Modeling and approximation of copulas for complex data structures using Cramér-von Mises statistic

Eckhard Liebscher

University of Applied Sciences Merseburg, Germany

Abstract

This talk addresses the recent challenges associated with analysing multivariate data. The objective is to approximate the copula of the random vector under consideration using an appropriate model. Goodness-of-fit tests often reveal that many null hypotheses are not rejected. This situation is due to the high complexity of the multivariate distribution and the limited size of the sample. In this talk, we present a method for overcoming this problem. Rather than seeking an exact fit, we search for a reasonable approximation of the sample copula. The aim is to find a copula from a parametric family or from families which approximates best the sample copula.

The specific challenges of complex data sets of a random vector $X = (X^{(1)}, \dots, X^{(d)})$ are:

(i) the distribution of at least one $X^{(j)}$ has discrete components,

(ii) asymmetry of the copula and different correlations,

(iii) high-dimensional random vectors,

(iv) copulas with special structure.

Regarding (i), the investigations are limited to positive random vectors, where zero values occur with positive probability (zero-inflated data). In this case we are not able to approximate the whole copula which is uniquely determined only on the codomain of the marginal distributions. We therefore work with a surrogate function $\widetilde{C}$ for C.

Let C denote the copula of the sample. The copula C is modelled by a parametric family $\mathcal{M} = \{C_\theta\}_{\theta\in\Theta}$ of copulas. Θ is the compact parameter space. Against the above background, we assume that C does not belong to $\mathcal{M}$. We consider the Cramér-von-Mises divergence as a measure of discrepancy between the copula C and the model class $\mathcal{M}$. This has the advantage that copula densities, which can have complicated formulas, do not need to be evaluated. The parameter θ_0 of the best approximation is estimated by the approximate minimum distance estimator $\hat{\theta}_n$. Under certain regularity conditions, we have proved that $\hat{\theta}_n$ is a consistent estimator for θ_0. Moreover we present an asymptotic normality result for $\hat{\theta}_n$. This was achieved by applying a central limit theorem for U-statistics. We show how comparisons of the approximation by several copula models can be done.

We provide a method for treating the distribution of high-dimensional random vectors by using product copulas which are mostly asymmetric. Finally, the use of the methods is illustrated by means of examples.

Keywords

asymmetric copulas, approximation of copulas, zero-inflated data, high-dimensional random vectors.

References

[1] Liebscher, E. (2009). Semiparametric Estimation of the Parameters of Multivariate Copulas. *Kybernetika 45*, 972-991.

[2] Liebscher, E. (2025). Selection of parametric copula models in the approximation of copulas using Cramér-von Mises divergence. In: A. Steland, E. Rafajlowicz, N. Parolya (Eds.): *Stochastic Models, Statistics and Their Applications, TU Delft 2024*. (pp. 19–31). Springer.