Empirical and non-parametric copula models with the `cort` package

O. Laverny*

Résumé

The R package `cort` implements object-oriented classes and methods to estimate, simulate and visualize certain types of non-parametric copulas. The classic empirical checkerboard copula, and the empirical chekerboard copula with known multivariate margins are avalailables. The core of the package is the Cort algorithm, for Copula Recursive Trees. This algorithm perform a recursive local splitting of the unit hypercube, much like the CART regression algorithm, to fit a piecewise constant copula density. We also provide an efficient way of mixing copulas, allowing to bag the previous algorithms into forests. The algorithms implemented in this package and the corresponding implementation choices will be developped.
Mots-clefs : Dependence structure, copulas, estimation, non-parametric, checkerboard, cort, piecewise linear distribution functions.

Développement

The R package `cort` [4, 3] implements object-oriented classes and methods to estimate, simulate and visualize certain types of non-parametric copulas [8].

Copulas are distribution functions on the unit hypercube that have uniform margins. Although the estimation of copulas is a widely-treated subject, most performing estimators available in the literature are based on restricted, parametric estimation : vine copulas [6] and graphical models [5] for example are potential solutions but under restrictive assumptions. Classical density estimators such as kernels or wavelets do not satisfy marginal copula constraints. There also exist several tree-structured piece wise constant density estimators, but they do not always lead to proper copulas when applied on pseudo-observations or true copula samples. The new models that are implemented in this package try to solve these issues.

The Copula recursive tree, or Cort, is a flexible, consistent, piece-wise linear estimator for a copula, leveraging the patchwork copula formalization [2] and a specific piece-wise constant density estimator, the density estimation tree [7]. While the patchwork structure imposes the grid, this estimator is data-driven and constructs the grid recursively from the data, minimizing a chosen distance on the copula space. Furthermore, while the addition of the copula constraints makes the available solutions for density estimation unusable, our estimator is only concerned with dependence and guarantees the uniformity of margins. The R package `cort` [4, 3] provides a useful implementation of this model and several potential refinements, allowing for fast computations of Cort trees, and parallel computations of Cort forests.

Although the state of the art `copula` package has functions to estimate the empirical copula, we provide a structured set of S4 classes that allows estimation of empirical copulas, checkerboard copulas, Cort copula and bagging of all of these. A specific class exists for bagging Cort models, which implementation runs in parallel, to fasten the computations, using the `future`[1] package. Most of the underlying machinery and computations are written in `C++` for speed.

*Institut Camille Jordan UMR 5208, SCOR SE, `oskar.laverny@univ-lyon1.fr`

Références

[1] Henrik Bengtsson. A unifying framework for parallel and distributed processing in r using futures, aug 2020. URL `https://arxiv.org/abs/2008.00553`.

[2] Fabrizio Durante, Juan Fernández-Sánchez, José Juan Quesada-Molina, and Manuel Ubeda-Flores. Convergence results for patchwork copulas. *European Journal of Operational Research*, 247(2) : 525–531, 2015. doi : 10.1016/j.ejor.2015.06.028.

[3] Oskar Laverny. Empirical and non-parametric copula models with the 'cort' r package. *Journal of Open Source Software*, 5(56) :2653, 2020. doi : 10.21105/joss.02653. URL `https://doi.org/10.21105/joss.02653`.

[4] Oskar Laverny, Véronique Maume-Deschamps, Esterina Masiello, and Didier Rullière. Dependence structure estimation using copula recursive trees, 2020. URL `https://arxiv.org/abs/2005.02912`.

[5] Yinan Li, Xiao Liu, and Fang Liu. Panda : Adaptive noisy data augmentation for regularization of undirected graphical models, 2019. URL `https://arxiv.org/abs/1810.04851`.

[6] Thomas Nagler and Claudia Czado. Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. *Journal of Multivariate Analysis*, 151 :69–89, 2016. doi : 10.1016/j.jmva.2016.07.003.

[7] Parikshit Ram and Alexander G Gray. Density estimation trees. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 627–635, 2011. doi : 10.1145/2020408.2020507.

[8] Abe Sklar. Fonction de répartition dont les marges sont données. *Inst. Stat. Univ. Paris*, 8 : 229–231, 1959.