Estimating bivarite tails, a copula based approach.

Véronique Maume-Deschamps, université Lyon 1 - ISFA,
Joint work with E. Di Bernardino and C. Prieur.

AST&Risk (ANR Project)

Framework

Goal : estimating the tail of a bivariate distribution function.

Idea : a general extension of the Peaks-Over-Threshold method.

Tools :

- a *two-dimensional version of the Pickands-Balkema-de Haan Theorem*,
- Juri & Wüthrich's and Charpentier & Juri's approach of the tail dependence.
- dependence modeled by copulas.

Asymptotic dependence as well as asymptotic independence are considered.

Summary of results

- Construction of a two-dimensional tail estimator, study of its asymptotic properties.
- A parameter that describes the nature of the tail dependence is introduced and estimated.

Other possible approaches

- Multivariate generalized Pareto distribution developed by Falk and Reiss and Rootzen and Tajvidi but the estimation of scaling parameters has to be addressed first. Our work is an alternative contribution to the Generalized Pareto distribution approach.
- Ledford and Tawn models.

$\Rightarrow$ alternative model based on regularity conditions of the copula and on the explicit description and estimation of the dependence structure in the joint tail.

Upper-tail dependence copula

$C(u, v)$ is a 2-dimensional copula and $C^*(u, v)$ is the associated survival copula.
Let X and Y be uniformly distributed on $[0, 1]$. Let u be a threshold in $[0, 1)$ such that $C^*(1 - u, 1 - u) > 0$.
The Upper-tail dependence copula at level $u \in [0, 1)$ is

$$C_u^{up}(x, y) := \mathbb{P}[X \leq \overline{F}_{X,\,u}^{-1}(x), Y \leq \overline{F}_{Y,\,u}^{-1}(y) \,|\, X > u, Y > u],$$

$\forall\, (x, y) \in [0, 1]^2$, where $\overline{F}_{X,\,u}$, $\overline{F}_{Y,\,u}$ are the distribution of X and Y conditioned on $\{X > u, Y > u\}$:

$$\overline{F}_{X,\,u}(x) := \mathbb{P}[X \leq x \,|\, X > u, Y > u] = 1 - \frac{C^*(1 - x \vee u, 1 - u)}{C^*(1 - u, 1 - u)}.$$

Limit of the upper-tail dependence copula

Assume that

$$\frac{\partial C^*(1-u, 1-v)}{\partial u} < 0 \text{ and } \frac{\partial C^*(1-u, 1-v)}{\partial v} < 0, \text{ for all } u, v \in [0,1).$$

Assume that there is a positive function G such that

$$\lim_{u \to 1} \frac{C^*(x\,(1-u),\, y\,(1-u))}{C^*(1-u,\, 1-u)} = G(x,y), \quad \text{for all } x,\, y > 0.$$

Limit of the upper-tail dependence copula

Property

Then for all $(x, y) \in [0,1]^2$

$$\lim_{u \to 1} C_u^{up}(x,y) = x + y - 1 + G(g_X^{-1}(1-x), g_Y^{-1}(1-y)) := C^{*G}(x,y), \quad (1)$$

where $g_X(x) := G(x,1)$, $g_Y(y) := G(1,y)$ *and there is a constant* $\theta > 0$ *such that, for* $x > 0$

$$G(x,y) = x^\theta g_Y(\frac{y}{x}) \text{ for } \frac{y}{x} \in [0,1], \text{ and } \quad y^\theta g_X(\frac{x}{y}) \text{ for } \frac{y}{x} \in (1, \infty).$$

Proof: adapt a result by Charpentier and Juri (2006) - they were concerned with the lower tail copula.

Standing assumptions

- X and Y are two continuous real valued random variables, with marginal distributions, F_X, F_Y, and copula C.
- $F_X \in MDA(H_{\xi_1})$, $F_Y \in MDA(H_{\xi_2})$
- C satisfies the above assumptions.

$V_{\xi_1, a_1(\cdot)}$ (resp. $V_{\xi_2, a_2(\cdot)}$) is the univariate GPD distribution with parameters ξ_1 (resp. ξ_2) and $a_1(\cdot)$ (resp. $a_2(\cdot)$) of X (resp. Y).

A two dimensional Pickands- Balkema-de Haan Theorem

Theorem

Under the standing assumptions,

$$\sup_{\mathscr{A}} \Big| \mathbb{P}\big[X - u \le x, Y - F_Y^{-1}(F_X(u)) \le y \big| X > u, Y > F_Y^{-1}(F_X(u))\big]$$

$$- C^{*\,G}\big(1 - g_X(1 - V_{\xi_1, a_1(u)}(x)), 1 - g_Y(1 - V_{\xi_2, a_2(F_Y^{-1}(F_X(u)))}(y))\big) \Big| \xrightarrow[u \to x_{F_X}]{} 0,$$

$\mathscr{A} := \{(x, y) : 0 < x \le x_{F_X} - u, 0 < y \le x_{F_Y} - F_Y^{-1}(F_X(u))\}$, *with*
$x_{F_X} := \sup\{x \in \mathbb{R} \,|\, F_X(x) < 1\}$, $x_{F_Y} := \sup\{y \in \mathbb{R} \,|\, F_Y(y) < 1\}$.

Proof: generalize the proof by Jury and Wüthrich in the case of a symmetric copula (and same marginal distributions).

Stable tail dependence function

Assume that the bivariate distribution function F has stable tail dependence function l:

$$\lim_{t\to 0} \frac{1}{t} \mathbb{P}[1 - F_X(X) \leq t\,x \text{ or } 1 - F_Y(Y) \leq t\,y] := l(x,y)$$

or equivalently

$$\lim_{t\to 0} \frac{1}{t} \mathbb{P}[1 - F_X(X) \leq t\,x, 1 - F_Y(Y) \leq t\,y] := R(x,y) = x + y - l(x,y).$$

Relation with the upper tail dependence coefficient

Recall the upper tail dependence coefficient:

$$\lambda := \lim_{t \to 0} \mathbb{P}[F_X^{-1}(X) > 1 - t \,|\, F_Y^{-1}(Y) > 1 - t].$$

Asymptotic dependence: $\lambda > 0$.
Asymptotic independence: $\lambda = 0$.
$\lambda = R(1, 1)$.

Estimators

The tail dependence function R is estimated by:

$$\widehat{R}(x,y) = \frac{1}{k_n} \sum_{i=1}^{n} 1_{\{R(X_i) > n - k_n x + 1,\, R(Y_i) > n - k_n y + 1\}},$$

where $k_n \to \infty$, $k_n/n \to 0$ and $R(X_i)$ is the rank of X_i among $(X_1, \ldots, X_n)$, $R(Y_i)$ is the rank of Y_i among $(Y_1, \ldots, Y_n)$, for $i = 1, \ldots, n$.

Estimators

The functions g_X, g_Y, G are estimated by

$$\widehat{g}_X(x) = \frac{\widehat{R}(x,1)}{\widehat{R}(1,1)}, \quad \widehat{g}_Y(x) = \frac{\widehat{R}(1,y)}{\widehat{R}(1,1)}, \quad \text{and} \quad \widehat{G}(x,y) = \frac{\widehat{R}(x,y)}{\widehat{R}(1,1)},$$

Estimators

The coefficient θ is estimated by

$$\widehat{\theta}_{\frac{y}{x}} = \frac{\log(\widehat{G}(x,y)) - \log(\widehat{g}_Y(\frac{y}{x}))}{\log(x)} \text{ if } \frac{y}{x} \in [0,1],$$

$$\widehat{\theta}_{\frac{y}{x}} = \frac{\log(\widehat{G}(x,y)) - \log(\widehat{g}_X(\frac{x}{y}))}{\log(y)} \text{ if } \frac{y}{x} \in (1,\infty).$$

Convergence results (asymptotically dependent case)

Theorem 2.2 in Einmahl *et al.* (2006) leads to the following consistency result.

Property

Under our standing assumptions, for v_n *such that* $v_n/\sqrt{k_n} \to 0$, *for* $n \to \infty$, *and* $\lambda > 0$,

$$v_n \sup_{0<x,y\leq 1} \left| \widehat{G}(x,y) - G(x,y) \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

$$v_n \sup_{0<x\leq 1} \left| \widehat{g}_X(x) - g_X(x) \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0, \quad v_n \sup_{0<y\leq 1} \left| \widehat{g}_Y(y) - g_Y(y) \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

with $k_n \to \infty$, $k_n/n \to 0$ *and* $k_n = o\left(n^{\frac{2\alpha}{1+2\alpha}}\right)$.

Convergence results (asymptotically independent case)

The case $\lambda = R(1,1) = 0$ requires second order conditions.
As in Draisma *et al.* (2004), we assume that:

$$\lim_{t \to 0} \frac{\frac{C^*(tx, ty)}{C^*(t,t)} - G(x,y)}{q_1(t)} := Q(x,y),$$

for all $x, y \geq 0$, $x + y > 0$, with

- q_1 is some positive function and Q is neither a constant nor a multiple of G.
- The above convergence is uniform on $\{x^2 + y^2 = 1\}$.
- Denote $q(t) := \mathbb{P}[1 - F_X(X) < t, 1 - F_Y(Y) < t]$.

Convergence results (asymptotically independent case)

Property

Under our standing assumptions and second order conditions, for a sequence k_n such that $a_n := n\, q(k_n/n) \to \infty$

$$\psi_n \sup_{0<x,y\leq 1} \left| \widehat{G}(x,y) - G(x,y) \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

$$\psi_n \sup_{0<x\leq 1} \left|\widehat{g}_X(x) - g_X(x)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0, \quad \psi_n \sup_{0<y\leq 1} \left|\widehat{g}_Y(y) - g_Y(y)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

with $\psi_n = o(\sqrt{a_n})$.

A new tail estimator

Main ingredients of the estimator.

- A threshold u
- Define $\widehat{u}_Y = \widehat{F}_Y^{-1}(\widehat{F}_X(u))$, with $\widehat{F}_X(u)$ the empirical distribution function and $\widehat{F}_Y^{-1}$ the empirical quantile function of Y.
- $\widehat{k}_X$, $\widehat{\sigma}_X$ (resp. $\widehat{k}_Y$, $\widehat{\sigma}_Y$) the MLE based on the excesses of X (resp. Y).
- $\widehat{F_X}^*(x)$ (resp. $\widehat{F_Y}^*(y)$) the univariate tail estimator (see McNeil (1999)):

$$\widehat{F_X}^*(x) = (1 - \widehat{F}_X(u)) V_{\widehat{k},\widehat{\sigma}}(x-u) + \widehat{F}_X(u), \quad \text{for } x > u.$$

- $\widehat{F}_1^*(u, y) = \exp\{-\widehat{l}_n(-\log(\widehat{F_X}(u)), -\log(\widehat{F_Y}^*(y)))\}$, and $\widehat{F}_2^*(x, \widehat{u}_Y) = \exp\{-\widehat{l}_n(-\log(\widehat{F_X}^*(x)), -\log(\widehat{F_Y}(\widehat{u}_Y)))\}$.

A new tail estimator

We estimate $F(x,y)$ by

$$\widehat{F}^*(x,y) = \left(\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i>u,\, Y_i>\widehat{u}_Y\}}\right)\left(1 - \widehat{g}_X\left(1 - V_{\widehat{\xi}_X,\widehat{\sigma}_X}(x-u)\right)\right.$$
$$- \widehat{g}_Y\left(1 - V_{\widehat{\xi}_Y,\widehat{\sigma}_Y}(y-\widehat{u}_Y)\right) + \widehat{G}\left(1 - V_{\widehat{\xi}_X,\widehat{\sigma}_X}(x-u), 1 - V_{\widehat{\xi}_Y,\widehat{\sigma}_Y}(y-\widehat{u}_Y)\right)\Big)$$
$$+ \widehat{F}_1^*(u,y) + \widehat{F}_2^*(x,\widehat{u}_Y) - \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i\le u,\, Y_i\le \widehat{u}_Y\}},$$

A new tail estimator

In case the second threshold is known (for example if the marginal laws are the same), we estimate $F(x, y)$ by

$$\widetilde{F}^*(x,y) = \left(\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i > u,\ Y_i > u_Y\}}\right)\left(1 - \widehat{g}_X(1 - V_{\widehat{\xi}_X, \widehat{\sigma}_X}(x-u)) - \widehat{g}_Y(1 - V_{\widehat{\xi}_Y, \widehat{\sigma}_Y}(y-u_Y)) + \widehat{G}\left(1 - V_{\widehat{\xi}_X, \widehat{\sigma}_X}(x-u), 1 - V_{\widehat{\xi}_Y, \widehat{\sigma}_Y}(y-u_Y)\right)\right) + \widehat{F}_1^*(u,y) + \widehat{F}_2^*(x,u_Y) - \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \le u,\ Y_i \le u_Y\}},$$

Assumptions on the marginals

The assumptions below are assumed both for F_X and F_Y.
First order assumptions F is in the maximum domain of attraction of Fréchet, that is $\exists \alpha > 0$ such that $\overline{F}(x) = x^{-\alpha}L(x)$ with L a *slowly varying* function.

Assumptions on the marginals

The assumptions below are assumed both for F_X and F_Y.
First order assumptions F is in the maximum domain of attraction of Fréchet, that is $\exists \alpha > 0$ such that $\overline{F}(x) = x^{-\alpha} L(x)$ with L a *slowly varying* function.

Second order assumptions as in Smith (1987), we assume that L satisfies

$$\text{SR2} \qquad \frac{L(tx)}{L(x)} = 1 + k(t)\phi(x) + o(\phi(x)),\ \forall\ t > 0,\ \text{as } x \to \infty$$

with ϕ positive and $\phi(x) \xrightarrow[x \to +\infty]{} 0$.

Hypothesis on the threshold

Hypothesis on the threshold: the ones used by Smith to obtain the univariate convergence of the MLE GPD estimator and the POT univariate estimator $\widehat{F_X}^*(x)$.
Let n be the sample size, let $u_n := \overline{f}(n)$ (threshold sequence) and $z_n := f(n)$. We assume that $\overline{f}(n) \xrightarrow[n\to\infty]{} \infty$, $f(n) \xrightarrow[n\to\infty]{} \infty$ and several relations on the asymptotic behavior of u_n and z_n.

Convergence results (asymptotically dependent case)

$\lambda > 0$, standing assumptions, first and second order conditions on the marginal laws and hypothesis on the thresholds above.

Theorem

$$\left|\sqrt{k_n}(F^*(x_n, y_n) - \widetilde{F}^*(x_n, y_n))\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

with $x_n = \overline{f}_1(n) f_1(n)$, $y_n = \overline{f}_2(n) f_2(n)$.

Moreover if $\widehat{\overline{f}}_2(n)$ *satisfies the thershlod conditions in probability then*

$$\left|\sqrt{k_n}(F^*(x_n, \widehat{y}_n) - \widehat{F}^*(x_n, \widehat{y}_n))\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

with $\widehat{y}_n = \widehat{\overline{f}}_2(n) f_2(n)$.

We have $k_n \to \infty$, $k_n/n \to 0$, $k_n = o(n^{\frac{2\alpha}{1+2\alpha}})$, $\alpha > 0$.

Convergence results (asymptotically independent case)

$\lambda = 0$, standing assumptions, first and second order conditions on the marginal laws, second order condition on the join distribution and Smith's hypothesis on the thresholds.

Theorem

$$\left|\sqrt{a_n}\left(F^*(x_n, y_n) - \widetilde{F}^*(x_n, y_n)\right)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

where $x_n = \overline{f}_1(n)f_1(n)$, $y_n = \overline{f}_2(n)f_2(n)$. *Moreover if* $\widehat{\overline{f}}_2(n)$ *satisfies the threshold conditions in probability then*

$$\left|\sqrt{a_n}\left(F^*(x_n, \widehat{y}_n) - \widehat{F}^*(x_n, \widehat{y}_n)\right)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

with $\widehat{y}_n = \widehat{\overline{f}}_2(n)f_2(n)$ *and* $a_n = n\,q(k_n/n) \to \infty$.
$k_n/n \to 0$, $\sqrt{a_n}\, q_1(q^{\leftarrow}(a_n/n)) \to 0$ *and* $k_n = o(n^{\frac{2\alpha}{1+2\alpha}})$, *for some* $\alpha > 0$.

Real data

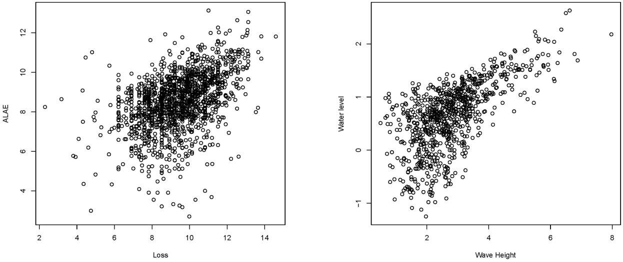

Figure: Logarithmic scale (left) ALAE versus Loss; (right) Wave heights versus Water level.

Real data

Stability of our estimation compared to the one of $\widehat{\mathscr{F}_1^*}$, as well as the estimation of parameter θ of these real cases.

$$\widehat{\mathscr{F}_1^*}(y_1, y_2) = \exp\{-\widehat{l}(-\log(\widehat{F}^*_{Y_1}(y_1)), -\log(\widehat{F}^*_{Y_2}(y_2)))\},$$

is known to produce a significant bias for asymptotically independent data.

Loss / ALEA data: asymptotic independent case

Loss-ALAE data: Each claim consists of an indemnity payment (the loss, X) and an allocated loss adjustment expense (ALAE, Y). We estimate $F(2.10^5, 10^5)$.

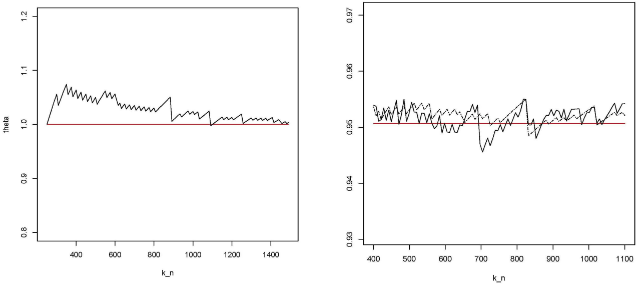

Figure: (left) $\widehat{\theta}_{0.04}$; (right) $\widehat{F}^*(2.10^5, 10^5)$ (full line), $\widehat{\mathscr{F}}_1^*(2.10^5, 10^5)$ (dashed line), with the empirical probability indicated with a horizontal line.

Wave height vs Water level: asymptotic independent case

Wave height versus Water level data: recorded during 828 storm events spread over 13 years in front of the Dutch coast near the town of Petten.

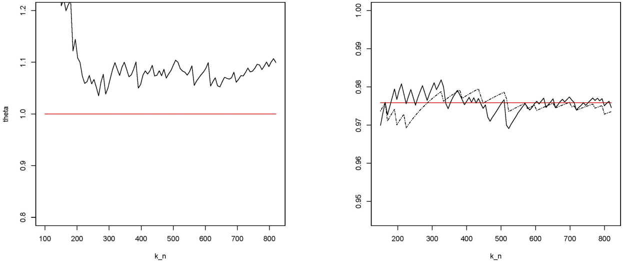

Figure: (left) $\widehat{\theta}_{\frac{0.1}{0.11}} = \widehat{\theta}_{0.91}$; (right) $\widehat{F}^*(5.93, 1.87)$ (full line), $\widehat{\mathscr{F}_1^*}(5.93, 1.87)$ (dashed line), with the empirical probability indicated with a horizontal line.

Summary

⋆ a new and different approach for estimating bivariate tails,

⋆ we need neither Ledford & Tawn assumptions nor unit Fréchet margins,

⋆ as for L & T estimate, it is particularly interesting when dealing with asymptotic independence.

Ideas for future developments

⋆ get the optimal rate, a central limit theorem?

⋆ use the bivariate tail estimator $\widehat{F}^*(x,y)$ to obtain estimation of bivariate upper-quantile curves, for high levels α.

⋆ application to the estimation of bivariate Value-at-Risk for large α :

$$VaR_\alpha(\widehat{F}) := \{(x,y) \in (\overline{f}_1(n), +\infty) \times (\widehat{\overline{f}}_2(n), +\infty) : \widehat{F}^*(x,y) = \alpha\}.$$

Thanks for your attention