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>
<thead>
<tr>
<th>1.</th>
<th>Introduction and motivations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>Piecewise linear copulas and the CORT estimator</td>
</tr>
<tr>
<td>3.</td>
<td>Asymptotic behavior</td>
</tr>
<tr>
<td>4.</td>
<td>Bagging and cross-validation</td>
</tr>
<tr>
<td>5.</td>
<td>Simulation Study</td>
</tr>
<tr>
<td>6.</td>
<td>Take away points</td>
</tr>
</tbody>
</table>
Introduction and motivations
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 structures.
Suppose that $X$ is a (continuous) random vector of dimension $d$ with c.d.f $F$ and marginals c.d.f $(F_i)_{i \in \{1, \ldots, d\}}$. Then Sklar’s theorem$^1$ gives us the copula of $X$ as:

$$C(u) = F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))$$

(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.


Density estimation trees

In regression, the CART algorithm from Breiman\textsuperscript{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\textsuperscript{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:

\[ f(x) = \sum_{\ell \in \mathcal{L}} \frac{f_\ell}{\lambda(\ell)} 1_{x \in \ell} \]

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


Piecewise linear copulas and the CORT estimator
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 $\lambda$ 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\(^5\)

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)**

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

We use the integrated square error between densities:

$$\| c_{p,\mathcal{L}} - c \|_2^2 \propto \| p \|_{\mathcal{L}}^2 - 2 \langle p, f \rangle_{\mathcal{L}},$$

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:

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

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

7/19
Joint optimisation for the breakpoint and the weights

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

$$L((a, b], x, D) = \prod_{j \in D} \left\{(a_j, x_j], (x_j, b_j]\right\} \times \prod_{j \in \{1, \ldots, d\} \backslash D} \left\{(a_j, b_j]\right\}.$$

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 $D = \{1, \ldots, d\}$.

**Definition (Final optimisation problem)**

The global optimisation problem we want to solve is:

$$\arg \min_{D \in \mathcal{P}(\{1, \ldots, d\})} \min_{x \in \mathbb{I}} \min_{p \in \mathcal{C}_{\mathcal{L}_x, D}} \|p\|_{\mathcal{L}_x, D}^2 - 2 \langle p, f_{\mathcal{L}_x, D} \rangle_{\mathcal{L}_x, D}.$$
The recursive procedure

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

\[
\arg \min_{D \in P(\{1, \ldots, d\})} \min_{x \in \mathbb{I}} -\|f_{\mathcal{L}_x, D}\|_{\mathcal{L}_x, D}^2
\]

(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., \textit{nloptr})

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

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

\[
\arg \min_{p \in \mathcal{C}_{\mathcal{L}}} \|p\|_{\mathcal{L}}^2 - 2 \langle p, f_{\mathcal{L}} \rangle_{\mathcal{L}}
\]

via a quadratic programming solver (e.g., \textit{osqp}), with initial values \( f_{\mathcal{L}} \).

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

**Hypothesis ($H_j$)**

\[(U_j \perp \perp U_{-j}) \mid U \in \ell \text{ and } U_j \mid U \in \ell \sim U(\ell_j)\]

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

**Definition (Test statistic)**

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

\[I_j = \|e_{j,n} - f_{f,\mathcal{L}}^{(n)}\|_2^2\]

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

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

$$x^* = \arg \min_{x \in I} -\| f_{\mathcal{L}_x} \|^2_{\mathcal{L}_x}$$

(ii) Compute:

$$\hat{I}_j = \sum_{k \in \mathcal{L}_{x^*}, \{1, \ldots, d\} \setminus \{j\}} \left( \frac{f_k^2}{\lambda(k)} + \sum_{\ell \in \mathcal{L}_{x^*}, \{1, \ldots, d\}, \ell \subset k} \left( \frac{f_\ell^2}{\lambda(\ell)} - 2 \frac{f_k f_\ell}{\lambda(k)} \right) \right).$$

Argument: the cut will be on the same $x$ in dimensions other than $j$ whether 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
Ram and Gray\textsuperscript{8} gave the consistency of $f_{f,L}^{(n)}$. Assuming the maximum diameter of leaves goes to 0 as $n$ goes to $\infty$, we have:

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

Denoting $q$ s.t:

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

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

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

Definition (Integrated constraint influence)

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

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 \( p = \mathcal{P}_\mathcal{C}(f) \) exist and is unique. Since \( q \in \mathcal{C} \), we have that \( d_\mathcal{L}(f, p)^2 \leq d_\mathcal{L}(f, q)^2 \). \( \square \)
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 \( \infty \), the estimator \( c_{p,\mathcal{L}}^{(n)} \) is consistent, i.e:

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

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

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

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

*We also proposed a weighed version of the forest.*


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.

<table>
<thead>
<tr>
<th></th>
<th>Empirical</th>
<th>Cb(m=10)</th>
<th>Cb(m=5)</th>
<th>Beta</th>
<th>Cort</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{J})</td>
<td>0.00501</td>
<td>-40.6</td>
<td>-17.6</td>
<td>-186</td>
<td>-1387</td>
</tr>
<tr>
<td>(\hat{K})</td>
<td>Inf</td>
<td>Inf</td>
<td>Inf</td>
<td>-0.544</td>
<td>-5.15</td>
</tr>
<tr>
<td>(\hat{M})</td>
<td>2.89e-05</td>
<td>2.53e-05</td>
<td>5.25e-05</td>
<td>3.43e-05</td>
<td>7.06e-05</td>
</tr>
<tr>
<td>(\hat{N})</td>
<td>-0.000662</td>
<td>-0.000667</td>
<td>-0.000639</td>
<td>-0.000655</td>
<td>-0.000634</td>
</tr>
</tbody>
</table>

\(\hat{M}\) 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\textsuperscript{12} and paper\textsuperscript{13}.
