A copula estimation through recursive partitioning of the unit hypercube

O. Laverny [1,2]
E. Masiello [1], V. Maume-Deschamps [1] and D. Rullière [3]

[1] Institut Camille Jordan, UMR 5208, Université Claude Bernard Lyon 1, Lyon, France

[2] SCOR SE

[3] Mines Saint-Etienne, UMR CNRS 6158, LIMOS, Saint-etienne, France

June 23, 2022

Table of contents

Introduction and motivations

Introduction

A reinsurer such as SCOR observes many (positives) random variables that represent losses from several portfolios of risks.

A major topic is the evaluation and modeling of the dependency between those risks. Indeed, the properties of this dependency have a huge impact on the behavior of the total loss, and therefore on the capital management of the company.

The practical goal of this work was to produce relevant non-parametric estimators of dependence stuctures.

Copulas Basics

Suppose that $\boldsymbol{X}$ is a (continuous) random vector of dimension d with c.d.f F and marginals c.d.f $(F_i)_{i \in \{1,\dots,d\}}$. Then Sklar's theorem[1] gives us the copula of $\boldsymbol{X}$ as :

$$C(\boldsymbol{u}) = F\left(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\right)$$

(i) C is a c.d.f with uniform margins on $[0, 1]$.

(ii) It characterizes the dependence structure of F in the sense that F is completely characterized by C and the F_i's.

Few non-parametric estimators exists ! *But the estimation of copula is still a wide-treated subject: see, e.g., Deheuvel*[2] *and the* `copula` *R package.*

[1] A Sklar. "Fonctions de repartition an dimensions et leurs marges". In: *Publ. inst. statist. univ. Paris* 8 (1959), pp. 229–231.

[2] Paul Deheuvels. "La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance". In: *Bulletins de l'Académie Royale de Belgique* 65.1 (1979), pp. 274–292.

Density estimation trees

In regression, the CART algorithm from Breiman[3] selects a covariate and a univariate breakpoint, minimizing a loss, and assign to each leaf the mean response inside the leaf.

In density estimation, the DET from Ram and Gray[4] selects a dimension and a breakpoint minimizing a loss, and assign to each leaf the frequency of observations, producing histogram with varying bin sizes:

$$f(x) = \sum_{\ell \in \mathcal{L}} \frac{\mathrm{f}_\ell}{\lambda(\ell)} \mathbf{1}_{x \in \ell}$$

What loss can we use ? Will this yield a copula if applied to pseudo-observations ?

[3]Leo Breiman. “Out-of-Bag Estimation”. 1996.

[4]Parikshit Ram and Alexander G. Gray. “Density Estimation Trees”. In: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '11*. The 17th ACM SIGKDD International Conference. San Diego, California, USA: ACM Press, 2011, p. 627. ISBN: 978-1-4503-0813-7.

Piecewise linear copulas and the CORT estimator

Definition

Definition (Piecewise linear copula)

Let $\mathbb{I} = [0,1]^d$ be the unit hypercube and $\mathcal{L}$ a partition of $\mathbb{I}$. A Piecewise linear copula is defined by its distribution function:

$$\forall u \in \mathbb{I},\ C_{p,\mathcal{L}}(u) = \sum_{\ell \in \mathcal{L}} p_\ell \lambda_\ell(u)$$

(i) $\lambda_\ell(u) = \frac{\lambda([0,u]\cap\ell)}{\lambda(\ell)}$ where λ is the Lebesgue measure.

(ii) p is a vector of positives weights summing to one.

Corresponding density : $c_{p,\mathcal{L}}(u) = \sum_{\ell \in \mathcal{L}} \frac{p_\ell}{\lambda(\ell)} \mathbf{1}_{u\in\ell}$.

This is a simple form of patchwork copulas[5]

[5] Fabrizio Durante, Juan Fernández Sánchez, and Carlo Sempi. "Multivariate Patchwork Copulas: A Unified Approach with Applications to Partial Comonotonicity". In: *Insurance: Mathematics and Economics* 53.3 (Nov. 2013), pp. 897–905. ISSN: 01676687.

Knowing the partition, finding weights is a QP.

We restrict the leaves in $\mathcal{L}$ to be hyperboxes of the form $[a, b]$.

Property (Copula constraints are linear in the weights)

$$C_{p,\mathcal{L}} \text{ is a proper copula} \iff p \in \mathcal{C}_{\mathcal{L}} = \{p \in \mathbb{R}^{|\mathcal{L}|} : \; Bp = g \text{ and } p \geq 0\}.$$

We use the integrated square error between densities:

$$\|c_{p,\mathcal{L}} - c\|_2^2 \propto \|p\|_{\mathcal{L}}^2 - 2\langle p, \mathrm{f}\rangle_{\mathcal{L}}, \text{ where } \langle x, y\rangle_{\mathcal{L}} = \sum_{\ell \in \mathcal{L}} \frac{x_\ell y_\ell}{\lambda(\ell)}.$$

Definition (Quadratic program)

Knowing the partition $\mathcal{L}$, define p^* as the solution to the quadratic program:

$$\arg\min_{p \in \mathcal{C}_{\mathcal{L}}} \|p\|_{\mathcal{L}}^2 - 2\langle p, \mathrm{f}\rangle_{\mathcal{L}}.$$

Remark: projection of f onto $\mathcal{C}_{\mathcal{L}}$ regarding $\|.\|_{\mathcal{L}}^2$.

Joint optimisation fo the breakpoint and the weights

For a set of dimensions D in $\mathcal{P}(\{1, ..., d\})$, let $L(\ell, x, D)$ be the partition of the leaf ℓ splitted on a point x in dimensions D, i.e:

$$L((a,b], x, D) = \bigtimes_{j \in D} \{(a_j, x_j], (x_j, b_j]\} \bigtimes_{j \in \{1,...,d\} \setminus D} \{(a_j, b_j]\}\,.$$

Define then the full partition by:

$$\mathcal{L}_{x,D} = \mathcal{L} \setminus \{\ell(x)\} \cup L(\ell(x), x, D).$$

We will omit the parameter D if $D = \{1, .., d\}$.

Definition (Final optimisation problem)

The global optimisation problem we want to solve is :

$$\underset{\substack{D \in \mathcal{P}(\{1,...,d\}) \\ x \in \mathbb{I} \\ p \in \mathcal{C}_{\mathcal{L}_{x,D}}}}{\arg\min} \quad \|p\|^2_{\mathcal{L}_{x,D}} - 2 \langle p, \mathrm{f}_{\mathcal{L}_{\mathrm{x,D}}} \rangle_{\mathcal{L}_{x,D}}$$

The recursive procedure

(i) Solve the *density* problem to find the splitting point:

$$\underset{\substack{D \in \mathcal{P}(\{1,\dots,d\}) \\ x \in \mathbb{I}}}{\arg\min} \quad -\|\mathrm{f}_{\mathcal{L}_{\mathrm{x,D}}}\|^2_{\mathcal{L}_{x,D}}$$

(a) Find the splitting dimensions D first *using a test inspired by Bowman*[6].

(b) Minimize greedily on x via a non-linear programming solver (e.g., `nloptr`)

(ii) Recurse on each ℓ in $\mathcal{L}_{x,D}$ by rescaling ℓ to $\mathbb{I}$ and solving the same problem to obtain the final partition $\mathcal{L}$.

(iii) Then, with $\mathcal{L}$ fixed, solve the projection:

$$\underset{p \in \mathcal{C}_{\mathcal{L}}}{\arg\min} \; \|p\|^2_{\mathcal{L}} - 2\,\langle p, \mathrm{f}_{\mathcal{L}} \rangle_{\mathcal{L}}$$

via a quadratic programming solver (e.g., `osqp`), with initial values $\mathrm{f}_{\mathcal{L}}$.

[6] A.W. Bowman. "Density Based Tests for Goodness-of-Fit". In: *Journal of Statistical Computation and Simulation* 40.1-2 (Feb. 1992), pp. 1–13. ISSN: 0094-9655, 1563-5163.

Finding the splitting dimensions D (for $U \sim C$)

Hypothesis ($\mathcal{H}_j$)

$$(U_j \perp\!\!\!\perp U_{-j})\,|U \in \ell \text{ and } U_j|U \in \ell \sim \mathcal{U}(\ell_j)$$

Test statistic: a quadratic loss on densities, based on Bowman[7]. Suppose that $\ell = \mathbb{I}$, containing n obs. of the random vairable $U \sim F$, for F the restriction of C to ℓ, rescaled to $\mathbb{I}$. Then:

Definition (Test statistic)

Denote by $f^{(n)}_{\mathrm{f},\mathcal{L}}$ the piecewise constant density that will be estimated on data $U_1, ..., U_n \sim F$, and $e_{j,n}(x) = \mathbb{E}(f^{(n)}_{\mathrm{f},\mathcal{L}}(x)|\mathcal{H}_j)$. The test statistic is given by :

$$\mathcal{I}_j = \|e_{j,n} - f^{(n)}_{\mathrm{f},\mathcal{L}}\|_2^2$$

where $\mathcal{L}$, $e_{j,n}$ and $f^{(n)}_{\mathrm{f},\mathcal{L}}$ are stochastic objetcs, depending on U.

[7] A.W. Bowman. "Density Based Tests for Goodness-of-Fit". In: *Journal of Statistical Computation and Simulation* 40.1-2 (Feb. 1992), pp. 1–13. ISSN: 0094-9655, 1563-5163.

Test procedure

We weakened the test by assuming that the next split is enough to test $\mathcal{H}_j$. This gives a test procedure as follows:

(i) Solve

$$x^* = \underset{x \in \mathbb{I}}{\arg\min} \; -\|\mathrm{f}_{\mathcal{L}_\mathrm{x}}\|^2_{\mathcal{L}_x}$$

(ii) Compute:

$$\widehat{\mathcal{I}}_j = \sum_{k \in \mathcal{L}_{x^*,\{1,\dots,d\}\setminus\{j\}}} \left(\frac{\mathrm{f}_k^2}{\lambda(k)} + \sum_{\substack{\ell \in \mathcal{L}_{x^*,\{1,..,d\}} \\ \ell \subset k}} \left(\frac{\mathrm{f}_\ell^2}{\lambda(\ell)} - 2\frac{\mathrm{f}_k \mathrm{f}_\ell}{\lambda(k)} \right) \right).$$

Argument : the cut will be on the same x in dimensions other than j wheter or not we work under $\mathcal{H}_j$.

(iii) Compare to a Monte-carlo simulation of its distribution under the null to exclude the dimension j if necessary.

Asymptotic behavior

Previous result

Ram and Gray[8] gave the consistency of $f_{\mathrm{f},\mathcal{L}}^{(n)}$. Assuming the maximum diameter of leaves goes to 0 as n goes to ∞, we have :

$$\mathbb{P}\left(\lim_{n\mapsto+\infty}\|f_{\mathrm{f},\mathcal{L}}^{(n)} - f\|_2^2 = 0\right) = 1.$$

Denoting q s.t:

$$\forall \ell \in \mathcal{L},\ q_\ell = \int_\ell c(u)du,$$

this results writes $d_{\mathcal{L}}(\mathrm{f}, q)^2 \to 0$, a.s.

Furthermore, by construction, $q \in \mathcal{C}$.

[8] Parikshit Ram and Alexander G. Gray. "Density Estimation Trees". In: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '11*. The 17th ACM SIGKDD International Conference. San Diego, California, USA: ACM Press, 2011, p. 627. ISBN: 978-1-4503-0813-7.

Constraint influence

Definition (Integrated constraint influence)

$$\|c_{p,\mathcal{L}}^{(n)} - f_{\mathrm{f},\mathcal{L}}^{(n)}\|_2^2 = d_{\mathcal{L}}(p, \mathrm{f})^2$$

This quantity measures how much f and p are far from each other. But since f is closer and closer to q, which is in the set that f is projected on to give p, we have :

Property (Asymptotical effect of constraints)

The integrated constraint influence is asymptotically 0.

Proof.

$\mathcal{C}$ *is convex, closed and non-empty. Hence* $p = \mathcal{P}_{\mathcal{C}}(\mathrm{f})$ *exist and is unique. Since* $q \in \mathcal{C}$, *we have that* $d_{\mathcal{L}}(\mathrm{f}, p)^2 \leq d_{\mathcal{L}}(f, q)^2$. □

Cort's consistency

Property (Consistency)

For c the density of the true copula, assuming the diameter of the leaves goes to 0 as n goes to ∞, the estimator $c_{p,\mathcal{L}}^{(n)}$ is consistent, i.e :

$$\mathbb{P}\left(\lim_{n\to+\infty}\|c_{p,\mathcal{L}}^{(n)} - c\|_2^2 = 0\right) = 1$$

The proof leverages the true weights: $\forall \ell \in \mathcal{L},\ q_\ell = \int_\ell c(u)du$, and compares them to p and f by noting that $q \in \mathcal{C}_{\mathcal{L}}$.

Bagging and cross-validation

A simple forest

Regression: random forests from Breiman[9] use an out-of-bag procedure.

Density estimation: kernels uses *leave-one-out* for bandwidths. See Sain & al[10] and Wu[11] for cross-validation in density estimation. We develop an *out-of-bag* procedure.

Definition (Out-of-bag "density" and metrics)

$$c_{oob}(u) = \frac{1}{N(u)} \sum_{j=1}^{N} c^{(j)}(u) \mathbf{1}_{\text{u was not in the training set of } c^{(j)}}$$

$$J_{oob}(c_N) = \|c_N\|_2^2 - \frac{2}{n} \sum_{i=1}^{n} c_{oob}(u_i) \text{ and } KL_{oob}(c_N) = -\frac{1}{n} \sum_{i=1}^{n} ln(c_{oob}(x_i))$$

We also proposed a weighed version of the forest.

[9] Leo Breiman. "Out-of-Bag Estimation". 1996.

[10] Stephan R Sain, Keith A Baggerly, and David W Scott. "Cross-validation of multivariate densities". In: *Journal of the American Statistical Association* 89.427 (1994), pp. 807–817.

[11] Kaiyuan Wu, Wei Hou, and Hongbo Yang. "Density estimation via the random forest method". In: *Communications in Statistics-Theory and Methods* 47.4 (2018), pp. 877–889.

Simulation Study

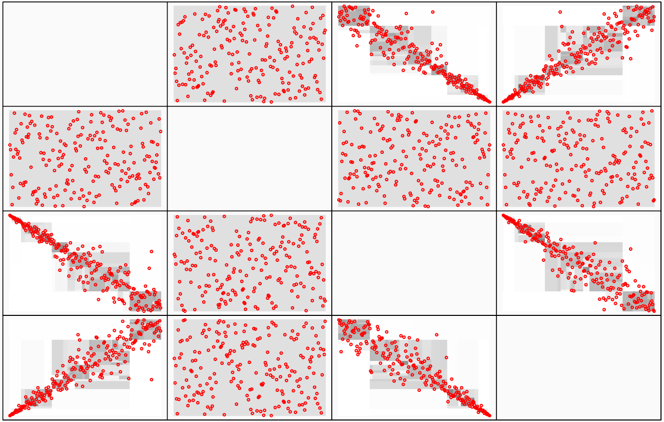

Figure 1: In gray scale, we observe a bivariate histogram corresponding to the estimated tree. Red points are input data, a simulation of 200 points from a 3-variate Clayton copula with $\theta = 8$, the second marginal is uniform and the fourth is flipped.

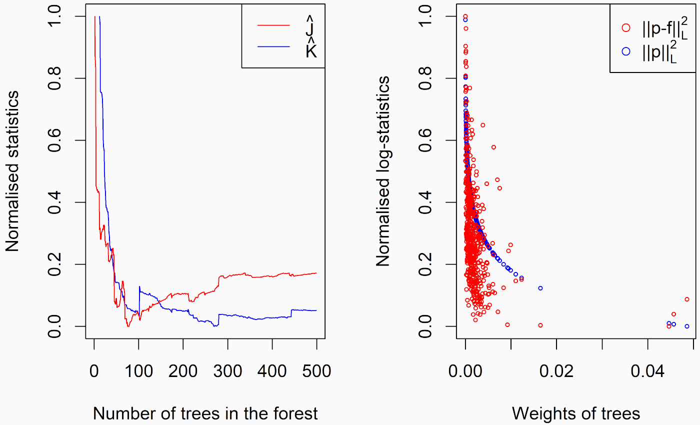

Figure 2: Left : Out-of-bag Kullback-Leibler and ISE as a function of the number of trees. Right: Constraint influence and L2-Norm in function of weights

Table 1: Bagging results on several out-of-bag performance metrics.

	Empirical	Cb(m=10)	Cb(m=5)	Beta	Cort
$\hat{J}$	0.00501	-40.6	-17.6	-186	**-1387**
$\hat{K}$	Inf	Inf	Inf	-0.544	**-5.15**
$\hat{M}$	2.89e-05	**2.53e-05**	5.25e-05	3.43e-05	7.06e-05
$\hat{N}$	-0.000662	**-0.000667**	-0.000639	-0.000655	-0.000634

$\hat{M}$ and $\hat{N}$ are other out-of-bag metrics based on the c.d.f. rather than the p.d.f.

Note: Checkerboards are also piecewise linear copulas, with fixed regular partition.

Take away points

Take away and potential improvements

Some take away points:

(i) The copula property of a copula estimator is a prerequisite for further analysis (sampling, etc.)

(ii) Piecewise linear distribution function are handy models for copula modeling because the copula constraints are linear in the weights.

(iii) Fitting piecewise linear distribution functions with trees is quite simple and fast

(iv) Such models can easily be bagged, boosted, cross-validated. . .

(v) The main issue is the degree of freedom in weights took away by the copula constraint.

Detailed code and examples are available in our R package[12] and paper[13].

[12] OL. "Empirical and Non-Parametric Copula Models with the Cort R Package". In: *Journal of Open Source Software* 5.56 (2020), p. 2653.

[13] OL, Esterina Masiello, Véronique Maume-Deschamps, and Didier Rullière. "Dependence Structure Estimation Using Copula Recursive Trees". In: *Journal of Multivariate Analysis* 185 (2021), p. 104776.