varphi$  denoting the standard normal density. Using this we can readily compute the multivariate Gaussian copula density

$$\begin{aligned} c(u_1, \dots, u_N) &= \frac{\partial C(u_1, \dots, u_N)}{\partial u_1 \dots \partial u_N} \\ &= \frac{\varphi_{\Sigma}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n))}{\prod_{i=1}^n \varphi(\Phi^{-1}(u_i))} \end{aligned}$$

Now, since the marginal of a multivariate normal distribution is a normal distribution with an appropriately reduced correlation matrix, we have that

$$c(u_2, \dots, u_N)_{u_1=1} = \frac{\varphi_{\Sigma_{-u_1}}(\Phi^{-1}(u_2), \dots, \Phi^{-1}(u_n))}{\prod_{i=2}^n \varphi(\Phi^{-1}(u_i))}$$

where  $\Sigma_{-u_1}$  is a reduced correlation matrix that results from removing the first row and column from  $\Sigma$ .

## 2.2 Parameter Estimation

Estimation of  $\Sigma$  via a direct maximum likelihood approach can be difficult in high dimension. However, a relationship between the correlation matrix and Kendall's tau provides an alternative. Kendall's tau is a well known measure of concordance for bivariate random vectors (Kruskal, 1958)

$$\rho_{\tau}(X, Y) = E \left[ \text{sign} \left( (X - \tilde{X})(Y - \tilde{Y}) \right) \right]$$

where the pair  $(\tilde{X}, \tilde{Y})$  has the same distribution as  $(X, Y)$ . Remarkably, for elliptical copula distributions (including the normal and student-t), if  $\rho$  denotes the copula correlation coefficient of  $X$  and  $Y$  then

$$\rho_{\tau}(X, Y) = \frac{2}{\pi} \arcsin \rho$$

Using this property, Linsskog et al. (2003) suggest the following simple procedure for estimating  $\Sigma$  both for the multivariate normal and student-t copulas. First, given  $M$  instances, an empirical estimate of Kendall's tau for all pairs of variables  $X_i, X_j$  is computed using

$$\begin{aligned} \hat{\rho}_{\tau}(X_i, X_j) &= \frac{n(n-1)}{2} \times \\ &\sum_{1 \leq m_1 < m_2 \leq M} \text{sign} \left[ \begin{array}{l} (X_i[m_1] - X_i[m_2]) \times \\ (X_j[m_1] - X_j[m_2]) \end{array} \right] \end{aligned}$$

where  $X_i[m_1]$  denotes the value of  $X_i$  in the  $m_1$ 'th instances. This yields an unbiased and consistent estimator of Kendall's tau. Using the method of moments,  $\Sigma_{ij}$  is then given by  $\sin(\frac{\pi}{2} \hat{\rho}_{\tau}(X_i, X_j))$ . One numerical issue to consider is that theoretically (though this did not happen in our experiments), this element-wise computation of  $\Sigma$  may result in a correlation matrix that is not positive definite, in which case appropriate adjustment methods can be used (e.g., Rousseeuw & Molenberghs (1993)). See Demarta & McNeil (2005) for more details on this estimation approach.

If we assume a uniform correlation structure so that the diagonal values of  $\Sigma$  are fixed to 1 and the off diagonal elements all equal to  $\rho$ , then direct optimization via a conjugate gradient method is straightforward and fairly efficient. The explicit form of the derivatives can be found in Zezula (2009).

## 3 Multivariate Frank Copula

Frank's Archimedean copula for  $N$  dimensions is defined as

$$\begin{aligned} C(u_1, \dots, u_N) &= \\ &-\frac{1}{\theta} \log \left( 1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1) \dots (e^{-\theta u_N} - 1)}{(e^{-\theta} - 1)^{N-1}} \right) \end{aligned}$$

To derive the copula density, we compute the  $N$ -order partial derivative of  $C(u_1, \dots, u_N)$ . Using  $A$  to denote the term within the logarithm, we have we have

$$\begin{aligned}\frac{\partial A}{\partial u_i} &= -\theta e^{-\theta u_i} \frac{\prod_{j \neq i} (e^{-\theta u_j} - 1)}{(e^{-\theta} - 1)^{N-1}} \\ &= -\theta(A-1) \frac{e^{-\theta u_i}}{(e^{-\theta u_i} - 1)} = -\theta(A-1)Q_i,\end{aligned}$$

where we used  $Q_i$  to denote the quotient  $\frac{e^{-\theta u_i}}{(e^{-\theta u_i} - 1)}$ . Using this we have

$$\begin{aligned}\frac{\partial C(u_1, \dots, u_N)}{\partial u_1} &= -\frac{1}{\theta} \frac{1}{A} \frac{\partial A}{\partial u_1} \\ &= \frac{(A-1)}{A} Q_1 = \left(1 - \frac{1}{A}\right) Q_1.\end{aligned}$$

From which we have

$$\frac{\partial^2 C(u_1, \dots, u_N)}{\partial u_1 \partial u_2} = \frac{1}{A^2} \frac{\partial A}{\partial u_2} Q_1 = -\theta \frac{(A-1)}{A^2} Q_1 Q_2.$$

And

$$\begin{aligned}\frac{\partial^3 C(u_1, \dots, u_N)}{\partial u_1 \partial u_2 \partial u_3} &= -\theta Q_1 Q_2 \frac{\partial A}{\partial u_3} \times \frac{A^2 - (A-1)2A}{A^4} = \\ &= -\theta^2 Q_1 Q_2 Q_3 \frac{(A^2 - 3A + 2)}{A^3}.\end{aligned}$$

Continuing we have

$$\begin{aligned}\frac{\partial^4 C(u_1, \dots, u_N)}{\partial u_1 \partial u_2 \partial u_3 \partial u_4} &= -\theta^2 Q_1 Q_2 Q_3 \frac{\partial A}{\partial u_4} \times \frac{(2A-3)A^3 - (A^2 - 3A + 2)3A^2}{A^6} \\ &= -\theta^2 Q_1 Q_2 Q_3 \frac{\partial A}{\partial u_4} \times \frac{-A^2 + 6A - 6}{A^4} \\ &= -\theta^3 Q_1 Q_2 Q_3 Q_4 \times \frac{A^3 - 7A^2 + 12A - 6}{A^4}.\end{aligned}$$

And

$$\begin{aligned}\frac{\partial^5 C(u_1, \dots, u_N)}{\partial u_1 \partial u_2 \partial u_3 \partial u_4 \partial u_5} &= +\theta^4 Q_1 Q_2 Q_3 Q_4 Q_5 \times \\ &\quad \frac{(A-1)(-A^3 + 14A^2 - 36A + 24)}{A^5}.\end{aligned}$$

While cumbersome, this can be easily continued. However, the above formulas are sufficient for our work where we only consider networks with up to 4 parents (local copulas with up to 5 variables).

To find the maximum likelihood parameters of the density, we rely on the building block

$$\begin{aligned}\frac{\partial A}{\partial \theta} &= \frac{\partial A - 1}{\partial \theta} = (A-1) \frac{\partial \log(A)}{\partial \theta} \\ &= (A-1) \left[ \sum_i -u_i Q_i + \frac{N-1}{e^{-\theta} - 1} \right].\end{aligned}$$

Using this, the derivatives of the copula density forms derived above are straightforward.

## References

- Bowman, A. and Azzalini, A. *Applied Smoothing Techniques for Data Analysis*. Oxford University Press, 1997.
- Demarta, S. and McNeil, A. The t copula and related copulas. *International Statistical Review*, 73 (1):111–129, 2005.
- Kruskal, W. Ordinal measures of association. *Journal of the American Statistical association*, 53: 814–861, 1958.
- Lindskog, F., McNeil, A., and Schmock, U. Kendall’s tau for elliptical distributions. *Credit Risk - measurement, evaluation and management*, pp. 149–156, 2003.
- Parzen, E. On estimation of a probability density function and mode. *Annals of Math. Statistics*, 33: 1065–1076, 1962.
- Rousseeuw, P. and Molenberghs, G. Transformation of non positive semidefinite correlation matrices. *Communications in Statistics - Theory and Methods*, 22:965–084, 1993.
- Sklar, A. Fonctions de repartition a n dimensions et leurs marges. *Publications de l’Institut de Statistique de L’Universite de Paris*, 8:229–231, 1959.
- Zezula, I. On Multivariate Gaussian Copulas. *Journal of Statistical Planning and Inference*, 2009.