Local robust estimation of the Pickands dependence function

Armelle Guillou

with Mikael Escobar-Bach & Yuri Goegebeur

Annals of Statistics, 46, 2806-2843, 2018

General Framework

- $X_1, \ldots, X_n \sim^{i.i.d.} F(\mu, \sigma^2)$
- "Mean" behavior (TCL):

$$\sqrt{n}\,\frac{\frac{X_1+\ldots+X_n}{n}-\mu}{\sigma} \xrightarrow{d} \mathcal{N}(0,1), \quad n \to \infty$$

- "Extreme" behavior: $X_{1,n} \leqslant \ldots \leqslant X_{n,n}$

$$\mathbb{P}\left(\frac{X_{n,n}-b_n}{a_n} \leqslant x\right) = F^n(a_n x + b_n) \longrightarrow H_\gamma(x), \quad n \to \infty$$

General Framework

$$H_\gamma(x) = \begin{cases} \exp\Big(-(1+\gamma x)^{-\frac{1}{\gamma}}\Big) & \text{if } \gamma \neq 0 \\ \exp\Big(-\exp(-x)\Big) & \text{if } \gamma = 0 \end{cases}$$

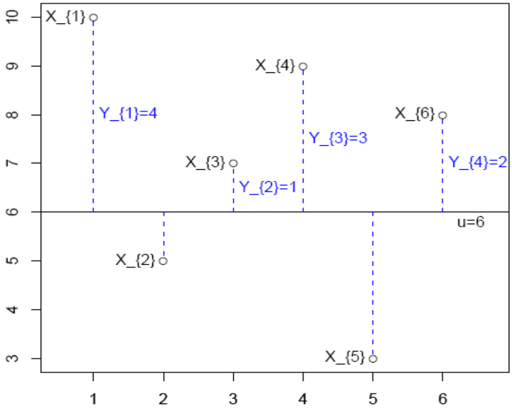
γ: extreme value index

$$\begin{cases} \gamma > 0 & \text{Fréchet, distribution of Pareto-type: } 1 - F(x) = x^{-\frac{1}{\gamma}} \ell_F(x) \\ \gamma < 0 & \text{Weibull, } X \text{ has a finite endpoint} \\ \gamma = 0 & \text{Gumbel, exponential decreasing tail} \end{cases}$$

"Peaks-Over-Thresholds" (POT) approach

$u_n < \tau_F := \sup\{x : F(x) < 1\}$ non random threshold

$Y_1, \ldots, Y_{N_n}$ the excesses above u_n



"Peaks-Over-Thresholds" (POT) approach

$$F_{u_n}(x) = \mathbb{P}\Big(X \leqslant u_n + x \Big| X > u_n\Big) = \frac{F(u_n + x) - F(u_n)}{1 - F(u_n)}$$

- Pickands (1975)

$$\sup_{x \in [0, \tau_F - u_n[} \left| F_{u_n}(x) - G_{\gamma, \sigma(u_n)}(x) \right| \to 0, \text{ as } n \to \infty$$

$$\text{where} \quad G_{\gamma,\sigma}(x) = \begin{cases} 1 - (1 + \frac{\gamma x}{\sigma})^{-\frac{1}{\gamma}} & \text{if } \gamma \neq 0 \\ 1 - \exp(-\frac{x}{\sigma}) & \text{if } \gamma = 0 \end{cases}$$

Introduction

Modelling dependence among extremes is of primary importance in practical applications where extreme phenomena occurr

Sklar (1959):

$$\mathbb{P}\left(Y^{(1)} \leqslant y_1, Y^{(2)} \leqslant y_2\right) = C\left(F_1(y_1), F_2(y_2)\right)$$

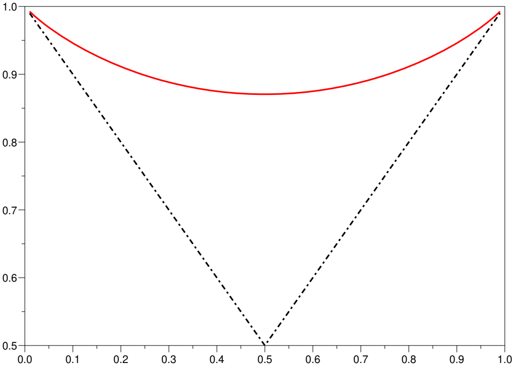
- C characterizes the dependence between $Y^{(1)}$ and $Y^{(2)}$
- C is called an extreme value copula if and only

$$C(y_1, y_2) = \exp\left(\log(y_1 y_2) A\left(\frac{\log(y_2)}{\log(y_1 y_2)}\right)\right)$$

where A: $[0,1] \to [1/2, 1]$ is the Pickands dependence function

Introduction: Pickands dependence function

A is convex and satisfies $\max\{t, 1-t\} \leqslant A(t) \leqslant 1$



$$C(y_1, y_2) = \exp\left(\log(y_1 y_2) A\left(\frac{\log(y_2)}{\log(y_1 y_2)}\right)\right)$$

Motivation 1: covariate

Aim: Extend to the case where the pair $(Y^{(1)}, Y^{(2)})$ is recorded along with a random covariate $X \in \mathbb{R}^p$

- the copula function depend on the covariate
- the marginal distribution functions depend on the covariate

$$\mathbb{P}\left(F_1(Y^{(1)}|x) \leqslant y_1, F_2(Y^{(2)}|x) \leqslant y_2 \middle| X = x\right) = C_x(y_1, y_2),$$

where C_x admits a representation of the form

$$C_x(y_1, y_2) = \exp\left(\log(y_1 y_2) A\left(\frac{\log(y_2)}{\log(y_1 y_2)} \middle| x\right)\right)$$

Portier & Segers (2018): $C_x = C$

Motivation 2: contamination

Basu et al. (1998): density power divergence between two densities g and h

$$\Delta_\alpha(g,h) := \begin{cases} \int_{\mathbb{R}} \left[h^{1+\alpha}(y) - \left(1+\frac{1}{\alpha}\right) h^\alpha(y) g(y) + \frac{1}{\alpha} g^{1+\alpha}(y) \right] dy, & \alpha > 0 \\ \int_{\mathbb{R}} \log \frac{g(y)}{h(y)} g(y) dy, & \alpha = 0 \end{cases}$$

$$h = h_\theta \qquad Z_1, \ldots, Z_n \text{ iid } g$$

The MDPDE $\widehat{\theta}$ minimizes the empirical version

$$\widehat{\Delta}_\alpha(\theta) := \begin{cases} \int_{\mathbb{R}} h^{1+\alpha}(y) dy - \left(1+\frac{1}{\alpha}\right) \frac{1}{n} \sum_{i=1}^n h^\alpha(Z_i) & \alpha > 0, \\ -\frac{1}{n} \sum_{i=1}^n \log h(Z_i) & \alpha = 0 \end{cases}$$

Simple framework: Case of know margins

Assume:

- $F_1(.|x)$ and $F_2(.|x)$ are standard exponential distribution functions

- $A_0(.|x) =$ true conditional Pickands dependence function associated to this pair $(Y^{(1)}, Y^{(2)})$

$$\begin{aligned} G(y_1, y_2|x) &:= \mathbb{P}\left(Y^{(1)} > y_1, Y^{(2)} > y_2 \middle| X = x\right) \\ &= \exp\left(-(y_1 + y_2)A_0\left(\frac{y_2}{y_1 + y_2}\middle| x\right)\right) \end{aligned}$$

Simple framework: Case of know margins

Consider

$$Z_t := \min\left(\frac{Y^{(1)}}{1-t}, \frac{Y^{(2)}}{t}\right)$$

it is clear that

$$\mathbb{P}(Z_t > z | X = x) = e^{-zA_0(t|x)}, \quad \forall z > 0 \text{ and } x \in \mathbb{R}^p$$

$$\Longrightarrow \mathcal{L}(Z_t | X = x) \text{ is } \mathcal{E}xp(A_0(t|x))$$

Simple framework: Case of know margins

$(Z_{t,i}, X_i), i = 1, \ldots, n$, be independent copies of the random pair (Z_t, X)

Robust estimator for $A_0(t|x)$ by fitting this exponential distribution function locally to $Z_{t,i}$, $i = 1, ..., n$, by means of the MDPD criterion, adjusted to locally weighted estimation, i.e. we minimize for $\alpha > 0$

$$\widehat{\Delta}_{\alpha,x,t}(A(t|x)) \quad = \quad \frac{[A(t|x)]^{\alpha}}{n} \sum_{i=1}^{n} K_h(x - X_i) \left\{ \frac{1}{1+\alpha} - \left(1 + \frac{1}{\alpha}\right) e^{-\alpha A(t|x) Z_{t,i}} \right\}$$

- $K_h(.) := K(./h)/h^p$ where K is a joint density on $\mathbb{R}^p$
- $h = h_n$ is a positive non-random sequence satisfying $h_n \to 0$ as $n \to \infty$

Simple framework: Our aim

The MDPDE for $A(t|x)$ satisfies the estimating equation

$$\widehat{\Delta}^{(1)}_{\alpha,x,t}(A(t|x)) = 0$$

Aim: To show the weak convergence of the stochastic process

$$\left\{\sqrt{nh^p}\left(\widehat{A}_{\alpha,n}(t|x) - A_0(t|x)\right), t \in [0,1]\right\},$$

in the space of all continuous functions on $[0,1]$, denoted as $\mathcal{C}([0,1])$, when $n \to \infty$.

Simple framework: Case of know margins

Our starting point: the estimating equation

$$\widehat{\Delta}^{(1)}_{\alpha,x,t}(A(t|x)) = 0$$

Taylor series expansion around $A_0(t|x)$

$$\begin{aligned} 0 &= \widehat{\Delta}^{(1)}_{\alpha,x,t}(A_0(t|x)) + \left(\widehat{A}_{\alpha,n}(t|x) - A_0(t|x)\right)\widehat{\Delta}^{(2)}_{\alpha,x,t}(A_0(t|x)) \\ &+ \frac{1}{2}\left(\widehat{A}_{\alpha,n}(t|x) - A_0(t|x)\right)^2 \widehat{\Delta}^{(3)}_{\alpha,x,t}(\widetilde{A}(t|x)) \end{aligned}$$

$$\Longrightarrow \sqrt{nh^p}\left(\widehat{A}_{\alpha,n}(t|x) - A_0(t|x)\right)$$

$$= \frac{-\sqrt{nh^p}\,\widehat{\Delta}^{(1)}_{\alpha,x,t}(A_0(t|x))}{\widehat{\Delta}^{(2)}_{\alpha,x,t}(A_0(t|x)) + \frac{1}{2}\widehat{\Delta}^{(3)}_{\alpha,x,t}(\widetilde{A}(t|x))\left(\widehat{A}_{\alpha,n}(t|x) - A_0(t|x)\right)}$$

Simple framework: Preliminary step

Key statistic:

$$T_n(K, a, t, \lambda, \beta, \gamma | x) := \frac{a^\gamma}{n} \sum_{i=1}^{n} K_h(x - X_i) Z_{t,i}^{\beta} e^{-\lambda a Z_{t,i}}$$

for $a \in [1/2, 1]$, $t \in [0, 1]$, $\lambda, \beta \geqslant 0$ and $\gamma \in \mathbb{R}$

Simple framework: Assumptions

Assumption $(\mathcal{D})$. *There exist $M_f > 0$ and $\eta_f > 0$ such that*

$$|f(x) - f(z)| \leqslant M_f \|x - z\|^{\eta_f}$$

for all $(x, z) \in S_X \times S_X$.

Assumption $(\mathcal{A}_0)$. *There exist $M_{A_0} > 0$ and $\eta_{A_0} > 0$ such that*

$$|A_0(t|x) - A_0(t|z)| \leqslant M_{A_0} \|x - z\|^{\eta_{A_0}}$$

for all $(x, z) \in B_{x_0}(r) \times B_{x_0}(r), r > 0$ and $t \in [0, 1]$.

Assumption $(\mathcal{K}_1)$. *K is a bounded density function on $\mathbb{R}^p$ with support S_K included in the unit ball of $\mathbb{R}^p$ with respect to the norm $\|.\|$.*

Simple framework: Preliminary step

Lemma 1. *Assume that for all* $t \in [0,1]$, $x \to A_0(t|x)$ *and the density function* f *are both continuous at* $x_0 \in Int(S_X)$ *non-empty. Under Assumption* $(\mathcal{K}_1)$, *if* $h \to 0$ *and* $nh^p \to \infty$, *then for* $a \in [1/2, 1]$, $\lambda, \beta \geqslant 0$, $\gamma \in \mathbb{R}$ *and* x_0 *such that* $f(x_0) > 0$, *we have*

$$T_n(K, a, t, \lambda, \beta, \gamma | x_0) \xrightarrow{\mathbb{P}} a^{\gamma} \Gamma(\beta + 1) \frac{A_0(t|x_0)}{(\lambda a + A_0(t|x_0))^{\beta+1}} f(x_0)$$

as $n \to \infty$, *where* Γ *is the gamma function defined as*

$$\Gamma(r) := \int_0^{\infty} t^{r-1} e^{-t} dt, \qquad \forall r > 0$$

Simple framework: Preliminary step

Theorem. *Let* $\gamma \in \mathbb{R}$ *and* $(\lambda, \beta) \in (0, \infty) \times \mathbb{R}_+$ *or* $(\lambda, \beta) = (0, 0)$. *Under the assumptions of Lemma 1 and if* $(\mathcal{D})$ *and* $(\mathcal{A}_0)$ *hold with* $\sqrt{nh^p} h^{\min(\eta_f, \eta_{A_0})} \to 0$, *then the process*

$$\left\{ \sqrt{nh^p} \left(T_n(K, A_0(t|x_0), t, \lambda, \beta, \gamma | x_0) - \Gamma(\beta + 1) \frac{[A_0(t|x_0)]^{\gamma - \beta}}{(\lambda + 1)^{\beta + 1}} f(x_0) \right), t \in [0, 1] \right\}$$

weakly converges in $\mathcal{C}([0, 1])$ *towards a tight centered Gaussian process* $\{B_t, t \in [0, 1]\}$

Simple framework: Preliminary step

Limiting distribution of

$$\mathbb{T}_n := \left(T_n(K, A_0(t_1|x_0), t_1, \lambda_1, \beta_1, \gamma_1|x_0), \ldots, T_n(K, A_0(t_m|x_0), t_m, \lambda_m, \beta_m, \gamma_m|x_0)\right)^T,$$

Theorem 1. *Under the assumptions of Lemma 1, we have*

$$\sqrt{nh^p}\left(\mathbb{T}_n - \mathbb{E}[\mathbb{T}_n]\right) \rightsquigarrow \mathcal{N}_m(0, \Sigma)$$

Existence and uniform consistency of $\widehat{A}_{\alpha,n}(t|x_0)$

Theorem 2. *Let $\alpha > 0$. Under the assumptions of Theorem 1, with probability tending to 1, there exists a sequence $\left(\widehat{A}_{\alpha,n}(t|x_0)\right)_{n\in\mathbb{N}}$ of solutions for the estimating equation such that*

$$\sup_{t\in[0,1]} \left|\widehat{A}_{\alpha,n}(t|x_0) - A_0(t|x_0)\right| = o_{\mathbb{P}}(1)$$

Weak convergence of the stochastic process

Theorem 3. *Let* $\left(\widehat{A}_{\alpha,n}(t|x_0)\right)_{n\in\mathbb{N}}$ *be the consistent sequence defined in Theorem 2. Under the assumptions of Theorem 1, the process*

$$\left\{\sqrt{nh^p}\left(\widehat{A}_{\alpha,n}(t|x_0) - A_0(t|x_0)\right), t \in [0,1]\right\}$$

weakly converges in $\mathcal{C}([0,1])$ *towards a tight centered Gaussian process* $\{N_t, t \in [0,1]\}$

General framework: Case of unknown margins

$F_1(.|x)$ and $F_2(.|x)$ are unknown conditional dfs

- Consider the triplets

$$\left(-\log\left(F_{n,1}(Y_i^{(1)}|X_i)\right), -\log\left(F_{n,2}(Y_i^{(2)}|X_i)\right), X_i\right)$$

- Compute

$$\check{Z}_{n,t,i} := \min\left(\frac{-\log\left(F_{n,1}(Y_i^{(1)}|X_i)\right)}{1-t}, \frac{-\log\left(F_{n,2}(Y_i^{(2)}|X_i)\right)}{t}\right)$$

- Define the key statistic

$$\check{T}_n(K,a,t,\lambda,\beta,\gamma|x_0) := \frac{a^\gamma}{n}\sum_{i=1}^{n} K_h(x_0-X_i)\check{Z}_{n,t,i}^{\beta} e^{-\lambda a \check{Z}_{n,t,i}}$$

General framework: Case of unknown margins

The MDPDE satisfies the estimating equation

$$\check{\Delta}^{(1)}_{\alpha,x,t}(A(t|x_0)) = 0,$$

where

$$\check{\Delta}_{\alpha,x,t}(a) := \frac{a^{\alpha}}{n}\sum_{i=1}^{n} K_h(x_0 - X_i)\left\{\frac{1}{1+\alpha} - \left(1+\frac{1}{\alpha}\right)e^{-\alpha a \check{Z}_{n,t,i}}\right\}.$$

Final goal: weak convergence of the stochastic process

$$\left\{\sqrt{nh^p}\left(\check{A}_{\alpha,n}(t|x_0) - A_0(t|x_0)\right), t \in [0,1]\right\}$$

Decompose

$$\sqrt{nh^p}\left(\breve{T}_n - \mathbb{E}[\breve{T}_n]\right)(K, a, t, \lambda, \beta, \gamma | x_0)$$

into two terms

$$\begin{aligned} & \left\{\sqrt{nh^p}\,(T_n - \mathbb{E}[T_n])\,(K, a, t, \lambda, \beta, \gamma | x_0)\right\} \\ + & \\ & \left\{\sqrt{nh^p}\left([\breve{T}_n - T_n] - \mathbb{E}[\breve{T}_n - T_n]\right)(K, a, t, \lambda, \beta, \gamma | x_0)\right\} \end{aligned}$$

General framework: Empirical kernel estimator of $F_j(.|x)$

$$F_{n,j}(y|x) := \frac{\sum_{i=1}^{n} K_c(x - X_i)\,\mathbb{1}_{\{Y_i^{(j)} \leqslant y\}}}{\sum_{i=1}^{n} K_c(x - X_i)}, \; j = 1, 2$$

Assumption $(\mathcal{F})$. *There exist* $M_{F_j} > 0$ *and* $\eta_{F_j} > 0$ *such that*

$$|F_j(y|x) - F_j(y|z)| \leqslant M_{F_j} \|x - z\|^{\eta_{F_j}},$$

for all $y \in \mathbb{R}$ *and all* $(x, z) \in S_X \times S_X$ *and* $j = 1, 2$.

Assumption $(\mathcal{K}_2)$. *K satisfies Assumption* $(\mathcal{K}_1)$ *and belongs to the linear span (the set of finite linear combinations) of functions* $k \geqslant 0$ *satisfying the following property: the subgraph of k,* $\{(s, u) : k(s) \geqslant u\}$*, can be represented as a finite number of Boolean operations among sets of the form* $\{(s, u) : q(s, u) \geqslant \varphi(u)\}$*, where q is a polynomial on* $\mathbb{R}^p \times \mathbb{R}$ *and* φ *is an arbitrary real function.*

General framework: Preliminary results

Lemma 2. *Assume that there exists* $b > 0$ *such that* $f(x) \geqslant b, \forall x \in S_X \subset \mathbb{R}^p$, f *is bounded, and* $(\mathcal{K}_2)$ *and* $(\mathcal{F})$ *hold. Consider a sequence c tending to 0 as* $n \to \infty$ *such that for some* $q > 1$

$$\frac{|\log c|^q}{nc^p} \longrightarrow 0.$$

Also assume that there exists an $\varepsilon > 0$ *such that for n sufficiently large*

$$\inf_{x \in S_X} \lambda(\{u \in B_0(1) : x - cu \in S_X\}) > \varepsilon, \qquad (1)$$

where λ *denotes the Lebesgue measure. Then for any* $0 < \eta < \min(\eta_{F_1}, \eta_{F_2})$, *we have*

$$\sup_{(y,x) \in \mathbb{R} \times S_X} |F_{n,j}(y|x) - F_j(y|x)| = o_{\mathbb{P}}\left(\max\left(\sqrt{\frac{|\log c|^q}{nc^p}}, c^\eta\right)\right), \text{ for } j = 1, 2.$$

General framework: Preliminary results

Theorem 4. *Assume that there exists* $b > 0$ *such that* $f(x) \geqslant b, \forall x \in S_X \subset \mathbb{R}^p$, f *is bounded, and* $(\mathcal{K}_2)$, $(\mathcal{D})$ *and* $(\mathcal{F})$ *hold together with condition (1). Consider two sequences h and c tending to 0, such that for* $nh^p \to \infty$ *and for some* $q > 1$ *and any* $0 < \eta < \min(\eta_{F_1}, \eta_{F_2})$

$$\sqrt{nh^p}\, r_n := \sqrt{nh^p} \max\left(\sqrt{\frac{|\log c|^q}{nc^p}}, c^\eta\right) \longrightarrow 0,$$

as $n \to \infty$*. Then, for all* $\gamma \in \mathbb{R}$ *and* $(\lambda, \beta) \in (0, \infty) \times \mathbb{R}_+$ *or* $(\lambda, \beta) = (0, 0)$*, we have*

$$\sup_{t \in [0,1], a \in [1/2, 1]} \sqrt{nh^p} \left| \check{T}_n - T_n - \mathbb{E}\left[\check{T}_n - T_n\right]\right| (K, a, t, \lambda, \beta, \gamma | x_0) = o_{\mathbb{P}}(1).$$

Theorem 5. *Let* $\alpha > 0$. *Under the assumptions of Theorem 4 and* $(\mathcal{A}_0)$, *with probability tending to 1, there exists a sequence* $\left(\check{A}_{\alpha,n}(t|x_0)\right)_{n\in\mathbb{N}}$ *of solutions for the estimating equation such that*

$$\sup_{t\in[0,1]} \left|\check{A}_{\alpha,n}(t|x_0) - A_0(t|x_0)\right| = o_{\mathbb{P}}(1).$$

Moreover, for this consistent sequence, if $\sqrt{nh^p}\, h^{\min(\eta_f,\eta_{A_0})} \to 0$, *the process*

$$\left\{\sqrt{nh^p}\left(\check{A}_{\alpha,n}(t|x_0) - A_0(t|x_0)\right), t \in [0,1]\right\},$$

weakly converges in $\mathcal{C}([0,1])$ *towards the tight centered Gaussian process* $\{N_t, t \in [0,1]\}$ *defined in Theorem 3.*

Small simulation study

The conditional distribution function of $(Y^{(1)}, Y^{(2)})$ given $X = x$ is a mixture model of the form

$$F_\varepsilon(y_1, y_2|x) = (1-\varepsilon)F_\ell(y_1, y_2|x) + \varepsilon F_c(y_1, y_2|x),$$

where $\varepsilon \in [0,1]$ represents the fraction of contamination in the dataset.

$$F_\ell(y_1, y_2|x) := \exp\left\{-\left(y_1^{-1/x} + y_2^{-1/x}\right)^x\right\}, \text{ for } y_1, y_2 \geqslant 0$$

and

$$A_0(t|x) = \left(t^{1/x} + (1-t)^{1/x}\right)^x,$$

where the covariate X is a uniformly distributed random variable on $[0,1]$

- complete dependence as $x \downarrow 0$
- independence as $x = 1$

First type of contamination

Given $X = x$, the distribution function F_c is

$$F_c(y_1, y_2|x) = \frac{1}{2}\left\{e^{-y_1^{-1}} + e^{-y_2^{-1}}\right\} \mathbb{1}_{\{y_1 \geqslant 0, y_2 \geqslant 0\}}$$

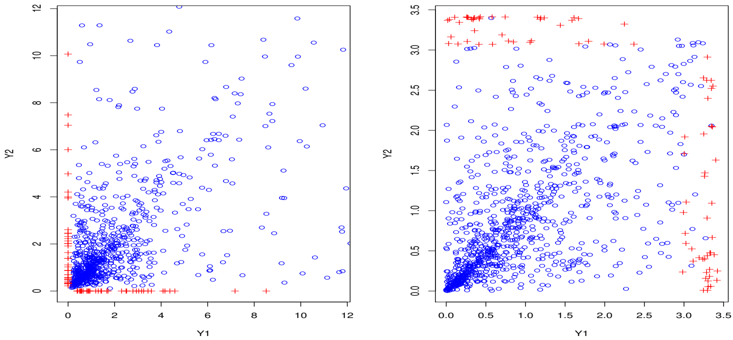


Figure 1: On the left the original data and on the right the data after transformation into (approximate) unit exponentials. Here ε is set to the value 0.1.

The distribution function F_c has completely dependent unit exponential margins

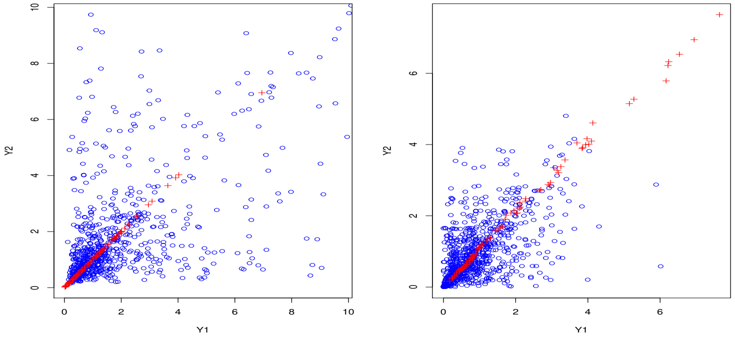


Figure 2: On the left the original data and on the right the data after transformation into (approximate) unit exponentials. Here ε is set to the value 0.1.

Cross validation for the sequence c

A random selection of size $n_r := n \wedge 1000$ from the original observations is obtained: $\{(Y_{i,r}^{(1)}, Y_{i,r}^{(2)}, X_{i,r})\}_{i=1,\dots,n_r}$

$$c_j := \arg\min_{\tilde{c}_j \in \mathcal{C}} \sum_{i=1}^{n_r} \sum_{k=1}^{n_r} \left[\mathbb{1}_{\left\{ Y_{i,r}^{(j)} \leqslant Y_{k,r}^{(j)} \right\}} - \widetilde{F}_{n_r,-i,j}(Y_{k,r}^{(j)} | X_i) \right]^2, \; j = 1,2,$$

where $\widetilde{F}_{n_r,-i,j}(y|x) := \dfrac{\sum_{k=1,k\neq i}^{n_r} K_{\tilde{c}_j}(x - X_{k,r}) \mathbb{1}_{\{Y_{k,r}^{(j)} \leqslant y\}}}{\sum_{k=1,k\neq i}^{n_r} K_c(x - X_{k,r})}$

Small simulation study: Our estimator $\check{A}_{\alpha,n}$

Cross validation for the sequence h

$$h := \arg\min_{\tilde{h}\in\mathcal{H}} \frac{1}{n_r M}\sum_{i=1}^{n_r}\sum_{j=1}^{M}\check{A}_{\alpha,n,(-i)}\left(t_j|X_{i,r}\right)^{\alpha}\left(\frac{1}{1+\alpha}-\left(1+\frac{1}{\alpha}\right)e^{-\alpha\check{A}_{\alpha,n,(-i)}\left(t_j|X_{i,r}\right)\check{Z}_{n,t_j,i,r}}\right)$$

where $\check{A}_{\alpha,n,(-i)}(t|x)$ denotes the estimator of $A_0(t|x)$ obtained on all but observation i, $\check{Z}_{n,t_j,i,r}$ is as $\check{Z}_{n,t_j,i}$ but now calculated for $(Y_{i,r}^{(1)}, Y_{i,r}^{(2)}, X_{i,r})$, and

$$\check{A}_{0,n}(t|x) := \frac{\sum_{i=1}^{n} K_{\tilde{h}}(x-X_i)}{\sum_{i=1}^{n} K_{\tilde{h}}(x-X_i)\check{Z}_{t,i}}$$

Small simulation study

- $\mathcal{C} = \{0.06, 0.12, 0.18, 0.24, 0.3\}$ and $\mathcal{H} = \{0.02, 0.03, 0.04, 0.05, 0.06\}$
- bi-quadratic function $K(x) := \frac{15}{16}(1 - x^2)^2 \mathbb{1}_{[-1,1]}(x)$
- procedure repeated $N = 200$ times, and sample sizes $n = 1000$ and 5000
- **Indicator of efficiency**

$$\text{MISE}(\varepsilon, \alpha|x) \quad := \quad \frac{1}{NM} \sum_{i=1}^{N} \sum_{m=1}^{M} \left[\check{A}^{(i)}_{\alpha,\varepsilon,n}(t_m|x) - A_0(t_m|x) \right]^2$$

where $\check{A}^{(i)}_{\alpha,\varepsilon,n}(t_m|x)$ is our estimator obtained with the $i-$th sample when the contamination is ε.

First contamination: $\text{MISE}(\varepsilon, \alpha | x), n = 5000$

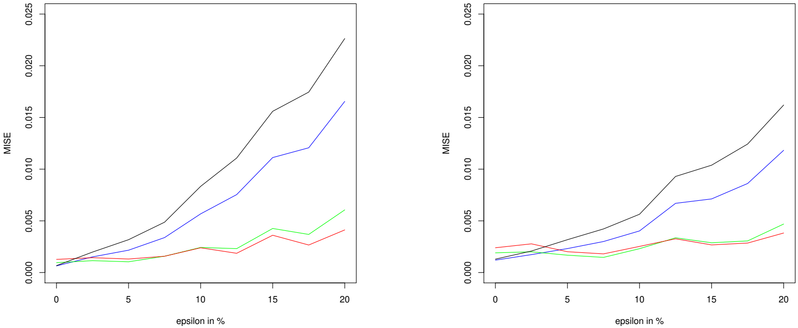


Figure 3: $\alpha = 0$ (black), $\alpha = 0.1$ (blue), $\alpha = 0.5$ (green) and $\alpha = 1$ (red). Here $x = 0.1$ (left) and 0.5 (right).

Second contamination: MISE$(\varepsilon, \alpha | x)$, $n = 5000$

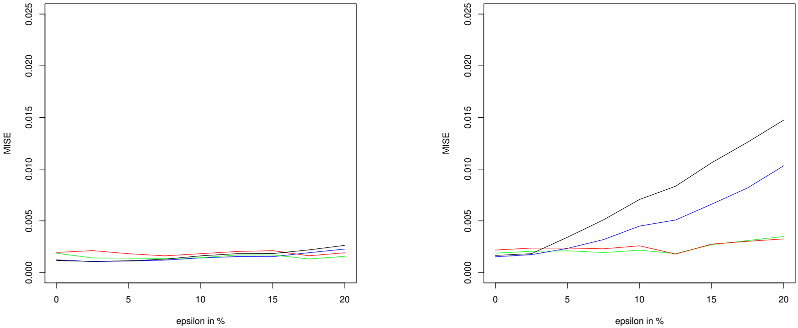


Figure 4: $\alpha = 0$ (black), $\alpha = 0.1$ (blue), $\alpha = 0.5$ (green) and $\alpha = 1$ (red). Here $x = 0.5$ (left) and 0.9 (right).

Coverage probabilities of 90% confidence intervals

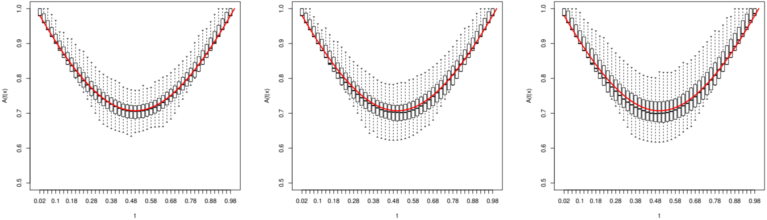
$x = 0.1$	α	$t = 0.3$				$t = 0.5$				$t = 0.7$			
		0	0.1	0.5	1	0	0.1	0.5	1	0	0.1	0.5	1
$n = 5000$	$\varepsilon = 0.0$	0.95	0.95	0.96	0.98	0.95	0.96	0.96	0.97	0.95	0.96	0.96	0.98
	$\varepsilon = 0.1$	0.14	0.28	0.84	0.90	0.06	0.12	0.69	0.82	0.13	0.29	0.84	0.90
	$\varepsilon = 0.2$	0.02	0.05	0.55	0.83	0.00	0.01	0.16	0.51	0.01	0.03	0.56	0.80

$x = 0.5$	α	$t = 0.3$				$t = 0.5$				$t = 0.7$			
		0	0.1	0.5	1	0	0.1	0.5	1	0	0.1	0.5	1
$n = 5000$	$\varepsilon = 0.0$	0.97	0.99	0.98	0.97	0.97	0.97	0.97	0.95	0.96	0.97	0.97	0.98
	$\varepsilon = 0.1$	0.30	0.53	0.91	0.95	0.39	0.56	0.90	0.91	0.31	0.54	0.93	0.94
	$\varepsilon = 0.2$	0.04	0.10	0.70	0.87	0.06	0.11	0.66	0.80	0.05	0.08	0.69	0.86

$x = 0.5$	α	$t = 0.3$				$t = 0.5$				$t = 0.7$			
		0	0.1	0.5	1	0	0.1	0.5	1	0	0.1	0.5	1
$n = 5000$	$\varepsilon = 0.0$	0.96	0.96	0.97	0.96	0.94	0.93	0.96	0.96	0.96	0.97	0.96	0.96
	$\varepsilon = 0.1$	0.93	0.98	0.96	0.96	0.48	0.69	0.94	0.96	0.95	0.99	0.99	0.98
	$\varepsilon = 0.2$	0.92	0.98	0.96	0.95	0.20	0.33	0.88	0.96	0.87	0.97	0.96	0.94

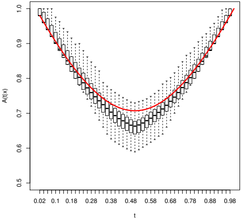
$x = 0.9$	α	$t = 0.3$				$t = 0.5$				$t = 0.7$			
		0	0.1	0.5	1	0	0.1	0.5	1	0	0.1	0.5	1
$n = 5000$	$\varepsilon = 0.0$	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00	0.98	0.98	0.99	0.99
	$\varepsilon = 0.1$	0.61	0.77	0.97	0.98	0.15	0.39	0.95	0.97	0.60	0.82	0.98	0.99
	$\varepsilon = 0.2$	0.26	0.44	0.92	0.95	0.01	0.09	0.74	0.91	0.24	0.42	0.91	0.96

Second contamination: $x = 0.5$, $\varepsilon = 0\%$

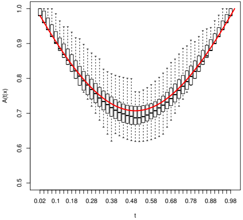


$\alpha = 0.1$ $\alpha = 0.5$ $\alpha = 1$

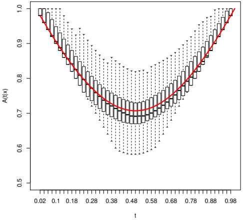
Second contamination: $x = 0.5$, $\varepsilon = 10\%$



$\alpha = 0.1$

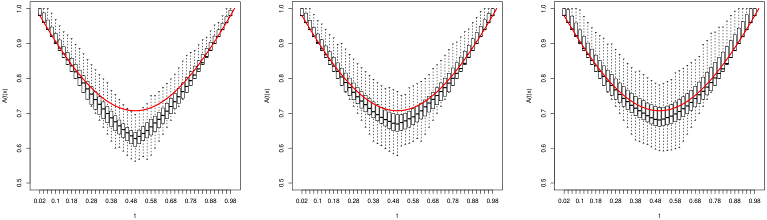


$\alpha = 0.5$



$\alpha = 1$

Second contamination: $x = 0.5$, $\varepsilon = 20\%$



$\alpha = 0.1$ $\alpha = 0.5$ $\alpha = 1$

Application to air pollution

The dataset contains daily measurements on, among others, maximum temperature, and ground level ozone, carbon monoxide and particulate matter concentrations, for the time period 1999 to 2013

Focus on ground level ozone and particulate matter concentrations
$\longrightarrow$ calculate component-wise monthly maximum of daily maximum concentrations
$\longrightarrow$ estimate the Pickands dependence function conditional on the covariates time and location (latitude, longitude)

Application to air pollution: Houston, April 2002

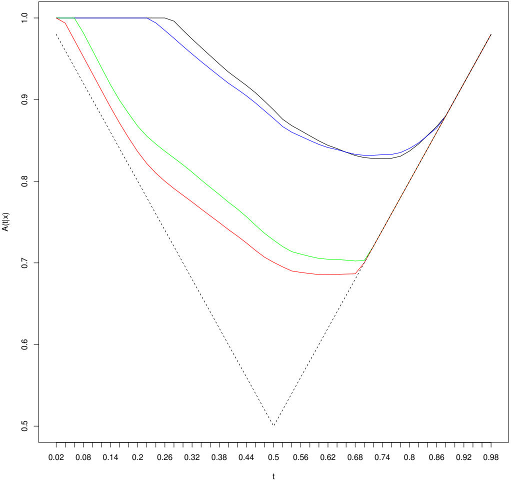


Figure 5: Conditional Pickands dependence function in April 2002, $\alpha = 0$ (black), $\alpha = 0.1$ (blue), $\alpha = 0.5$ (green) and $\alpha = 1$ (red).