

# Dissimilarity functions for rank-based hierarchical clustering of continuous variables\*

Fabrizio Durante<sup>†</sup>

The selection of a suitable stochastic model is particularly demanding in the high-dimensional setting so that, in order to guide the process of model building, various clustering procedures have been proposed to identify sub-groups of random variables that have a common behaviour. Such methods have become particularly popular in the analysis of comovements of financial time series, also in view of their impact for portfolio diversification.

Motivated by this problem, we present a theoretical framework for a (copula-based) notion of dissimilarity between subsets of continuous random variables and study its main properties. In essence, the dissimilarity measure will be of probabilistic nature, i.e. it will depend on the joint probability distribution function of the involved variables, and will be related to the degree of comonotonicity among the variables.

Specifically, we focus on those properties of dissimilarity measures that are prone to the hierarchical agglomerative methods, such as reducibility. Moreover, we discuss advantages and disadvantages of the use of measures that differ from classical linkage methods.

This is a joint work with Marta Di Lascio and Sebastian Fuchs, and the related preprint is available at <https://arxiv.org/abs/2007.04799>.

---

\*This work has been supported by the project “Stochastic Models for Complex Systems” by Italian MIUR (PRIN 2017, Project no. 2017JFFHSH).

<sup>†</sup>Dipartimento di Scienze dell’Economia, Università del Salento, Lecce (Italy), e-mail: [fabrizio.durante@unisalento.it](mailto:fabrizio.durante@unisalento.it).