Rank-based M-Estimation for Tail Dependence and Independence

Michaël Lalancette[1]

Joint work with Sebastian Engelke[2] and Stanislav Volgushev[1]

[1]Department of Statistical Sciences, University of Toronto

[2]Research Center for Statistics, University of Geneva

July 4, 2019

1. Introduction: Bivariate tails and Asymptotic independence
2. Non-parametric estimation of c
3. Parametric estimation of c

- Estimation of a multivariate EVD is usually done in two steps
- 1. Estimate its margins
  2. Estimate its dependence structure
- Strategies to estimate the extremal dependence structure depend on whether Asymptotic dependence on Asymptotic independence
- Our objective: Propose a unifying way to do so in both situations

Introduction (with math)

- Suppose (X, Y) has a continuous bivariate CDF F with marginals F_1 and F_2
- Under regularity conditions in the tails (satisfied if F is in a MDA),

$$L(x, y) := \lim_{t \to 0} \frac{1}{t} \mathbb{P}\left(F_1(X) \geq 1 - tx \quad \text{or} \quad F_2(Y) \geq 1 - ty\right) \tag{1}$$

  exists for every $(x, y) \in [0, \infty)^2$
- Equivalently, we may study the function

$$\begin{aligned} R(x, y) &:= \lim_{t \to 0} \frac{1}{t} \mathbb{P}\left(F_1(X) \geq 1 - tx, F_2(Y) \geq 1 - ty\right) \\ &= x + y - L(x, y) \end{aligned} \tag{2}$$

- L is called *stable tail dependence function*, or simply the L-function ([Huang, 1992, de Haan and Ferreira, 2006])
- If F is in a MDA, L, R and the exponent measure are all equivalent

Asymptotic independence

- A popular approach to estimate the extremal dependence between X and Y is to assume eq. (1)/eq. (2) and estimate L or R
- Essentially, L allows an approximation of the probability of "at least one exceedance", and R, **if it is not 0**, allows for an approximation of the probability of joint exceedances
- But it may be that $\mathbb{P}\left(F_1(X) \geq 1 - tx, F_2(Y) \geq 1 - ty\right) = o(t)$ for every $(x, y) \in [0, \infty)^2$, which makes $R(x, y) = 0$
- X and Y are then said to be *asymptotically independent* (AI). Otherwise, they are deemed *asymptotically dependent* (AD)
- Note that if $(X, Y) \in \mathcal{D}(G)$ for a bivariate EVD G, X and Y are AI iff the two components of G are independent

Another representation of the extremal dependence

- Instead of the function R, assume the existence of

$$c(x,y) := \lim_{t\to 0} \frac{1}{q(t)} \mathbb{P}\left(F_1(X) \geq 1 - tx, F_2(Y) \geq 1 - ty\right)$$

  for a scaling function q that makes the limit non-zero
- Then there exists a *coefficient of tail dependence* $\eta \in (0,1]$ such that
  1. q is $(1/\eta)$–RV at 0
  2. c is $(1/\eta)$–homogeneous ($c(ax,ay) = a^{1/\eta} c(x,y)$)
- [Ledford and Tawn, 1997, Draisma et al., 2004]
- Essentially, q describes the strength of tail dependence and c describes the shape of the joint tail, **but they are not completely unrelated**
- Advantage: includes both AI ($q(t) = o(t)$) and *asymptotic dependence* (AD) ($q(t) \sim t$), but is not trivial in either case
- Note: For c to be unique, we assume $c(1,1) = 1$

The estimator

- Let $(X_1, Y_1), ..., (X_n, Y_n)$ be independent copies of (X, Y)
- The definition of c suggests the "estimator"

$$\hat{c}_n(x, y) := \frac{1}{q(k/n)} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1} \left\{ \hat{F}_1(X_i) \geq 1 - \frac{k}{n}x, \hat{F}_2(Y_i) \geq 1 - \frac{k}{n}y \right\},$$

  where $\hat{F}_j$ are the empirical CDF's
- This is a rank-based estimator (can be rewritten as a function of ranks)
- It appears in [Draisma et al., 2004]

Asymptotic normality of $\hat{c}_n$

- Assume that as $t \to 0$,

$$\frac{1}{q(t)} \mathbb{P}\left(F_1(X) \geq 1 - tx, F_2(Y) \geq 1 - ty\right) = c(x, y) + O(q_1(t))$$

  uniformly over $\{(x, y) \in [0, \infty)^2 : x^2 + y^2 = 1\}$
- Assume that q is positive measurable, c is not everywhere 0 and $q_1(t) = O\left(\frac{1}{\log(1/t)}\right)$
- For an intermediate sequence k, define another sequence by $m = nq(k/n)$. Assume that $m \to \infty$ and $mq_1(k/n)^2 \to 0$

Theorem (L, Engelke and Volgushev (2019))

Then, there exist Gaussian processes $W^{(1)}$ and $W^{(2)}$ on $[0, \infty)^2$ such that

1. *Under AI, $\sqrt{m}\,(\hat{c}_n - c) \rightsquigarrow W^{(1)}$ (in $\ell^\infty\left([0, T]^2\right)$).*
2. *Under AD, $\sqrt{m}\,(\hat{c}_n - c) \rightsquigarrow W^{(2)}$ (in the topo. of hypi-convergence for locally bounded functions ([Bücher et al., 2014])).*

Important remarks

- Weak assumptions (no smoothness on c, bias bounded by a very slow term)
- Basically,

$$\sqrt{m}\left(\hat{c}_n(x,y) - c(x,y)\right) = \underbrace{\text{Something}}_{\rightsquigarrow W^{(1)}} + \sqrt{m}\left(c(\hat{x}_n, \hat{y}_n) - c(x,y)\right),$$

  where $\hat{x}_n$ and $\hat{y}_n$ are based on the empirical quantiles of X and of Y
- "Something" is what one would obtain with known marginal distributions F_1, F_2. It is a fairly standard empirical process
- The other term comes from the error in estimating the marginals
- Under AD, it converges to a non trivial limit
- Under AI, it disappears because convergence of $\hat{x}_n$ and $\hat{y}_n$ is faster than convergence of "Something" to $W^{(1)}$ (based on more data)

Example 1: Inverted max-stable distributions

Suppose that $(1/X, 1/Y)$ has a bivariate max-stable distribution, with L-function L. Then under a mild smoothness assumption on L, (X, Y) satisfies our assumptions, with

$$q(t) = t^{L(1,1)}, \quad c(x, y) = x^{\dot{L}_1(1,1)} y^{\dot{L}_2(1,1)}, \quad q_1(t) = \frac{1}{\log(1/t)}$$

Example 2: A random scale construction

Suppose $R \sim \text{Pareto}(\alpha)$, $W_j \sim \text{Pareto}(1)$, where R, W_1, W_2 are independent. Then $(X, Y) = R(W_1, W_2)$ satisfies our assumptions

Range of α	$q(t)$	$c(x, y)$	$q_1(t)$
$(0,1)$	$K_\alpha t$	$(1 - r(\alpha))(x \wedge y) + r(\alpha)(x \wedge y)^{1/\alpha}(x \vee y)^{1-1/\alpha}$	$t^{1/\alpha - 1}$
1	$\frac{K_\alpha t}{\log(1/t) + \log\log(1/t)}$	$(x \wedge y)\left(1 + \frac{1}{2}\log\left(\frac{x \vee y}{x \wedge y}\right)\right)$	$\frac{1}{\log(1/t)}$
$(1,2)$	$K_\alpha t^\alpha$	$(x \wedge y)(x \vee y)^{\alpha - 1}$	$t^{(\alpha-1)\wedge(2-\alpha)}$
2	$K_\alpha t^2 \log(1/t)$	xy	$\frac{1}{\log(1/t)}$
$(2, \infty)$	$K_\alpha t^2$	xy	$t^{\alpha - 2}$

$$r(\alpha) = \frac{\alpha}{2}\left(1 - (2-\alpha)(1-\alpha)^{1/\alpha - 1}\right) \in (0,1)$$

Only thing to know: $\alpha < 1 \Rightarrow$ AD and $\alpha \geq 1 \Rightarrow$ AI

Why a parametric estimator?

- Parametric models often allow for a nice interpretation
- The non-parametric estimator $\hat{c}_n$ is not a proper function c
- More importantly, recall that $\hat{c}_n$ depends on the unknown scaling function q (through $m = nq(k/n)$)
- The following parametric estimation procedure fixes this problem

The M-estimator we need

- Assume a parametric family of functions $\{c_\theta : \theta \in \Theta \subset \mathbb{R}^p\}$
- Idea: Choose θ as to minimize

$$\left\| \int_{[0,T]^2} g(x,y) c_\theta(x,y)\, dx\, dy - \int_{[0,T]^2} g(x,y) \hat{c}_n(x,y)\, dx\, dy \right\|,$$

  where $g : [0, T]^2 \to \mathbb{R}^q$ is a vector of arbitrary weight functions
- [Einmahl et al., 2012] proposed an identical approach to estimate the L-function, but recall that the L-function is not informative under AI
- Problem: $\hat{c}_n$ can only be calculated up to the unknown scaling m
- Solution: Since

$$m\hat{c}_n(x,y) = \sum_{i=1}^{n} \mathbb{1}\left\{ \hat{F}_1(X_i) \geq 1 - \frac{k}{n}x, \hat{F}_2(Y_i) \geq 1 - \frac{k}{n}y \right\}$$

  can be calculated, simply multiply the second integral by m
- To adjust, multiply left integral by a new unknown parameter

- We obtain the following objective function:

$$\Psi_n(\theta, \sigma) := \left\| \sigma \int_{[0,T]^2} g(x,y) c_\theta(x,y)\, dx\, dy - m \int_{[0,T]^2} g(x,y) \hat{c}_n(x,y)\, dx\, dy \right\|$$

- By minimizing this objective function, we hope that c_θ will estimate c and σ will estimate m

Asymptotic normality of the M-estimator

- Suppose that the true function generating the data is c_{θ_0}, $\theta_0 \in \Theta$, and that the map

$$(\theta, \xi) \mapsto \xi \int_{[0,T]^2} g(x,y) c_\theta(x,y)\, dx\, dy$$

is continuously differentiable at $(\theta_0, 1)$ with full-rank Jacobian.

- Assume the setting of the previous theorem

Theorem (L, Engelke and Volgushev (2019))

Then if $(\hat{\theta}_n, \hat{\sigma}_n)$ is an estimator such that $\Psi_n(\hat{\theta}_n, \hat{\sigma}_n) = o_P(\sqrt{m})$,

$$\sqrt{m}\left(\left(\hat{\theta}_n, \frac{\hat{\sigma}_n}{m}\right) - (\theta_0, 1)\right) \rightsquigarrow N\left(0, \Sigma(\theta_0)\right).$$

Summary

- We use a higher order representation of the tail dependence that naturally encompasses AD and AI
- It generalizes the L and R functions
- We obtain asymptotically normal estimators of the shape of tail dependence (represented by c)
- Because c is homogeneous of order $1/\eta$, our estimator of c yields an estimator for η

Thank you!

Bücher, A., Segers, J., and Volgushev, S. (2014).
When Uniform Weak Convergence Fails: Empirical Processes for Dependence Functions and Residuals via Epi- and Hypographs.
Ann. Stat., 42:1598–1634.

de Haan, L. and Ferreira, A. (2006).
Extreme Value Theory.
Springer.

Draisma, G., Drees, H., Ferreira, A., and de Haan, L. (2004).
Bivariate Tail Estimation: Dependence in Asymptotic Independence.
Bernoulli, 10:251–280.

Einmahl, J., Krajina, A., and Segers, J. (2012).
An M-Estimator for Tail Dependence in Arbitrary Dimensions.
Ann. Stat., 40:1764–1793.

Huang, X. (1992).
Statistics of Bivariate Extreme Value Theory.
PhD thesis, Erasmus University Rotterdam.

Ledford, A. and Tawn, J. (1997).
Modelling dependence within joint tail regions.
J. R. Stat. Soc. B, 59(2):475–499.