Approximate Optimal Transport for Continuous Densities with Copulas

Jinjin Chi, Jilin University
Jihong Ouyang, Jilin University
Ximing Li, Jilin University
Yang Wang, Hefei University of Technology
Meng Wang, Hefei University of Technology

May 26, 2019

Supplement

A D-vine decomposes a copula density $\mathbf{cop}$ into a product of pair copulas [1]. More formally, a D-vine is a nest set of trees $T = \{T_1, ..., T_d, ..., T_{D-1}\}$, where each T_d contains a node set N_d and an edge set E_d. It decomposes a D-dimensional copula as follows:

$$\mathbf{cop}(z_1, ..., z_D) = \prod_{k=1}^{D-1} \prod_{j=1}^{D-k} \mathbf{cop}_{j,j+k|(j+1):(j+k-1)}((z_j|z_{j+1}, ..., z_{j+k-1}), (z_{j+k}|z_{j+1}, ..., z_{j+k-1})) \tag{1}$$

Note that for brevity we omit the CDF and copula paramters. A D-dimensional density p corresponding to a D-vine can be written as:

$$\begin{aligned} &p(z_1, ..., z_D) \\ &= \prod_{i=1}^{D} p(z_i) \times \prod_{k=1}^{D-1} \prod_{j=1}^{D-k} \mathbf{cop}_{j,j+k|(j+1):(j+k-1)}((z_j|z_{j+1}, ..., z_{j+k-1}), (z_{j+k}|z_{j+1}, ..., z_{j+k-1})) \end{aligned} \tag{2}$$

For example, two 3-dimensional densities $p(x)$ and $q(y)$ can be written as follows:

$$p(x_1, x_2, x_3) = \prod_{i=1}^{3} p_i(x_i)\mathbf{cop}(x_1, x_2)\mathbf{cop}(x_2, x_3)\mathbf{cop}(x_1, x_3|x_2) \tag{3}$$

$$q(y_1, y_2, y_3) = \prod_{i=1}^{3} q_i(y_i)\mathbf{cop}(y_1, y_2)\mathbf{cop}(y_2, y_3)\mathbf{cop}(y_1, y_3|y_2) \tag{4}$$

Then a joint distribution π with marginals $p(x)$ and $q(y)$ is:

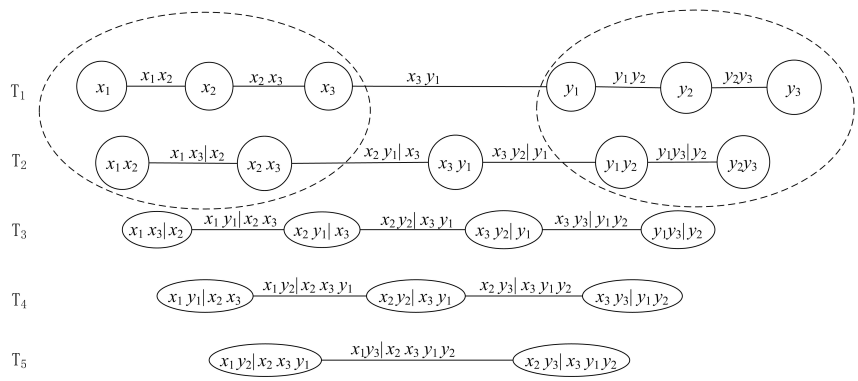

Figure 1: A D-vine with 6 variables, 5 trees and 15 edges showing conditioned (before |) and conditioning (after |) sets. Each node is a variable. Each edge associates with a pair-copula. The dotted boxes are two D-vines with 3 variables.

$$
\begin{aligned}
&\pi(x_1, x_2, x_3, y_1, y_2, y_3) \\
&= \prod_{i=1}^{3} p_i(x_i) q_i(y_i) \mathbf{cop}(x_1, x_2) \mathbf{cop}(x_2, x_3) \mathbf{cop}(x_1, x_3|x_2) \mathbf{cop}(y_1, y_2) \mathbf{cop}(y_2, y_3) \mathbf{cop}(y_1, y_3|y_2) \\
&\times \mathbf{cop}(x_3, y_1) \mathbf{cop}(x_2, y_1|x_3) \mathbf{cop}(x_3, y_2|y_1) \mathbf{cop}(x_1, y_1|x_2, x_3) \mathbf{cop}(x_2, y_2|x_3, y_1) \mathbf{cop}(x_3, y_3|y_1, y_2) \\
&\times \mathbf{cop}(x_1, y_2|x_2, x_3, y_1) \mathbf{cop}(x_2, y_3|x_3, y_1, y_2) \mathbf{cop}(x_1, y_3|x_2, x_3, y_1, y_2)
\end{aligned} \tag{5}
$$

Substituting Eq. 3 and Eq. 4 into Eq. 5 produces an alternative form:

$$
\begin{aligned}
&\pi(x_1, x_2, x_3, y_1, y_2, y_3) \\
&= p(x) q(y) \mathbf{cop}(x_3, y_1) \mathbf{cop}(x_2, y_1|x_3) \mathbf{cop}(x_3, y_2|y_1) \mathbf{cop}(x_1, y_1|x_2, x_3) \mathbf{cop}(x_2, y_2|x_3, y_1) \\
&\times \mathbf{cop}(x_3, y_3|y_1, y_2) \mathbf{cop}(x_1, y_2|x_2, x_3, y_1) \mathbf{cop}(x_2, y_3|x_3, y_1, y_2) \mathbf{cop}(x_1, y_3|x_2, x_3, y_1, y_2) \\
&= p(x) q(y) \prod_{i,j}^{3} \mathbf{cop}_{xy}(x_i, y_j)
\end{aligned} \tag{6}
$$

where $\mathbf{cop}_{xy}$ is the bivariate copula between each component of x and y.

Figure 1 is a D-vine with 6 variables, which describes the decomposition process.

References

[1] Jeffrey Dissmann, Eike C Brechmann, Claudia Czado, and Dorota Kurowicka. Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59:52–69, 2013.