Proceedings of the 6th International Conference on Statistics: Theory and Applications (ICSTA'24)
Barcelona, Spain – August 19 - 21, 2024
Paper No. 148
DOI: 10.11159/icsta24.148

On the Estimation of Joint Density Functions via Copulas

Serge B. Provost [1], Yishan Zang [1]
[1]The University of Western Ontario
Department of Statistical and Actuarial Sciences, London, Canada
provost@stats.uwo.ca; yzang8@uwo.ca

Extended Abstract

Copulas are mainly used for modeling dependency features in multivariate distributions. In light of Sklar's theorem, they enable one to express the joint distribution of two or more random variables in terms of a specific correlation structure and the marginal distributions so that the effects of the dependence between the variables and the marginals can be separated. Copulas have found applications in various research fields, including reliability theory, machine learning, finance, geodesy, signal processing, medicine and hydrology.

Since marginal distributions can readily be determined by making use of an array of parametric, semi-parametric and nonparametric techniques, the principal challenge when it comes to estimating a joint density function via Skar's result, consists of estimating the distribution of the copula associated with the pseudo-observations. This will be achieved by making use of four copula density estimation methodologies.

First, bivariate least-squares approximations of successive degrees are obtained from the empirical copula values. The resulting polynomials are then differentiated with respect to both variables so as to obtain copula density estimates. Such estimates turn out to be similar over a wide range of degrees, which allows one to select a density estimate as a yardstick or reference density for comparison purposes with respect to those secured by making use of other methodologies.

Another technique for obtaining a copula density consists of differentiating Berstein polynomial approximations of Deheuvels' empirical copula. The degree of the polynomial approximation is determined by minimizing the integrated squared difference between polynomial approximations of various degrees and the density function previously chosen as a yardstick.

Modified kernel density estimates whose kernels are assumed to have a finite support are also proposed. Given a random sample of n bivariate observations, we advocate to place the pseudo-observations in the middle of the cells of the n x n grid of the unit square and to center the kernels at these points so as to mitigate boundary issues. Then, the selection of an appropriate bandwidth is based on the integrated squared differences between kernel density estimates of differing bandwidths and a reliable reference copula density estimate.

Lastly, a methodology that is based on the differentiation of linearized empirical copulas is proposed. It involves a spacing parameter that is selected in such a way that the resulting density estimate and a reference copula density exhibit similar distributional features. Additionally, a moment-based bivariate polynomial approximation is introduced to smooth out the resulting density function.

Although distinct in nature, the proposed approaches to the estimation of copula density functions were found to produce similar density estimates, which to a certain extent, validates their suitability. Joint density functions are finally secured by applying a corollary of Sklar's theorem, which incidentally provides more flexibility than any direct joint density estimation technique. Illustrative examples involving actual and simulated data will be presented.