Smooth testing and clustering of copulas

Yves Ismaël Ngounou Bakam* and Denys Pommeret†

Résumé

A smooth test to simultaneously compare K copulas, where $K \geq 2$, is proposed. The K observed populations can be paired. The test statistic is based on the differences between moment sequences, called copula coefficients. These coefficients characterize the copulas, even in cases where the copula densities may not exist. The procedure involves a two-step data-driven procedure. In the initial step, the most significantly different coefficients are selected for all pairs of populations. The subsequent step utilizes these coefficients to identify populations that exhibit significant differences.

We then use this test to automatically construct clusters consisting of populations with significantly similar dependence structures. The procedure is data-driven and relies on the asymptotic level of the test.

We apply our test methodology and clustering algorithm, implemented in the `Kcop R` package [Bakam and Pommeret, 2022a] to real datasets.

Mots-clefs (3 à 5) : Copula coefficients – Data-driven – Smooth test – Legendre polynomials – Nonparametric clustering – `Kcop R` package.

Context

Investigating dependence structures between multidimensional variables in various domains can provide valuable insights and inform decision-making processes. Below are a few applications that can be considered :

— Finance : recognize and managing financial risks, enhancing portfolio diversification, evaluating derivative products, and evaluating financial contagion. It provides valuable insights for investors, portfolio managers, and financial regulators in a constantly evolving market environment.

— Marketing : understanding the dependence between product sales in different retail stores can help companies optimize their inventory management, pricing strategies, and marketing efforts. For instance, if certain products tend to sell together in one store but not in another, it may suggest differences in consumer preferences or market demand, leading to tailored marketing campaigns or adjustments in product placement.

— Economics : analyzing the dependence between income and consumption in different countries can offer insights into consumer behavior, economic trends, and policy effectiveness. It can help policymakers design targeted interventions to stimulate economic growth, reduce income inequality, or encourage responsible consumption habits.

— Accounting : investigating the dependence between reported items on corporate balance sheets across different countries can aid in assessing financial risk, compliance with accounting standards, and potential exposure to economic shocks. It can inform investors, regulators, and corporate decision-makers about the financial health and resilience of multinational corporations operating in diverse environments.

Moreover, non-parametric classification of these multidimensional variables offers a structured framework for investigating and interpreting various dependence structures within the data. This facilitates deeper insights into the interrelationships among variables, enhances model refinement, and

*Univ Rennes, Ensai, CNRS, CREST - UMR 9194, F-35000 Rennes, France , yves.ngounou@ensai.fr

†Aix-Marseille University, CNRS, Centrale Marseille, I2M, Campus de Luminy, 13288 Marseille cedex 9, France, denys.pommeret@univ-amu.fr

enables more informed decision-making. It underscores the importance of robust statistical methods, such as copula-based analysis, in extracting meaningful insights from complex datasets across diverse fields.

The purpose of this presentation (Ngounou et al., 2024) is to develop, initially a new method for simultaneously comparing copulas based on data Ngounou Bakam and Pommeret [2024]. We consider K continuous random vectors, namely

$$\mathbf{X}^{(1)} = (X_1^{(1)}, \ldots, X_p^{(1)}), \ldots, \mathbf{X}^{(K)} = (X_1^{(K)}, \ldots, X_p^{(K)}),$$

with joint cdf $\mathbf{F}^{(1)}, \ldots, \mathbf{F}^{(K)}$, and with associated copulas $C_1, \ldots, C_K$, respectively. Assume that we observe K iid samples from $X^{(1)}, \ldots, X^{(K)}$, possibly paired, denoted by

$$(X_{i,1}^{(1)}, \ldots, X_{i,p}^{(1)})_{i=1,\ldots,n_1}, \ldots, (X_{i,1}^{(K)}, \ldots, X_{i,p}^{(K)})_{i=1,\ldots,n_K}.$$

where we assume that for all $1 \leq \ell < m \leq K$, $\min(n_\ell, n_m) \to \infty$, and

$$n_\ell/(n_\ell + n_m) \to a_{\ell,m}, \text{ with } 0 < a_{\ell,m} < \infty.$$

We consider the problem of testing the equality

$$H_0 : C_1 = \cdots = C_K,$$

against H_1 : there exist $1 \leq k \neq k' \leq K$ such that $C_k \neq C_{k'}$.

The test statistic is based on the differences between moment sequences, called copula coefficients. These coefficients characterize the copulas, even in cases where the copula densities may not exist [Ngounou Bakam and Pommeret, 2024].

In the second part, we present an adapted version of the previous testing procedure to establish a data-driven method for clustering K populations into N subgroups, each characterized by a common dependence structure. The unknown number N of clusters will be automatically determined by the procedure and validated through our testing method. The full details is proposed inBakam and Pommeret [2022b].

Package Kcop

The `KcopTest` function conducts a nonparametric smooth test to simultaneously compare K (where K $>$ 1) copulas, while the `KcopClust` function executes a data-driven clustering procedure to group K multivariate populations of varying sizes into N subgroups distinguished by a shared dependence structure. The number N of clusters is unknown and determined automatically by our approach.

For more details, examples, and additional information, please refer to Bakam and Pommeret [2022a].

A comprehensive application involving the financial returns of companies listed in the S&P 500 index will be provided.

Références

Yves Ismaël Ngounou Bakam and Denys Pommeret. *Kcop : Smooth Test for Equality of Copula and Clustering Multivariate*, 2022a. URL https://CRAN.R-project.org/package=Kcop. R package version 1.0.0.

Yves Ismaël Ngounou Bakam and Denys Pommeret. Non-parametric clustering of multivariate populations with arbitrary sizes. *arXiv preprint arXiv :2211.06338*, 2022b.

Yves Ismaël Ngounou Bakam and Denys Pommeret. Smooth test for equality of copulas. *Electronic Journal of Statistics*, 18(1) :895–941, 2024.