ESTIMATING BIVARIATE TAILS

Clémentine PRIEUR[a]
joint work with
Elena DI BERNARDINO[b] and Véronique MAUME-DESCHAMPS[b]

[a]Université Joseph Fourier (Grenoble)
[b]ISFA, Université Lyon 1

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*,
- Yuri & Wüthrich's approach of the tail dependence.

Key words : Extreme Value Theory, Peaks Over Threshold method, Pickands-Balkema-de Haan Theorem, tail dependence.

Contents

Contents

Generalized Pareto distribution

Main idea of POT : use of the generalized Pareto distribution (1) to approximate the distribution of excesses over thresholds.

$$V_{k,\sigma}(x) := \begin{cases} 1 - \left(1 - \frac{kx}{\sigma}\right)^{\frac{1}{k}}, & \text{if } k \neq 0,\ \sigma > 0, \\ 1 - \mathrm{e}^{\frac{-x}{\sigma}}, & \text{if } k = 0,\ \sigma > 0, \end{cases} \qquad (1)$$

and $x \geq 0$ for $k \leq 0$ or $0 \leq x < \frac{\sigma}{k}$ for $k > 0$.

- Let $X_1, X_2, \ldots$ be a sequence of i.i.d random variables with unknown distribution function F.
- Fix a threshold u. For $x > u$, decompose F as

$$F(x) = \mathbb{P}[X \leq x] = (1 - \mathbb{P}[X \leq u])\, F_u(x-u) + \mathbb{P}[X \leq u],$$

where $F_u(x) = \mathbb{P}[X \leq x + u \,|\, X > u]$.

Fisher-Tippet Theorem

Theorem (Fisher-Tippet Theorem)

Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sequence with common d.f. F. If there exist a sequence of positive numbers $(a_n)_{n>0}$ and a sequence $(b_n)_{n>0}$ of real numbers such that

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{\max\{X_1, X_2, \ldots, X_n\} - b_n}{a_n} \leq x\right] = H_k(x), \quad x \in \mathbb{R}, \qquad (2)$$

for a non-degenerate distribution function $H_k(x)$, then $H_k(x)$ is a member of the Generalized Extreme Value Distribution family

$$H_k(x) = \begin{cases} \exp\left(-(1 - k\,x)^{\frac{1}{k}}\right), & \text{if } k \neq 0, \\ \exp\left(-\mathrm{e}^{-x}\right), & \text{if } k = 0, \end{cases}$$

where $1 - k\,x > 0$, $k \in \mathbb{R}$. We write $F \in MDA(H_k)$.

$k < 0$ Fréchet, $k = 0$ Gumbel, $k > 0$ Weibull.

One-dimensional Pickands-Balkema-de Haan Theorem

Let

- $F_u(x) = \mathbb{P}[X - u \leq x \,|\, X > u]$,
- $x_F := \sup\{x \in \mathbb{R} \,|\, F(x) < 1\}$ (i.e. x_F is the right endpoint of F).

Theorem (Pickands-Balkema-de Haan Theorem)

$$F \in MDA(H_k) \quad \Leftrightarrow \quad \lim_{u \to x_F} \sup_{0 \leq x < x_F - u} \left| F_u(x) - V_{k,\sigma(u)}(x) \right| = 0.$$

We deduce from the **Pickands-Balkema-de Haan** Theorem the **POT** estimate in the univariate case

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

References : MacNeil (1997,1999) and references therein.

Contents

Framework

Setting :

- X, Y two real valued r.v. with continuous df F_X and F_Y,
- the dependence between X and Y is described by a continuous and symmetric copula C.

Notation and definitions :

Survival Copula

$\forall\,(u_1, u_2) \in [0,1]^2,\; C^*(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2).$

Upper-tail dependence copula $X, Y \sim \mathcal{U}[0,1]$, with symmetric C, $u \in [0,1) \,/\, C^*(1-u, 1-u) > 0$. Then, $\forall\,(x,y) \in [0,1]^2$, one defines

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

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

Modeling upper tail, Yuri & Wütrich's approach

Theorem (Upper-tail Theorem; Juri and Wüthrich (2003))

Let C be a symmetric copula such that $C^*(1-u, 1-u) > 0$, for all $u > 0$. Furthermore, assume that there is a strictly increasing continuous function $g : [0, \infty) \to [0, \infty)$ such that

$$\lim_{u \to 1} \frac{C^*(x(1-u), 1-u)}{C^*(1-u, 1-u)} = g(x), \quad x \in [0, \infty).$$

Then, there exists a $\theta > 0$ such that $g(x) = x^\theta g\left(\frac{1}{x}\right)$ for all $x \in (0, \infty)$. Further, for all $(x, y) \in [0, 1]^2$

$$\lim_{u \to 1} C_u^{up}(x, y) = x + y - 1 + G(g^{-1}(1-x), g^{-1}(1-y)) := C^{*\,G}(x, y), \quad (3)$$

with $G(x, y) := y^\theta g\left(\frac{x}{y}\right)$ $\forall (x, y) \in (0, 1]^2$ and $G :\equiv 0$ on $[0, 1]^2 \setminus (0, 1]^2$.

Auxiliary result

Proposition (Embrechts, Kluppelberg & Mikosch, 1997)

$F_X \in MDA(H_k)$ is equivalent to the existence of a positive measurable function $a(\cdot)$ such that, for $1 - k\,x > 0$ and $k \in \mathbb{R}$,

$$\lim_{u \to x_F} \frac{1 - F_X(u + x\,a(u))}{1 - F_X(u)} = \begin{cases} (1 - k\,x)^{\frac{1}{k}}, & \text{if } k \neq 0, \\ \mathrm{e}^{-x}, & \text{if } k = 0. \end{cases} \quad (4)$$

$[(3) and (4)] \Rightarrow$ [a $2D$ version of the Pickands-Balkema-de Haan Theorem]

- Juri & Wüthrich (2003) for a symmetric C and if $F_X = F_Y$,
- Di Bernardino, Maume-Deschamps & P. (2010) for a symmetric C even if $F_X \neq F_Y$.

Symmetric copula C, $F_X \neq F_Y$

Theorem (2D Pickands-Balkema-de Haan Theorem)

X, Y real valued r.v. with continuous df $F_X \neq F_Y$, C a symmetric copula.
Assume $F_X \in MDA(H_{k_1})$, $F_Y \in MDA(H_{k_2})$ and $\exists\, g$ such that C satisfies the assumptions of the Upper-tail Theorem. Define

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

Then $\exists\, a_i(\cdot)$, $i = 1, 2$ as in (4) such that

$$\sup_{\mathscr{A}} \Big| \mathbb{P}\big[X - u \leq x, Y - u_Y \leq y \big| X > u, Y > u_Y\big]$$

$$- {C^*}^G\big(1 - g(1 - V_{k_1, a_1(u)}(x)), 1 - g(1 - V_{k_2, a_2(u_Y)}(y))\big) \Big| \xrightarrow[u \to x_{F_X}]{} 0.$$

Symmetric copula C, $F_X \neq F_Y$

From (3), the term
$C^{*\,G}\big(1-g(1-V_{k_1,a_1(u)}(x)),1-g(1-V_{k_2,a_2(u_Y)}(y))\big)$ is equal to

$$
\begin{aligned}
1-g(1-V_{k_1,a_1(u)}(x))-g(1-V_{k_2,a_2(u_Y)}(y))&\\
+G\big(1-V_{k_1,a_1(u)}(x),1-V_{k_2,a_2(u_Y)}(y)\big)&.
\end{aligned}
$$

Contents

A new bivariate tail estimator

Context : F bivariate df with continuous marginals F_X, F_Y. F is assumed to have a stable tail dependence function l that is $\forall x,\ y \geq 0$, the following limit exists

$$\lim_{t \to 0} t^{-1}\mathbb{P}\left(1 - F_X(X) \leq tx \text{ or } 1 - F_Y(Y) \leq ty\right) = l(x, y).$$

Then define

$$\lim_{t \to 0} t^{-1}\mathbb{P}\left(1 - F_X(X) \leq tx,\ 1 - F_Y(Y) \leq ty\right) = R(x, y).$$

We have $\forall x, y \geq 0,\ R(x, y) = x + y - l(x, y)$.

Asymptotic dependence $R(1, 1) \neq 0$.

Asymptotic independence $\forall x, y \geq 0,\ l(x, y) = x + y$. It is equivalent to $R(1, 1) = 0$.

Asymptotic dependence, symmetric C

Uper Tail Theorem of Juri & Wüthrich (2003) holds with

$$g(x) = \frac{x+1-l(x,1)}{2-l(1,1)} = \frac{R(x,1)}{R(1,1)}, \quad G(x,y) = \frac{x+y-l(x,y)}{2-l(1,1)} = \frac{R(x,y)}{R(1,1)}.$$

Moreover $\forall\, x > 0$, $g(x) = x\, g(1/x)$ that is $\theta = 1$.

We estimate $g(x)$ with the estimator of l in Einmahl, Krajina, Serger (2008) :

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

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

Estimating θ

We estimate $g(x)$ by $\hat{g}(x) = \frac{x+1-\hat{l}_n(x,1)}{2-\hat{l}_n(1,1)}$.

We estimate $G(x,y)$ by $\hat{G}(x,y) = \frac{x+y-\hat{l}_n(x,y)}{2-\hat{l}_n(1,1)}$.

Finally, we estimate the unknown parameter θ by

$$\hat{\theta} = \frac{\log \hat{g}(x) - \log \hat{g}(1/x)}{\log x}.$$

In practice, k is "optimized" for each value of x.

On simulations

Case 1 Burr(1) margins, $C(u, v)$ Gumbel, $x = 5$. 10 samples of size $n = 2000$.

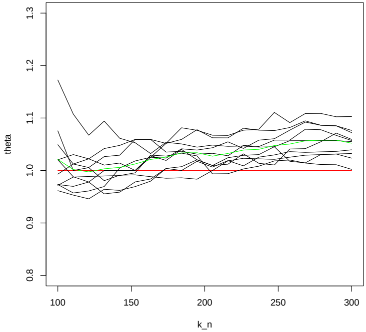

Figure: Copula Gumbel (parameter 2).

On simulations

Case 2 Burr(1) margins, $C(u, v)$ Survival Clayton, $x = 5$. 10 samples of size $n = 2000$.

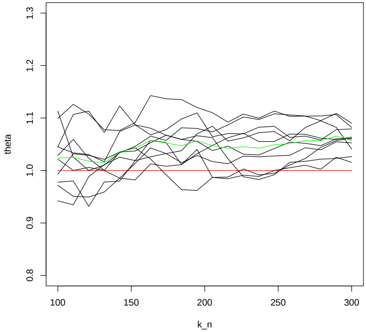

Figure: Copula Survival Clayton (parameter 1).

On simulations

Case 3 Burr(1) margins, $C(u,v) = u\,v$ (independent copula), $x = 3$. 10 samples of size $n = 2000$.

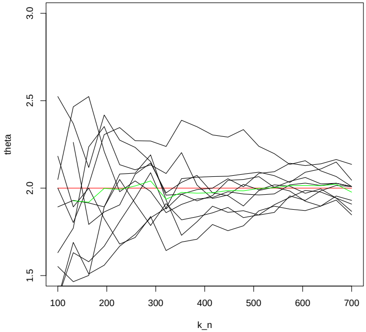

Figure: Independent Copula.

New tail estimator

For a threshold u define $\widehat{u}_Y = \widehat{F}_Y^{-1}(\widehat{F}_X(u))$.

Then, for $\widehat{k}_X$, $\widehat{\sigma}_X$ (resp. $\widehat{k}_Y$, $\widehat{\sigma}_Y$) the MLE based on the excesses of X (resp. Y), we estimate $F(x,y)$ by

$$\widehat{F}^*(x,y) = A_n\,(B_n + C_n) + \widehat{F}_1^*(u,y) + \widehat{F}_2^*(x,\widehat{u}_Y) - \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \le u,\, Y_i \le \widehat{u}_Y\}}$$

with

- $A_n = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i > u,\, Y_i > \widehat{u}_Y\}}$,
- $B_n = 1 - \widehat{g}_n(1 - V_{\widehat{k}_X,\widehat{\sigma}_X}(x-u)) - \widehat{g}_n(1 - V_{\widehat{k}_Y,\widehat{\sigma}_Y}(y - \widehat{u}_Y))$,
- $C_n = \widehat{G}_n(1 - V_{\widehat{k}_X,\widehat{\sigma}_X}(x-u), 1 - V_{\widehat{k}_Y,\widehat{\sigma}_Y}(y - \widehat{u}_Y))$,
- $\widehat{F}_1^*(u,y) = \exp\{-\widehat{l}_n(-\log(\widehat{F_X}(u)), -\log(\widehat{F_Y}^*(y)))\}$,
- $\widehat{F}_2^*(x,\widehat{u}_Y) = \exp\{-\widehat{l}_n(-\log(\widehat{F_X}^*(x)), -\log(\widehat{F_Y}(\widehat{u}_Y)))\}$.

New tail estimator

For a threshold u define $\widehat{u}_Y = \widehat{F}_Y^{-1}(\widehat{F}_X(u))$.

Then, for $\widehat{k}_X$, $\widehat{\sigma}_X$ (resp. $\widehat{k}_Y$, $\widehat{\sigma}_Y$) the MLE based on the excesses of X (resp. Y), we estimate $F(x, y)$ by

$$\widehat{F}^*(x,y) = A_n\,(B_n + C_n) + \widehat{F}_1^*(u,y) + \widehat{F}_2^*(x,\widehat{u}_Y) - \frac{1}{n}\sum_{i=1}^n 1_{\{X_i \leq u,\, Y_i \leq \widehat{u}_Y\}}$$

with

- $A_n = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i > u,\, Y_i > \widehat{u}_Y\}}$,
- $B_n = 1 - \widehat{g}_n(1 - V_{\widehat{k}_X,\widehat{\sigma}_X}(x-u)) - \widehat{g}_n(1 - V_{\widehat{k}_Y,\widehat{\sigma}_Y}(y - \widehat{u}_Y))$,
- $C_n = \widehat{G}_n(1 - V_{\widehat{k}_X,\widehat{\sigma}_X}(x-u), 1 - V_{\widehat{k}_Y,\widehat{\sigma}_Y}(y - \widehat{u}_Y))$,
- $\widehat{F}_1^*(u,y) = \exp\{-\widehat{l}_n(-\log(\widehat{F_X}(u)), -\log(\widehat{F_Y}^*(y)))\}$,
- $\widehat{F}_2^*(x,\widehat{u}_Y) = \exp\{-\widehat{l}_n(-\log(\widehat{F_X}^*(x)), -\log(\widehat{F_Y}(\widehat{u}_Y)))\}$.

New tail estimator

For a threshold u define $\widehat{u}_Y = \widehat{F}_Y^{-1}(\widehat{F}_X(u))$.

Then, for $\widehat{k}_X$, $\widehat{\sigma}_X$ (resp. $\widehat{k}_Y$, $\widehat{\sigma}_Y$) the MLE based on the excesses of X (resp. Y), we estimate $F(x, y)$ by

$$\widehat{F}^*(x, y) = A_n\ (B_n + C_n) + \widehat{F}_1^*(u, y) + \widehat{F}_2^*(x, \widehat{u}_Y) - \frac{1}{n}\sum_{i=1}^n 1_{\{X_i \leq u,\ Y_i \leq \widehat{u}_Y\}}$$

with

- $A_n = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i > u,\ Y_i > \widehat{u}_Y\}}$,
- $B_n = 1 - \widehat{g}_n(1 - V_{\widehat{k}_X, \widehat{\sigma}_X}(x - u)) - \widehat{g}_n(1 - V_{\widehat{k}_Y, \widehat{\sigma}_Y}(y - \widehat{u}_Y))$,
- $C_n = \widehat{G}_n(1 - V_{\widehat{k}_X, \widehat{\sigma}_X}(x - u), 1 - V_{\widehat{k}_Y, \widehat{\sigma}_Y}(y - \widehat{u}_Y))$,
- $\widehat{F}_1^*(u, y) = \exp\{-\widehat{l}_n(-\log(\widehat{F_X}(u)), -\log(\widehat{F_Y}^*(y)))\}$,
- $\widehat{F}_2^*(x, \widehat{u}_Y) = \exp\{-\widehat{l}_n(-\log(\widehat{F_X}^*(x)), -\log(\widehat{F_Y}(\widehat{u}_Y)))\}$.

New tail estimator

For a threshold u define $\widehat{u}_Y = \widehat{F}_Y^{-1}(\widehat{F}_X(u))$.

Then, for $\widehat{k}_X$, $\widehat{\sigma}_X$ (resp. $\widehat{k}_Y$, $\widehat{\sigma}_Y$) the MLE based on the excesses of X (resp. Y), we estimate $F(x, y)$ by

$$\widehat{F}^*(x,y) = A_n\ (B_n + C_n) + \widehat{F}_1^*(u,y) + \widehat{F}_2^*(x,\widehat{u}_Y) - \frac{1}{n}\sum_{i=1}^n 1_{\{X_i \le u,\ Y_i \le \widehat{u}_Y\}}$$

with

- $A_n = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i > u,\ Y_i > \widehat{u}_Y\}}$,
- $B_n = 1 - \widehat{g}_n(1 - V_{\widehat{k}_X,\widehat{\sigma}_X}(x-u)) - \widehat{g}_n(1 - V_{\widehat{k}_Y,\widehat{\sigma}_Y}(y - \widehat{u}_Y))$,
- $C_n = \widehat{G}_n(1 - V_{\widehat{k}_X,\widehat{\sigma}_X}(x-u), 1 - V_{\widehat{k}_Y,\widehat{\sigma}_Y}(y-\widehat{u}_Y))$,
- $\widehat{F}_1^*(u,y) = \exp\{-\widehat{l}_n(-\log(\widehat{F_X}(u)), -\log(\widehat{F_Y}^*(y)))\}$,
- $\widehat{F}_2^*(x,\widehat{u}_Y) = \exp\{-\widehat{l}_n(-\log(\widehat{F_X}^*(x)), -\log(\widehat{F_Y}(\widehat{u}_Y)))\}$.

New tail estimator

For a threshold u define $\widehat{u}_Y = \widehat{F}_Y^{-1}(\widehat{F}_X(u))$.

Then, for $\widehat{k}_X$, $\widehat{\sigma}_X$ (resp. $\widehat{k}_Y$, $\widehat{\sigma}_Y$) the MLE based on the excesses of X (resp. Y), we estimate $F(x, y)$ by

$$\widehat{F}^*(x, y) = A_n\,(B_n + C_n) + \widehat{F}_1^*(u, y) + \widehat{F}_2^*(x, \widehat{u}_Y) - \frac{1}{n}\sum_{i=1}^n 1_{\{X_i \le u,\, Y_i \le \widehat{u}_Y\}}$$

with

- $A_n = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i > u,\, Y_i > \widehat{u}_Y\}}$,
- $B_n = 1 - \widehat{g}_n(1 - V_{\widehat{k}_X, \widehat{\sigma}_X}(x - u)) - \widehat{g}_n(1 - V_{\widehat{k}_Y, \widehat{\sigma}_Y}(y - \widehat{u}_Y))$,
- $C_n = \widehat{G_n}(1 - V_{\widehat{k}_X, \widehat{\sigma}_X}(x - u), 1 - V_{\widehat{k}_Y, \widehat{\sigma}_Y}(y - \widehat{u}_Y))$,
- $\widehat{F}_1^*(u, y) = \exp\{-\widehat{l}_n(-\log(\widehat{F_X}(u)), -\log(\widehat{F_Y}^*(y)))\}$,
- $\widehat{F}_2^*(x, \widehat{u}_Y) = \exp\{-\widehat{l}_n(-\log(\widehat{F_X}^*(x)), -\log(\widehat{F_Y}(\widehat{u}_Y)))\}$.

New tail estimator

For a threshold u define $\widehat{u}_Y = \widehat{F}_Y^{-1}(\widehat{F}_X(u))$.

Then, for $\widehat{k}_X$, $\widehat{\sigma}_X$ (resp. $\widehat{k}_Y$, $\widehat{\sigma}_Y$) the MLE based on the excesses of X (resp. Y), we estimate $F(x,y)$ by

$$\widehat{F}^*(x,y) = A_n\,(B_n + C_n) + \widehat{F}_1^*(u,y) + \widehat{F}_2^*(x,\widehat{u}_Y) - \frac{1}{n}\sum_{i=1}^n 1_{\{X_i \le u,\, Y_i \le \widehat{u}_Y\}}$$

with

- $A_n = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i > u,\, Y_i > \widehat{u}_Y\}}$,
- $B_n = 1 - \widehat{g}_n(1 - V_{\widehat{k}_X,\widehat{\sigma}_X}(x-u)) - \widehat{g}_n(1 - V_{\widehat{k}_Y,\widehat{\sigma}_Y}(y-\widehat{u}_Y))$,
- $C_n = \widehat{G_n}(1 - V_{\widehat{k}_X,\widehat{\sigma}_X}(x-u), 1 - V_{\widehat{k}_Y,\widehat{\sigma}_Y}(y-\widehat{u}_Y))$,
- $\widehat{F}_1^*(u,y) = \exp\{-\widehat{l}_n(-\log(\widehat{F_X}(u)), -\log(\widehat{F_Y}^*(y)))\}$,
- $\widehat{F}_2^*(x,\widehat{u}_Y) = \exp\{-\widehat{l}_n(-\log(\widehat{F_X}^*(x)), -\log(\widehat{F_Y}(\widehat{u}_Y)))\}$.

New tail estimator

For a threshold u define $\widehat{u}_Y = \widehat{F}_Y^{-1}(\widehat{F}_X(u))$.

Then, for $\widehat{k}_X$, $\widehat{\sigma}_X$ (resp. $\widehat{k}_Y$, $\widehat{\sigma}_Y$) the MLE based on the excesses of X (resp. Y), we estimate $F(x,y)$ by

$$\widehat{F}^*(x,y) = A_n\ (B_n + C_n) + \widehat{F}_1^*(u,y) + \widehat{F}_2^*(x,\widehat{u}_Y) - \frac{1}{n}\sum_{i=1}^n 1_{\{X_i \le u,\ Y_i \le \widehat{u}_Y\}}$$

with

- $A_n = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i > u,\ Y_i > \widehat{u}_Y\}}$,
- $B_n = 1 - \widehat{g}_n(1 - V_{\widehat{k}_X,\widehat{\sigma}_X}(x-u)) - \widehat{g}_n(1 - V_{\widehat{k}_Y,\widehat{\sigma}_Y}(y-\widehat{u}_Y))$,
- $C_n = \widehat{G_n}(1 - V_{\widehat{k}_X,\widehat{\sigma}_X}(x-u), 1 - V_{\widehat{k}_Y,\widehat{\sigma}_Y}(y-\widehat{u}_Y))$,
- $\widehat{F}_1^*(u,y) = \exp\{-\widehat{l}_n(-\log(\widehat{F_X}(u)), -\log(\widehat{F_Y}^*(y)))\}$,
- $\widehat{F}_2^*(x,\widehat{u}_Y) = \exp\{-\widehat{l}_n(-\log(\widehat{F_X}^*(x)), -\log(\widehat{F_Y}(\widehat{u}_Y)))\}$.

Main steps of the construction

Distribution of excesses above u and u_Y :

$F_{u,u_Y}(x,y) := \mathbb{P}(X - u \leq x, Y - u_Y \leq y \,|\, X > u, Y > u_Y)$.

Define $\overline{F}(x,y) = \mathbb{P}(X > x, Y > y)$.

Then $\forall\, x > u,\ y > u_Y$,

$$F(x,y) = \overline{F}(u,u_Y)\, F_{u,u_Y}(x-u, y-u_Y) + F(u,y) + F(x,u_Y) - F(u,u_Y).$$

Main steps :

- using 2D Pickands-Balkema-de Haan Theorem, $F_{u,u_Y}(x-u, y-u_Y)$ is approximated by

$$C^{*\,G}\left(1 - g(1 - V_{k_X,\sigma_X(u)}(x-u)), 1 - g(1 - V_{k_Y,\sigma_Y(u_Y)}(y-u_Y))\right).$$

- we estimate $F(u,u_Y)$ and $\overline{F}(u,u_Y)$ by

$$\widehat{F}(u,u_Y) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq u,\, Y_i \leq u_Y\}}, \qquad \widehat{\overline{F}}(u,u_Y) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i > u,\, Y_i > u_Y\}}.$$

Main steps of the construction

Distribution of excesses above u and u_Y :
$F_{u,u_Y}(x,y) := \mathbb{P}(X-u \leq x, Y-u_Y \leq y \,|\, X>u, Y>u_Y)$.

Define $\overline{F}(x,y) = \mathbb{P}(X>x, Y>y)$.

Then $\forall\, x>u,\ y>u_Y$,

$$F(x,y) = \overline{F}(u,u_Y)\, F_{u,u_Y}(x-u, y-u_Y) + F(u,y) + F(x,u_Y) - F(u,u_Y).$$

Main steps :

- using 2D Pickands-Balkema-de Haan Theorem, $F_{u,u_Y}(x-u, y-u_Y)$ is approximated by

$$C^{*\,G}(1-g(1-V_{k_X,\sigma_X(u)}(x-u)), 1-g(1-V_{k_Y,\sigma_Y(u_Y)}(y-u_Y))).$$

- we estimate $F(u,u_Y)$ and $\overline{F}(u,u_Y)$ by

$$\widehat{F}(u,u_Y) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq u,\, Y_i \leq u_Y\}}, \quad \widehat{\overline{F}}(u,u_Y) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i > u,\, Y_i > u_Y\}}.$$

Main steps of the construction

Distribution of excesses above u and u_Y :
$F_{u,u_Y}(x,y) := \mathbb{P}(X - u \leq x, Y - u_Y \leq y \,|\, X > u, Y > u_Y)$.

Define $\overline{F}(x,y) = \mathbb{P}(X > x, Y > y)$.

Then $\forall\, x > u,\ y > u_Y$,

$$F(x,y) = \overline{F}(u,u_Y)\, F_{u,u_Y}(x-u, y-u_Y) + F(u,y) + F(x,u_Y) - F(u,u_Y).$$

Main steps :

• using 2D Pickands-Balkema-de Haan Theorem, $F_{u,u_Y}(x-u, y-u_Y)$ is approximated by

$$C^{*\,G}\left(1 - g(1 - V_{k_{\mathbf{X}},\sigma_{\mathbf{X}}(u)}(x-u)), 1 - g(1 - V_{k_{\mathbf{Y}},\sigma_{\mathbf{Y}}(u_{\mathbf{Y}})}(y-u_Y))\right).$$

• we estimate $F(u,u_Y)$ and $\overline{F}(u,u_Y)$ by

$$\widehat{F}(u,u_Y) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq u,\, Y_i \leq u_Y\}}, \quad \widehat{\overline{F}}(u,u_Y) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i > u,\, Y_i > u_Y\}}.$$

Main steps of the construction

Distribution of excesses above u and u_Y :
$F_{u,u_Y}(x,y) := \mathbb{P}(X-u\le x, Y-u_Y\le y \,|\, X>u, Y>u_Y)$.

Define $\overline{F}(x,y)=\mathbb{P}(X>x, Y>y)$.

Then $\forall\, x>u,\ y>u_Y$,

$$F(x,y)=\overline{F}(u,u_Y)\,F_{u,u_Y}(x-u,y-u_Y)+F(u,y)+F(x,u_Y)-F(u,u_Y).$$

Main steps :

• using 2D Pickands-Balkema-de Haan Theorem, $F_{u,u_Y}(x-u,y-u_Y)$ is approximated by

$$C^{*\,G}\left(1-g(1-V_{k_{\mathbf{X}},\sigma_{\mathbf{X}}(u)}(x-u)),1-g(1-V_{k_{\mathbf{Y}},\sigma_{\mathbf{Y}}(u_{\mathbf{Y}})}(y-u_Y))\right).$$

• we estimate $F(u,u_Y)$ and $\overline{F}(u,u_Y)$ by

$$\widehat{F}(u,u_Y)=\frac{1}{n}\sum_{i=1}^{n}1_{\{X_i\le u,\ Y_i\le u_Y\}},\qquad \widehat{\overline{F}}(u,u_Y)=\frac{1}{n}\sum_{i=1}^{n}1_{\{X_i>u,\ Y_i>u_Y\}}.$$

Main steps of the construction

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

with

- $\widehat{F}_X(u)$ (resp. $\widehat{F}_Y(u_Y)$) the empirical estimates of $F_X(u)$ (resp. $F_Y(u_Y)$),
- $\widehat{F}_X^*(x)$ (resp. $\widehat{F}_Y^*(y)$) the 1D POT estimates of $F_X(u)$ (resp. $F_Y(u_Y)$).

- we estimate u_Y by $\widehat{u}_Y := \widehat{F}_Y^{-1}(\widehat{F}_X(u))$.

Main steps of the construction

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

with

- $\widehat{F}_X(u)$ (resp. $\widehat{F}_Y(u_Y)$) the empirical estimates of $F_X(u)$ (resp. $F_Y(u_Y)$),
- $\widehat{F}_X^*(x)$ (resp. $\widehat{F}_Y^*(y)$) the 1D POT estimates of $F_X(u)$ (resp. $F_Y(u_Y)$).

- we estimate u_Y by $\widehat{u}_Y := \widehat{F}_Y^{-1}(\widehat{F}_X(u))$.

Main steps of the construction

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

with

- $\widehat{F}_X(u)$ (resp. $\widehat{F}_Y(u_Y)$) the empirical estimates of $F_X(u)$ (resp. $F_Y(u_Y)$),
- $\widehat{F}_X^*(x)$ (resp. $\widehat{F}_Y^*(y)$) the 1D POT estimates of $F_X(u)$ (resp. $F_Y(u_Y)$).

- we estimate u_Y by $\widehat{u}_Y := \widehat{F}_Y^{-1}(\widehat{F}_X(u))$.

Main steps of the construction

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

with

- $\widehat{F}_X(u)$ (resp. $\widehat{F}_Y(u_Y)$) the empirical estimates of $F_X(u)$ (resp. $F_Y(u_Y)$),
- $\widehat{F}_X^*(x)$ (resp. $\widehat{F}_Y^*(y)$) the 1D POT estimates of $F_X(u)$ (resp. $F_Y(u_Y)$).

- we estimate u_Y by $\widehat{u}_Y := \widehat{F}_Y^{-1}(\widehat{F}_X(u))$.

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

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

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

Remark : Let R_ρ be the set of $\rho-$regularly varying functions. Then, excluding trivial cases $\phi \in R_\rho$, for some $\rho \leq 0$, and $k(t) = c\, h_\rho(t)$, with $h_\rho(t) = \int_1^t u^{\rho-1} \mathrm{d}u$.

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

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

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

Remark : Let R_ρ be the set of $\rho-$regularly varying functions. Then, excluding trivial cases $\phi \in R_\rho$, for some $\rho \leq 0$, and $k(t) = c\, h_\rho(t)$, with $h_\rho(t) = \int_1^t u^{\rho-1} \mathrm{d}u$.

Univariate convergence results

Theorem (MLE Convergence Theorem, (Smith, 1987))

Assume L satisfies SR2. Let $Z_1, \dots, Z_{m_n}$ i.i.d from an unknown distribution function $F_{u_{m_n}}$ where $\lim_{n\to\infty} m_n = \infty$, $\lim_{n\to\infty} \frac{m_n}{n} = 0$. For each m_n we define a threshold $u_{m_n} := \overline{f}(m_n) \xrightarrow[n\to\infty]{} \infty$ such that

$$\frac{\sqrt{m_n}\, c\, \phi(\overline{f}(m_n))}{\alpha - \rho} \xrightarrow[n\to\infty]{} \mu \in (-\infty, \infty).$$

We define $k = -\alpha^{-1}$ and $\sigma_{m_n} = \overline{f}(m_n)\alpha^{-1}$. Then there exists a local maximum $(\widehat{\sigma}_{m_n}, \widehat{k}_{m_n})$ of the GPD log likelihood function, such that

$$\sqrt{m_n} \begin{pmatrix} \frac{\widehat{\sigma}_{m_n}}{\sigma_{m_n}} - 1 \\ \widehat{k}_{m_n} - k \end{pmatrix} \xrightarrow[n\to\infty]{d} \mathcal{N}\left(\begin{pmatrix} \frac{\mu(1-k)(1+2k\rho)}{1-k+k\rho} \\ \frac{\mu(1-k)k(1+\rho)}{1-k+k\rho} \end{pmatrix}; M^{-1} \right).$$

Univariate convergence results

The previous result is written conditionally on $N = m_n$. In practice the threshold u is fixed and N is considered as random. We give below a version of the *MLE Convergence Theorem*, unconditionally on N.

Corollary (Di Bernardino, Maume-Deschamps & P., 2010)

Assume L satisfies SR2. Let n be the sample size and $u_n := \overline{f}(n)$ the threshold, such that $\overline{f}(n) \xrightarrow[n\to\infty]{} \infty$. Let $N = N_n$ denote the random number of excesses above u_n. If

$$n(1 - F_X(u_n)) \xrightarrow[n\to\infty]{} \infty, \tag{5}$$

$$\sqrt{n(1 - F_X(u_n))}c\,\phi(u_n) \xrightarrow[n\to\infty]{} \mu(\alpha - \rho), \tag{6}$$

then the MLE Convergence Theorem holds also unconditionally on N.

Univariate convergence results

The previous result is written conditionally on $N = m_n$. In practice the threshold u is fixed and N is considered as random. We give below a version of the *MLE Convergence Theorem*, unconditionally on N.

Corollary (Di Bernardino, Maume-Deschamps & P., 2010)

Assume L satisfies SR2. Let n be the sample size and $u_n := \overline{f}(n)$ *the threshold, such that* $\overline{f}(n) \xrightarrow[n\to\infty]{} \infty$. *Let* $N = N_n$ *denote the random number of excesses above* u_n. *If*

$$n(1 - F_X(u_n)) \xrightarrow[n\to\infty]{} \infty, \tag{5}$$

$$\sqrt{n(1 - F_X(u_n))}c\,\phi(u_n) \xrightarrow[n\to\infty]{} \mu(\alpha - \rho), \tag{6}$$

then the MLE Convergence Theorem holds also unconditionally on N.

A univariate central limit theorem

Below follows a clt for the absolute error :

Theorem (Di Bernardino, Maume-Deschamps & P.)

Suppose L satisfies SR2. Let n be the sample size, $u_n := \bar{f}(n) \xrightarrow[n\to\infty]{} \infty$ *and* $z_n := f(n) \xrightarrow[n\to\infty]{} \infty$ *such that* $\forall\, s \in [0,1]$ $\;z_n^{-s\,\rho} \frac{\phi(u_n z_n^s)}{\phi(u_n)} \xrightarrow[n\to\infty]{} 1$.
Let $N = N_n$ *denote the random number of excesses above* u_n.
Assume moreover (5), (6) *and*

$$\frac{\log(z_n)}{\sqrt{n(1-F(u_n))}} \xrightarrow[n\to\infty]{} 0 \quad ,$$

$$z_n^{\alpha}(n(1-F(u_n)))^{-1/2} \xrightarrow[n\to\infty]{} 0 \quad . \tag{7}$$

Then $\dfrac{\sqrt{N}}{\log(f(n))\,\widehat{\overline{F}}_n(\bar{f}(n)\,f(n))}\left[F(\bar{f}(n)\,f(n)) - \widehat{F}^*(\bar{f}(n)\,f(n)\right] \xrightarrow[n\to\infty]{d} \mathcal{N}(\nu,\tau^2).$

Convergence results in bivariate framework

Let n be the sample size.
We choose thresholds $u_{1\,n} = \overline{f}_1(n)$ (*resp.* $u_{2\,n} = \overline{f}_2(n)$) for X (*resp.* Y) and sequences $z_{1\,n} = f_1(n)$ (*resp.* $z_{2\,n} = f_2(n)$) satisfying assumptions of the univariate clt.
We have

$$r_n \left| F(\overline{f}_1(n) f_1(n), \overline{f}_2(n) f_2(n)) - \widehat{F}^*(\overline{f}_1(n) f_1(n), \overline{f}_2(n) f_2(n)) \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Remark : we can replace $\overline{f}_2(n)$ by $\hat{\overline{f}}_2(n)$.

If C is twice continuously differentiable, in case of asymptotic dependence, we can take $\forall \varepsilon > 0$

$$r_n = \min \left\{ n^{1/3-\varepsilon}, \frac{\sqrt{N_{\mathbf{X}}}}{\log(f_1(n))\, \hat{\overline{F}}_{\mathbf{X}}\big(f_1(n)\,\overline{f}_1(n)\big)}, \frac{\sqrt{N_{\mathbf{Y}}}}{\log(f_2(n))\, \hat{\overline{F}}_{\mathbf{X}}\big(f_2(n)\,\overline{f}_2(n)\big)} \right\}.$$

Contents

Ledford & Tawn's second order model

Model :

Let (Z_1, Z_2) a bivariate random vector with Fréchet margins.

$\mathbb{P}(Z_1 > z_1, Z_2 > z_2) = z_1^{-c_1} z_2^{-c_2} \mathcal{L}(z_1, z_2)$ with $c_1, c_2 > 0$ and

$$\mathcal{L}(z_1, z_2) \sim g_1(z_1, z_2)(1 + g_2(z_1, z_2) z_1^{\rho_1} z_2^{\rho_2}) \text{ as } z_1, z_2 \to \infty,$$

with g_1 and g_2 homogeneous functions of order 0.

Notation :

- $\eta = (c_1 + c_2)^{-1}$,
- $\rho_1 + \rho_2 = \tau$, usually $\tau < 0$.

Ledford & Tawn's second order model

Asymptotic dependence if $\eta = 1$ and $\mathcal{L}(t) \nrightarrow 0$.

Asymptotic independence if $\eta < 1$ or if $\eta = 1$ and $\mathcal{L}(t) \to 0$.
Case exact independence $\eta = 1/2$ (in that case we have $\theta = 1/\eta = 2$).
Case positive association $1/2 < \eta < 1$ or $\eta = 1$ and $\mathcal{L}(t) \to 0$.
Case negative association $0 < \eta < 1/2$.

• "Ledfor & Tawn does not work for extreme sets that are not simultaneously extreme in all components."

• Note that there exist counter-examples to Ledford & Tawn models (Schlather, 2001).

• They always work with Fréchet margins, by proceding with the following transformations :
$\widehat{Z}_{1,i} = -1/\log \hat{F}_X(X_i)$, $\widehat{Z}_{2,i} = -1/\log \hat{F}_Y(Y_i)$.
What happens then with the rate when coming back to the initial distributions?

Ledford & Tawn's second order model

Asymptotic dependence if $\eta = 1$ and $\mathcal{L}(t) \nrightarrow 0$.

Asymptotic independence if $\eta < 1$ or if $\eta = 1$ and $\mathcal{L}(t) \rightarrow 0$.
Case exact independence $\eta = 1/2$ (in that case we have $\theta = 1/\eta = 2$).
Case positive association $1/2 < \eta < 1$ or $\eta = 1$ and $\mathcal{L}(t) \rightarrow 0$.
Case negative association $0 < \eta < 1/2$.

• "Ledfor & Tawn does not work for extreme sets that are not simultaneously extreme in all components."

• Note that there exist counter-examples to Ledford & Tawn models (Schlather, 2001).

• They always work with Fréchet margins, by proceding with the following transformations :
$\widehat{Z}_{1,i} = -1/\log \hat{F}_X(X_i)$, $\widehat{Z}_{2,i} = -1/\log \hat{F}_Y(Y_i)$.
What happens then with the rate when coming back to the initial distributions?

Ledford & Tawn's second order model

Asymptotic dependence if $\eta = 1$ and $\mathcal{L}(t) \nrightarrow 0$.

Asymptotic independence if $\eta < 1$ or if $\eta = 1$ and $\mathcal{L}(t) \rightarrow 0$.
Case exact independence $\eta = 1/2$ (in that case we have $\theta = 1/\eta = 2$).
Case positive association $1/2 < \eta < 1$ or $\eta = 1$ and $\mathcal{L}(t) \rightarrow 0$.
Case negative association $0 < \eta < 1/2$.

- "Ledfor & Tawn does not work for extreme sets that are not simultaneously extreme in all components."
- Note that there exist counter-examples to Ledford & Tawn models (Schlather, 2001).
- They always work with Fréchet margins, by proceding with the following transformations :
$\widehat{Z}_{1,i} = -1/\log \hat{F}_X(X_i)$, $\widehat{Z}_{2,i} = -1/\log \hat{F}_Y(Y_i)$.
What happens then with the rate when coming back to the initial distributions?

Ledford & Tawn's second order model

Asymptotic dependence if $\eta = 1$ and $\mathcal{L}(t) \nrightarrow 0$.

Asymptotic independence if $\eta < 1$ or if $\eta = 1$ and $\mathcal{L}(t) \to 0$.
Case exact independence $\eta = 1/2$ (in that case we have $\theta = 1/\eta = 2$).
Case positive association $1/2 < \eta < 1$ or $\eta = 1$ and $\mathcal{L}(t) \to 0$.
Case negative association $0 < \eta < 1/2$.

• "Ledfor & Tawn does not work for extreme sets that are not simultaneously extreme in all components."

• Note that there exist counter-examples to Ledford & Tawn models (Schlather, 2001).

• They always work with Fréchet margins, by proceding with the following transformations :
$\widehat{Z}_{1,i} = -1/\log \hat{F}_X(X_i)$, $\widehat{Z}_{2,i} = -1/\log \hat{F}_Y(Y_i)$.
What happens then with the rate when coming back to the initial distributions?

Contents

Model Survival Clayton-Fréchet, asymptotic dependence

$$C(u,v) = u+v-1+[(1-u)^{-1}+(1-v)^{-1}-1]^{-1} \text{ (Survival Clayton copula)},$$

$$F_X(x) = F_Y(x) = \exp(-1/x) \text{ (same margins, Fréchet distribution)}.$$

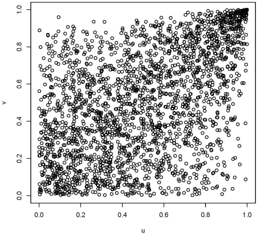

Figure: Copula Survival Clayton.

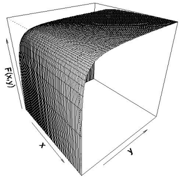

Figure: Bivariate distribution function $F_{X,Y}(x,y)$, with $F_X = F_Y$, for $x > 0, y > 0$.

We introduce

$$\widehat{\mathscr{F}_1}^*(x,y) = \exp\{-\widehat{l}_n(-log(\widehat{F_X}^*(x)), -log(\widehat{F_Y}^*(y)))\}, \qquad (8)$$

$$\widehat{\mathscr{F}_2}^*(x,y) = 1 - \widehat{l}_n(1-\widehat{F_X}^*(x), 1-\widehat{F_Y}^*(y)), \qquad (9)$$

with $\widehat{F_X}^*(x)$ (resp. $\widehat{F_Y}^*(y)$) 1D POT tail estimator for X (resp. Y).

method	$\overline{ERR_{abs}}$	$\overline{ERR_{rel}}$
classical 1	0.009907416	0.01207137
classical 2	0.01203755	0.01466676
L & T	0.02218138	0.02702618
Y & W	0.01566613	0.01908789

Table: $t = 100$ simulations of size $n = 1000$, $u_{1\,n} = u_{2\,n} = n^{1/3}/3 = 3.33333$, $z_{1\,n} = z_{2\,n} = \log n^{1/3} = 2.302585$

We introduce

$$\widehat{\mathscr{F}_1}^*(x,y) = \exp\{-\widehat{l}_n(-log(\widehat{F_X}^*(x)), -log(\widehat{F_Y}^*(y)))\}, \tag{8}$$

$$\widehat{\mathscr{F}_2}^*(x,y) = 1 - \widehat{l}_n(1-\widehat{F_X}^*(x), 1-\widehat{F_Y}^*(y)), \tag{9}$$

with $\widehat{F_X}^*(x)$ (resp. $\widehat{F_Y}^*(y)$) 1D POT tail estimator for X (resp. Y).

$method$	$\overline{ERR_{abs}}$	$\overline{ERR_{rel}}$
classical 1	0.009907416	0.01207137
classical 2	0.01203755	0.01466676
L & T	0.02218138	0.02702618
Y & W	0.01566613	0.01908789

Table: $t = 100$ simulations of size $n = 1000$, $u_{1\,n} = u_{2\,n} = n^{1/3}/3 = 3.33333$, $z_{1\,n} = z_{2\,n} = \log n^{1/3} = 2.302585$

Model Survival Clayton-Fréchet, asymptotic dependence

method	$F\left(f_1(n)\overline{f}_1(n), f_2(n)\overline{f}_2(n)\right)$	*empirical variance*
theoretic	0.8207367	
classical 1	0.8216137	0.0001566896
classical 2	0.8160857	0.0002055914
L & T	0.8143	0.000713136
Y & W	0.8310827	0.0002599203

Table: $t = 100$ simulations of size $n = 1000$

Model Survival Clayton-Burr, asymptotic dependence

$C(u,v) = u+v-1+[(1-u)^{-1}+(1-v)^{-1}-1]^{-1}$ (Survival Clayton copula),
$F_X(x) = 1-(1+x)^{-1}$, $F_Y(y) = 1-(1+y)^{-2}$ (Burr(1), Burr(2)).

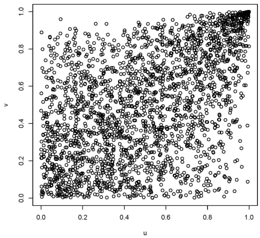

Figure: Copula Survival Clayton.

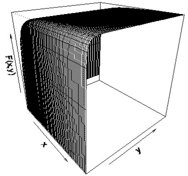

Figure: Bivariate distribution function $F_{X,Y}(x,y)$, with $F_X = F_Y$, for $x > 0, y > 0$.

Model Survival Clayton-Burr, asymptotic dependence

method	$\overline{ERR_{abs}}$	$\overline{ERR_{rel}}$
classical 1	0.01308886	0.01578057
classical 2	0.01285705	0.000192
L & T	0.01558348	0.01878820
Y & W	0.01685493	0.02128565

Table: $t = 100$ simulations of size $n = 1000$, $u_{1\,n} = n^{1/3}/3 = 3.33333$, $z_{1\,n} = \log n^{1/3} = 2.302585$, $u_{2\,n} = \sqrt{3.33333}$, $z_{2\,n} = \sqrt{2.302585}$

method	$F\left(f_1(n)\overline{f}_1(n), f_2(n)\overline{f}_2(n)\right)$	*empirical variance*
theoretic	0.8294288	
classical 1	0.8375733	0.0001816101
classical 2	0.836	0.000192
L & T	0.8210546	0.0005832912
Y & W	0.8313332	0.0006985493

Table: $t = 100$ simulations of size $n = 1000$

Model Independent-Burr, asymptotic independence

$$C(u, v) = u\, v \text{ (Independent copula)},$$
$$F_X(x) = 1 - (1 + x)^{-1},\ F_Y(y) = 1 - (1 + y)^{-2} \text{ (Burr(1), Burr(2))}.$$

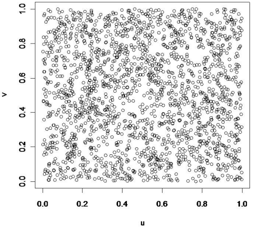

Figure: Copula Independent.

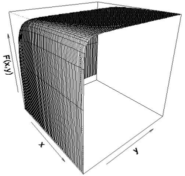

Figure: Bivariate distribution function $F_{X,Y}(x, y)$, with $F_X = F_Y$, for $x > 0, y > 0$.

Model Independent-Burr, asymptotic independence

method	$\overline{ERR}_{abs}$	$\overline{ERR}_{rel}$
classical 1	0.01039948	0.01297756
classical 2	0.02041998	0.01987981
L & T	0.00343821	0.004290557
Y & W	0.003974741	0.004960096

Table: $t = 100$ simulations of size $n = 1000$

method	$F\left(f_1(n)\overline{f}_1(n), f_2(n)\overline{f}_2(n)\right)$
theoretic	0.8013436
classical 1	0.811743
classical 2	0.820857
L & T	0.7979054
Y & W	0.8053183

Table: $t = 100$ simulations of size $n = 1000$, $u_{1\,n} = n^{1/3}/3 = 3.33333$, $z_{1\,n} = \log n^{1/3} = 2.302585$, $u_{2\,n} = \sqrt{3.33333}$, $z_{2\,n} = \sqrt{2.302585}$

Loss-ALAE

Data examined by Frees and Valdez (1998) with
X Pareto (1.122), Y Pareto (2.118), Copula Gumbel with parameter 1.4.

We then get $g(x) = \frac{1+x-(1+x^{1.4})^{1/1.4}}{2-2^{1/1.4}}$.

We choose

- $u_{1\,n} = 10000 \times n^{1/3} = 114471.4$, $z_{1\,n} = 1.7471 \Rightarrow u_{1\,n} \times z_{1\,n} = 200\,000$.
- $u_{2\,n} = \widehat{F}_Y\left(F_X(u_{1\,n})\right)$, $z_{2\,n} = 3 \Rightarrow u_{2\,n} \times z_{2\,n} = 100\,000$.

We get the estimate

$$\mathbb{P}\left(\text{Loss} \leq 200\,000,\ \text{ALAE} \leq 100\,000\right) = 0.9513696.$$

Hence $\mathbb{P}\left(\text{Loss} > 200\,000, \text{ALAE} > 100\,000\right) = 0.0067029.$
We compare with the empirical probability 0.006 (see Beirlant, Dierckx & Guillou, 2010).

Loss-ALAE

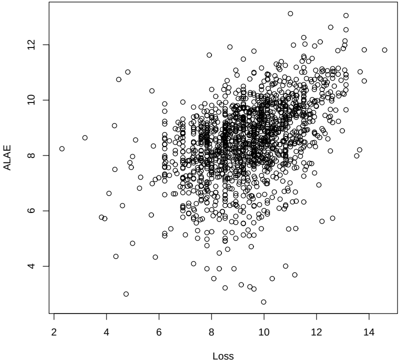

Figure: Loss-ALAE.

Loss-ALAE

Example : for $k_n = 840$ we get $\mathbb{P}(\text{Loss} \leq 200\,000, \text{ALAE} \leq 100\,000) = 0.9506583$, that is an absolute error equal to 8.436013×10^{-6} and a relative error equal to 8.835904×10^{-6}.

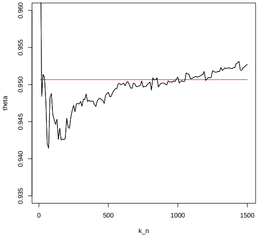

Figure: Sensibility with respect to $10 \leq k_n \leq 1500$.

Loss-ALAE

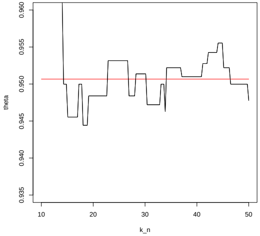

Figure: Zoom for $10 \leq k_n \leq 50$.

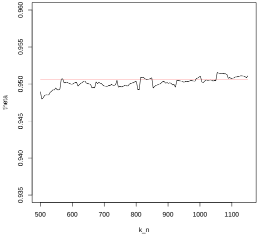

Figure: Zoom for $500 \leq k_n \leq 1150$.

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 :

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

Thank for your attention!