Construction and estimation of high dimensional copulas

Gildas Mazo
PhD work supervised by S. Girard and F. Forbes
Mistis, Inria and laboratoire Jean Kuntzmann, Grenoble, France

Séminaire Statistiques, Strasbourg, 14 octobre 2014
Copulas: why and what?

Multivariate vector \((X_1, \ldots, X_d) \sim F\)

- \(F\) is often nongaussian / tail dependent

Example

- flood \(\{X_1 > q, \ldots, X_d > q\}\)
- portfolio of financial assets \(X_1 + \cdots + X_d\)

Two steps modeling of \(F\):

1. Margins \(F_1, \ldots, F_d\)
2. Dependence

Copula:

- d.f. of \((F_1(X_1), \ldots, F_d(X_d))\)
- takes care of dependence
Challenge: modeling in high dimension with copulas

Two issues:

- **Construction of high-dimensional copulas:**
  difficult to describe the interplay between many variables

- **Estimation of copulas:**
  we have the sample of $F$

  $$
  (X_1^{(k)}, \ldots, X_d^{(k)}), \ k = 1, \ldots, n
  $$

  not that of $C$

  $$
  (F_1(X_1^{(k)}), \ldots, F_d(X_d^{(k)})), \ k = 1, \ldots, n
  $$
A short review of copulas
  Definition and main properties/ideas
  High dimensional copulas

Two novel high-dimensional copulas
  PBC copulas
  FDG copulas

Estimation
  PBC copulas
  FDG copulas

Numerical results
Definition and Sklar’s theorem

Definition (\(d\)-dimensional copula)
\(d.f.\) of \((U_1, \ldots, U_d), U_i \sim U([0, 1])\)

\[ C(u_1, \ldots, u_d) = P[U_1 \leq u_1, \ldots, U_d \leq u_d] \]

Theorem (Sklar, 1959)

\(\text{If } C \text{ copula and } F_1, \ldots, F_d \text{ are univariate } d.f. \text{ then} \)

\[ F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)) \]  \hspace{1cm} (1)

\(\text{defines a multivariate } d.f.\)

\(\text{Every } F \sim (X_1, \ldots, X_d) \text{ can be written as } (1)\)

\(\text{C is the } d.f. \text{ of } (F_1(X_1), \ldots, F_d(X_d))\)
Quantifying dependence

\((X_1^{(1)}, X_2^{(1)}), (X_1^{(2)}, X_2^{(2)})\) i.i.d. \(\sim (X_1, X_2)\) with margins \(F_1, F_2\)

Kendall’s tau:

\[
\tau = P \left[ (X_1^{(1)} - X_1^{(2)})(X_2^{(1)} - X_2^{(2)}) > 0 \right] - P \left[ (X_1^{(1)} - X_1^{(2)})(X_2^{(1)} - X_2^{(2)}) < 0 \right]
\]

Spearman’s rho:

\[
\rho = \text{cor}(F_1(X_1), F_2(X_2))
\]

Property:

\[
\tau = 4 \int_{[0,1]^2} CdC - 1, \quad \rho = 12 \int_{[0,1]^2} Cd\Pi - 3
\]
Tail dependence coefficients

Upper tail dependence coefficient:
\[
\lambda^{(U)} = \lim_{u \uparrow 1} P \left[ F_2(X_2) > u \mid F_1(X_1) > u \right]
\]

Lower tail dependence coefficient:
\[
\lambda^{(L)} = \lim_{u \downarrow 0} P \left[ F_2(X_2) \leq u \mid F_1(X_1) \leq u \right]
\]

Property:
\[
\lambda^{(L)} = \lim_{u \downarrow 0} \frac{C(u, u)}{u}, \quad \lambda^{(U)} = \lim_{u \uparrow 1} \frac{1 - 2u + C(u, u)}{1 - u}
\]
A short review of copulas
  Definition and main properties/ideas
  High dimensional copulas

Two novel high-dimensional copulas
  PBC copulas
  FDG copulas

Estimation
  PBC copulas
  FDG copulas

Numerical results
Important copulas in high dimension

- Tail asymmetry
- Flexibility
- Tractability

Diagram showing the relationships between tail asymmetry, flexibility, and tractability.
Important copulas in high dimension

Archimedean

flexibility

tractability
tail asymmetry
Important copulas in high dimension

Archimedean

flexibility

tractability

tail asymmetry

Vines
Important copulas in high dimension

- Archimedean
- Elliptical
- Vines

- Tractability
- Tail asymmetry
- Flexibility
- Tail asymmetry

Diagram showing the relationships between different copulas with attributes such as tractability and tail asymmetry.
Archimedean copulas
[Nelsen 1996], [Joe 2001], [McNeil Neslehova 2009]

\[
C(u_1, \ldots, u_d) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d))
\]

\(\psi\) univariate function, called the generator of \(C\)

Kendall’s tau:

\[
\tau = 1 + 4 \int_0^1 \frac{\psi^{-1}(t)}{(\psi^{-1})'(t)} dt
\]

Examples:

\[
\psi^{-1}(t) = (-\log t)^{\theta}, \quad \tau = (\theta - 1)/\theta, \quad \theta \geq 1 \text{ (Gumbel)}
\]

\[
\psi^{-1}(t) = (t^{-\theta} - 1)/\theta, \quad \tau = \theta/(\theta + 2), \quad \theta > 0 \text{ (Clayton)}
\]

Not flexible because exchangeable! \(\tau_{12} = \tau_{13} = \tau_{37} = \ldots\)
Vines

[Joe 2001], [Bedford Cook 2002], [Aas et al 2009]

Decomposition of the density into (conditional) bivariate densities

\[ c(u_1, \ldots, u_d) = \]

Drawable vines:

\[
\prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j|i+1,\ldots,i+j-1} \{ F(u_i|u_{i+1}, \ldots, u_{i+j-1}), F(u_{i+j}|u_{i+1}, \ldots, u_{i+j-1}) \}
\]

Canonical vines:

\[
\prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,j+1|1,\ldots,j-1} \{ F(u_j|u_1, \ldots, u_{j-1}), F(u_{j+i}|u_1, \ldots, u_{j-1}) \}
\]

- For each of above, 60 decompositions with \( d = 5 \), >1 million with \( d = 10 \), etc...
- Overfitting issues

Flexible, but not tractable
Elliptical copulas

[e.g., Frahm et al 2003]

Are the copulas of elliptical distributions: density $f$ writes

$$f(x) = |P|^{-1/2} g(x^T P^{-1} x),$$

$x \in \mathbb{R}^d$,

g density generator,

$P = (\theta_{ij})$ correlation matrix.

Examples:

$$g(t) \propto \exp(-t/2), \ t \geq 0, \ (\text{Gaussian})$$

$$g(t) \propto \left(1 + \frac{t}{\nu}\right)^{-(\nu+d)/2}, \ t \geq 0, \ \nu > 2, \ (\text{Student})$$

Kendall’s tau:

$$\tau_{ij} = \frac{2}{\pi} \arcsin(\theta_{ij})$$

Elliptical copulas are tail symmetric:

$$\lambda_{ij}^{(U)} = \lambda_{ij}^{(L)}$$
A short review of copulas
Definition and main properties/ideas
High dimensional copulas
Two novel high-dimensional copulas
   PBC copulas
   FDG copulas
Estimation
   PBC copulas
   FDG copulas
Numerical results

▶ Existing copulas have drawbacks
▶ Wish to construct copulas with all of
   ▶ tractability
   ▶ flexibility
   ▶ tail asymmetry
A short review of copulas
Definition and main properties/ideas
High dimensional copulas
Two novel high-dimensional copulas
   PBC copulas
   FDG copulas
Estimation
   PBC copulas
   FDG copulas
Numerical results

Product of Bivariate Copulas: build high-dimensional copulas by making products of bivariate copulas
Generalized product of copulas (Liebscher) is unhandy

\[ C(u_1, \ldots, u_d) = \prod_e \tilde{C}_e (g_{e1}(u_1), \ldots, g_{ed}(u_d)) \]

constraints: \[ \prod_e g_{ei}(u) = u, \quad i = 1, \ldots, d, \quad u \in [0, 1] \]

Almost not been used in practice (because of the constraints?)
Can we get a tractable/flexible model out of Liebscher?
Product of Bivariate Copulas (PBC)

Let $(\{U_1, \ldots, U_d\}, E)$ be a graph

**Assumption**

Set $g_{ei} = 1$ if $e \notin E$, does not depend on $e$ otherwise

**Theorem**

Liebscher reduces to PBC

$$C(u_1, \ldots, u_d) = \prod_{\{ij\} \in E} \tilde{C}_{ij}(u_i^{1/n_i}, u_j^{1/n_j}),$$

($n_i$ number of neighbors of the $i$-th variable)

No constraints on the parameters anymore!
Example

\[
C(u_1, u_2, u_3, u_4, u_5) = \bar{C}_{12}(u_1, u_2^{1/3}) \bar{C}_{23}(u_2^{1/3}, u_3^{1/2}) \bar{C}_{24}(u_2^{1/3}, u_4) \bar{C}_{35}(u_3^{1/2}, u_5)
\]
Properties and Limitations

- **Properties**
  - One parameter per edge in the graph
  - Benefits from the rich literature of bivariate copulas

- **Limitations**
  - Marginal independence:
    if \( \{ij\} \) does not belong to \( E \), then \( \tilde{C}_{ij} \equiv \Pi \)
  - Dependence weakens as graph gets more connected:
    if \( n_i \to \infty \), then \( \tilde{C}_{ik} \to \Pi \)

<table>
<thead>
<tr>
<th>coefficient</th>
<th>( \rho_{kl} )</th>
<th>( \tau_{kl} )</th>
<th>( \lambda_{kl} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>((n_k, n_l))</td>
<td>([-0.60, 0.60])</td>
<td>([-0.50, 0.50])</td>
<td>([0.00, 0.50])</td>
</tr>
<tr>
<td>(1,2)</td>
<td>([-0.30, 0.43])</td>
<td>([-0.21, 0.33])</td>
<td>([0.00, 0.50])</td>
</tr>
<tr>
<td>(2,2)</td>
<td>([-0.43, 0.43])</td>
<td>([-0.33, 0.33])</td>
<td>([0.00, 0.33])</td>
</tr>
<tr>
<td>(1,3)</td>
<td>([-0.19, 0.33])</td>
<td>([-0.13, 0.25])</td>
<td>([0.00, 0.33])</td>
</tr>
<tr>
<td>(2,3)</td>
<td>([-0.12, 0.27])</td>
<td>([-0.08, 0.20])</td>
<td>([0.00, 0.33])</td>
</tr>
<tr>
<td>(3,3)</td>
<td>([-0.43, 0.43])</td>
<td>([-0.33, 0.33])</td>
<td>([0.00, 0.33])</td>
</tr>
</tbody>
</table>
Triangle position

Can we do better?

if the data do not exhibit too much dependence:
A short review of copulas
Definition and main properties/ideas
High dimensional copulas

Two novel high-dimensional copulas
  PBC copulas
  FDG copulas

Estimation
  PBC copulas
  FDG copulas

Numerical results

one-Factor copulas with Durante Generators:
  - Factor property for tractability
  - “Durante class” for flexibility
Conditional independence property

[Krupskii Joe 2013]

\[ U_1, \ldots, U_d \text{ independent given a latent factor } U_0 \]

\[ C(u_1, \ldots, u_d) = \int_0^1 \prod_{i=1}^d \frac{\partial C_{0i}(u_0, u_i)}{\partial u_0} du_0 \]

Linking copulas:

\[(U_0, U_i) \sim C_{0i}\]

Can we calculate the integral for suitable linking copulas? (without removing flexibility)
“Durante” class of bivariate copulas

[Durante 2006]

(Nonparametric) class $\mathcal{D}_f$:

$$C_f(u, v) = \min(u, v)f(\max(u, v)),$$

$f$ generator
Coefficients:

$$\rho = 12 \int_0^1 x^2 f(x) dx - 3, \quad \tau = 4 \int_0^1 xf(x)^2 dx - 1,$$

$$\lambda^{(L)} = f(0), \text{ and } \lambda^{(U)} = 1 - f'(1)$$

Idea:

$$C_{0i} \equiv C_{fi} \in \mathcal{D}_{fi}$$
Expression of FDG copulas

- We can calculate the integral:

\[
C(u_1, \ldots, u_d) = 
\left[ \left( \prod_{j=2}^{d} u_{(j)} \right) \int_{u(d)}^{1} \prod_{j=1}^{d} f'_j(x) dx + f(1)(u(2)) \left( \prod_{j=2}^{d} f_j(u(j)) \right) \right. \\
+ \sum_{k=3}^{d} \left( \prod_{j=2}^{k-1} u_{(j)} \right) \left( \prod_{j=k}^{d} f_j(u(j)) \right) \int_{u(k-1)}^{u(k)} \prod_{j=1}^{k-1} f_j'(x) dx \bigg],
\]

\[u(1) := u_{\sigma(1)}, \quad f(1) := f_{\sigma(1)},\]

\[\sigma \text{ permutation of } (1, \ldots, d) \text{ such that } u_{\sigma(1)} \leq \cdots \leq u_{\sigma(d)}\]

- Bivariate margins \(C_{ij} \in D f_{ij}\) (“Durante”) with

\[f_{ij}(t) = f_i(t)f_j(t) + t \int_{t}^{1} f'_i(x)f'_j(x) dx\]
A remarkable tail dependence structure

Both in the upper and lower tails:

\[ \lambda_{ij}^{(\cdot)} = \lambda_i^{(\cdot)} \lambda_j^{(\cdot)} \]

Examples:

<table>
<thead>
<tr>
<th>Type</th>
<th>generator ( f_i(t) )</th>
<th>( \theta_i )</th>
<th>( \lambda_{ij}^{(L)} )</th>
<th>( \lambda_{ij}^{(U)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>((1 - \theta_i)t + \theta_i)</td>
<td>([0, 1])</td>
<td>(\theta_i\theta_j)</td>
<td>(\theta_i\theta_j)</td>
</tr>
<tr>
<td>CA</td>
<td>(t^{1-\theta_i})</td>
<td>([0, 1])</td>
<td>0</td>
<td>(\theta_i\theta_j)</td>
</tr>
<tr>
<td>S</td>
<td>(\frac{\sin(\theta_i t)}{\sin(\theta_i)})</td>
<td>((0, \frac{\pi}{2}])</td>
<td>0</td>
<td>((1 - \frac{\theta_i}{\tan \theta_i})(1 - \frac{\theta_j}{\tan \theta_j}))</td>
</tr>
<tr>
<td>E</td>
<td>(\exp\left(\frac{t^{\theta_i} - 1}{\theta_i}\right))</td>
<td>((0, \infty))</td>
<td>(\exp\left(-\frac{1}{\theta_i} - \frac{1}{\theta_j}\right))</td>
<td>0</td>
</tr>
</tbody>
</table>

F: [Fréchet 1958], CA: [Cuadras-Augé 1981]
S: sinus, E: exponential [Durante 2006]
We can calculate the extreme-value copulas of FDG

\[
\text{EVFDG}(u_1, \ldots, u_d) = \lim_{n \to \infty} \text{FDG}^n(u_1^{1/n}, \ldots, u_d^{1/n}) = \prod_{i=1}^{d} u_{(i)}^{\chi_i},
\]

with

\[
\chi_i = \left( \prod_{j=1}^{i-1} (1 - \lambda(j)) \right) \lambda(i) + 1 - \lambda(i) - \lambda_i = 1 - f_i'(1)
\]

Bivariate margins are Cuadras-Augé copulas [Cuadras-Augé 1981]:

\[
\text{EVFDG}_{#,ij}(u_i, u_j) = \min(u_i, u_j) \max(u_i, u_j)^{1-\lambda_i \lambda_j}
\]
FDG copulas seem to have a good triangle position
A short review of copulas
  Definition and main properties/ideas
  High dimensional copulas

Two novel high-dimensional copulas
  PBC copulas
  FDG copulas

Estimation
  PBC copulas
  FDG copulas

Numerical results
Two main methods for estimating copulas

**Setup** \((X_1, \ldots, X_d) \sim F\) with copula \(C \equiv C(\theta), \theta \in \Theta \subset \mathbb{R}^q\) for some integer \(q\).

We observe \((X_1^{(k)}, \ldots, X_d^{(k)})\) for \(k = 1, \ldots, n\).

**Methods**

- **Likelihood maximization**

  \[
  \max_{\theta} \sum_{k=1}^{n} \log c(\tilde{F}_1(X_1^{(k)}), \ldots, \tilde{F}_d(X_d^{(k)}); \theta)
  \]

  where \(\tilde{F}_i\) some approximation of \(F_i\)

  [Joe 2001], [Genest et al 1995]

- **Method-of-moment, based on dependence coefficients**

  [Favre Genest 2007], [Kluppelberg Kuhn 2009], [Oh Patton 2013]
A short review of copulas
  Definition and main properties/ideas
  High dimensional copulas

Two novel high-dimensional copulas
  PBC copulas
  FDG copulas

Estimation
  PBC copulas
  FDG copulas

Numerical results
Estimating PBC copulas by maximizing the likelihood

We want to maximize

\[ \log L(\theta) = \sum_{k=1}^{n} \log c(\tilde{F}_1(X_1^{(k)}), \ldots, \tilde{F}_d(X_d^{(k)}); \theta) \]

where the density is

\[ c(u_1, \ldots, u_d) = \frac{\partial^d C(u_1, \ldots, u_d)}{\partial u_1 \ldots \partial u_d} \]

and

\[ C(u_1, \ldots, u_d) = \prod_{\{ij\} \in E} \tilde{C}_{ij}(u_1^{1/n_i}, u_1^{1/n_j}). \]

Numerical derivation unhandy in high-dimension

How to compute the density?
A message-passing algorithm to compute the density

[Huang Jojic 2010]

\[ u = (u_1, \ldots, u_d), \]
\[ \tau^i_e \text{ subtree rooted at variable } i \text{ and containing edge } e, \]
\[ \tau^j_e \text{ subtree rooted at edge } e \text{ and containing variable } i. \]
Write \( C(u) = \prod_{e \in N(i)} T_{\tau^i_e}(u). \)

\[
\begin{align*}
\partial_{u_V} C(u) &= \partial_{u_i, u_{V \setminus i}} \left[ \prod_{e \in N(i)} T_{\tau^i_e}(u) \right] \\
&= \partial_{u_i} \left[ \prod_{e \in N(i)} \frac{\partial_{u_{\tau^i_e \setminus i}} T_{\tau^i_e}(u)}{\mu_{e \to i}(u)} \right]
\end{align*}
\]
A message-passing algorithm to compute the density

$\mu_{e \rightarrow i}(u) = \partial u_{\tau_e \setminus i} T_{\tau_e i}(u)$

$= \partial u_j \left[ \tilde{C}_e(u_{i}^{1/n_i}, u_{j}^{1/n_j}) \partial u_{\tau_j e \setminus j} T_{\tau_j e}(u) \right]$

Recursive algorithm
A short review of copulas
  Definition and main properties/ideas
  High dimensional copulas

Two novel high-dimensional copulas
  PBC copulas
  FDG copulas

Estimation
  PBC copulas
  FDG copulas

Numerical results
Issue: FDG copulas are not differentiable

Recall bivariate FDG:

\[ C_{ij}(u, v) = \min(u, v) f_{ij}(uv / \min(u, v)) \]

- No likelihood
- No density
- No partial derivatives

Attempt to calculate first partial derivative (Fréchet):

\[
\lim_{h \to 0} \frac{C_{ij}(u + h, u) - C_{ij}(u, u)}{h} = \begin{cases} 
(1 - \theta_i \theta_j)u + \theta_i \theta_j & \text{if } h < 0, \\
(1 - \theta_i \theta_j)u & \text{if } h > 0.
\end{cases}
\]
Let’s try method-of-moment instead

Match empirical with parametric dependence coefficients

Definition (Weighted Least-Squares estimator (WLS) based on dependence coefficients)

\[
\hat{\theta} = \arg \min_{\theta \in \Theta} (\hat{r} - r(\theta))^T \hat{W} (\hat{r} - r(\theta)),
\]

where

- \( r(\theta) = (r_{1,2}(\theta), \ldots, r_{d-1,d}(\theta)) \) vector of pairwise dependence coefficients
- \( \hat{r} = (\hat{r}_{1,2}, \ldots, \hat{r}_{d-1,d}) \) empirical counterpart
- \( \hat{W} \) is a weight matrix
No asymptotic properties are derived in general case

Particular cases:
- Single, real parameter [Favre Genest 2007]
- Elliptical copulas [Kluppelberg Kuhn 2009]
- “Smooth copulas” (i.e., with partial derivatives) [Oh Patton 2013]

None of these references are suitable for FDG copulas
How to derive asymptotic properties for nonsmooth copulas?

Use [Hoeffding 1948]

\[ \sqrt{n}(\hat{r} - r(\theta_0)) \rightarrow N(0, \Sigma) \]

and pass convergence to \( \sqrt{n}(\hat{\theta} - \theta_0) \) by homeomorphism condition:

**Assumption (main)**

\( \Theta \rightarrow r(\Theta), \theta \mapsto r(\theta) \) is one-to-one, continuous, and its inverse is continuous as well

**Theorem**

- \( \sqrt{n}(\hat{\theta} - \theta_0) \rightarrow N(0, \Xi) \) (even though \( C \) has no partial derivatives)
- \( r_{i,j}(\theta) = \lambda_{i,j}^{(U)}(\theta) = \lambda_i^{(U)}(\theta)\lambda_j^{(U)}(\theta) \) verifies the Assumption for extreme-value copulas associated to FDG
A short review of copulas
Definition and main properties/ideas
High dimensional copulas
Two novel high-dimensional copulas
PBC copulas
FDG copulas
Estimation
PBC copulas
FDG copulas
Numerical results

- Assess gain of efficiency for message-passing in PBC copulas estimation
- Check WLS estimator’s performance for FDG copulas
- Analysis of extreme hydrological events
PBC: is maximizing the full likelihood worth it?

- 100 datasets of dimension $d = 9$ and size $n = 500$
- Monitoring coefficients

\[
GE = 1 - \frac{\sum_{e=1}^{d-1} \text{Var}\left(\hat{\theta}_e^{FULL}\right)}{\sum_{e=1}^{d-1} \text{Var}\left(\hat{\theta}_e^{PW}\right)}
\]

\[
\text{MAE}_\rho = \frac{1}{d-1} \sum_{e=1}^{d-1} |\rho(\theta_e) - \rho(\hat{\theta}_e^{FULL})|
\]

\[
\text{MAE}_\tau = \frac{1}{d-1} \sum_{e=1}^{d-1} |\tau(\theta_e) - \tau(\hat{\theta}_e^{FULL})|
\]
It depends on the families

<table>
<thead>
<tr>
<th>Copula</th>
<th>GE</th>
<th>MAE$_\rho$</th>
<th>MAE$_\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBC AMH</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>PBC FGM</td>
<td>0.02</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>PBC Frank</td>
<td>0.21</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>PBC Gumbel</td>
<td>0.32</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>PBC Joe</td>
<td>0.29</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

AMH: Ali-Mikhail-Haq
FGM: Farlie-Gumbel-Morgenstern
FDG: how behaves WLS estimator on simulated data?

- 200 datasets of dimension \( d = 4, 50 \) and size \( n = 500 \) from FDG-CA, FDG-F, FDG-S, FDG-E

- Estimation by WLS estimator with Spearman’s rho

For each dataset and each model:

\[
\text{MAE}_r = \frac{1}{p} \sum_{i<j} |r_{i,j} - r(\hat{\theta}_i, \hat{\theta}_j)| \quad \text{and} \quad \text{RMAE} = \frac{1}{d} \sum_{i=1}^d \frac{|\hat{\theta}_i - \theta_{0i}|}{\theta_{0i}},
\]

(averaged over the replications)
Performs well, even on high-dimensional data

<table>
<thead>
<tr>
<th>Copula</th>
<th>MAE$_r$</th>
<th>RMAE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$d = 4$</td>
<td>$d = 50$</td>
</tr>
<tr>
<td>FDG-CA</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>FDG-F</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>FDG-sinus</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>FDG-exponential</td>
<td>0.03</td>
<td>0.03</td>
</tr>
</tbody>
</table>
Hydrological application

Assessment of flood critical levels
Critical levels

- Extreme event:

\[ \{ F_1(X_1) > q, \ldots, F_d(X_d) > q \} \]

- Return period $T$ and critical level $q$:

\[
T = \frac{1}{1 - M(q)}, \quad \text{where} \quad M(q) = P(\min(F_1(X_1), \ldots, F_d(X_d)) \leq q)
\]

A return period of $T = 30$ years and a critical level of $q = 70\%$ means that each $X_i$ exceeds its quantile of order 70% once every 30 years in average.
Estimated critical levels

Figure: Critical level $q$ as a function of the return period $T$. "empirical" stands for the empirical critical levels, and "independence" for the independence copula $C(u_1, \ldots, u_d) = \prod_{i=1}^{d} u_i$. 
Conclusion

- PBC
  - Liebscher leads to PBC
  - full likelihood of PBC can be computed via message-passing
  - PBC (and, therefore, Liebscher?) are of limited use in practice

- FDG
  - tractable, flexible, tail asymmetric
  - theoretically well grounded estimation

- Estimation: derivation of asymptotic properties without assuming partial derivatives
Achievements

Submitted papers:

▶ FDG copulas: A flexible and tractable class of one-factor copulas,
  http://hal.archives-ouvertes.fr/hal-00979147

▶ WLS estimator: Weighted least-squares inference based on dependence coefficients for multivariate copulas,
  http://hal.archives-ouvertes.fr/hal-00979151

▶ PBC: A class of multivariate copulas based on products of bivariate copulas,
  http://hal.archives-ouvertes.fr/hal-00910775

Softwares:

▶ PBC: Product of Bivariate Copulas (PBC),
  http://cran.r-project.org

▶ FDGcopulas: dealing with FDG copulas, out soon
Perspectives

- Do extreme-value copulas associated to FDG have the one-factor property?
- Derivation of EM algorithm for one-factor copulas
- Mixing different types of dependence coefficients in WLS estimator
- Extension of one-factor structure to spatial data (interpolation)
- Derivation of explicit / simple copula functionals (critical levels, value-at-risk, etc) from FDG copulas