Extreme Level Curves of Heavy-Tailed Distributions

Abdelaati DAOUIA [1], Laurent GARDES [2], Stéphane GIRARD [2] & Alexandre LEKINA [2]

[1] Toulouse School of Economics (GREMAQ), University of Toulouse. – [2] Team Mistis, INRIA Grenoble-Rhône-Alpes & Laboratoire Jean Kuntzmann

I. Statistical Framework

- Let $\{(X_i, Y_i),\ i = 1, \dots, n\}$ be independent copies of a random pair $(X, Y) \in \mathbb{R}^p \times \mathbb{R}$ where Y is a variable of interest associated with a covariate information X.
- Estimate for all $x \in \mathbb{R}^p$ fixed and for all $\alpha_n \to 0$, the extreme level curves defined as the graphs of the functions $x \in \mathbb{R}^p \mapsto q(\alpha_n|x) \in \mathbb{R}$ verifying

$$\bar{F}\left(q(\alpha_n|x)|x\right) \overset{def}{=} \mathbb{P}\left(Y > q(\alpha_n|x)|X = x\right) = \alpha_n,$$

when the conditional cumulative distribution function of Y given $X = x$ is heavy-tailed with tail index $\gamma(x)$, i.e,

$$q(\alpha_n|x) = \alpha_n^{-\gamma(x)}\ell(1/\alpha_n|x),$$

with $\gamma(.)$ an unknown and positive function of the covariate x called "the conditional tail index" and $\ell(.|x)$ a "slowly-varying function at infinity", i.e for all $\lambda > 0$,

$$\lim_{y\to\infty} \ell(\lambda y|x)/\ell(y|x) = 1.$$

II. Estimators

$$\hat{q}_n(\alpha_n|x) = \hat{\bar{F}}_n^{\leftarrow}(\alpha_n|x) = \inf\left\{t,\ \hat{\bar{F}}_n(t|x) \le \alpha_n\right\}.$$

$\hat{\bar{F}}_n^{\leftarrow}$ is the generalized inverse of an estimator of the conditional survival function

$$\hat{\bar{F}}_n(y|x) = \frac{1}{nh_n^p}\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)\mathbb{1}\{Y_i \ge y\} \Bigg/ \frac{1}{nh_n^p}\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) = \hat{\psi}_n(y,x)\,/\hat{g}_n(x).$$

- The kernel function $K(.)$ is positive, bounded and integrable on a compact support $S \subseteq \mathbb{R}^p$.
- The sequence of window-width $h_n \to 0$ as $n \to \infty$.
- The function $\hat{g}_n(.)$ is the classical kernel estimator of the point distribution function $g(.)$ of X.
- The function $\hat{\psi}_n(y,x)$ is an estimator of $\psi(y,x) = \bar{F}(y|x)g(x)$

III. Asymptotic distribution of $\hat{q}_n(\alpha_n|x)$

1. In [2], Collomb has established the weak consistency of the kernel estimator of the point distribution function $g(.)$.
2. If $\ell(.|x)$ is normalized and under some regularity conditions, we establish the asymptotic behavior of the estimator $\hat{\psi}_n(y_n, x)$ and we deduce the asymptotic distribution of the estimator $\hat{\bar{F}}_n(y_n|x)$ when the sequence y_n goes to infinity not too fast.
3. It follows that, for all $x \in \mathbb{R}^p$ and for all $j = 1, \dots, J$ such that $\alpha_{n,j} = \tau_j\alpha_n$, if $\alpha_n \to 0$, $nh_n^p\alpha_n \to \infty, nh_n^p\alpha_n \to \infty$ and $nh_n^{p+2}\alpha_n\log^2(\alpha_n) \to 0$ as $n \to \infty$, then we deduce that

$$\left\{\sqrt{nh_n^p\alpha_n}\left(\frac{\hat{q}_n(\alpha_{n,j}|x)}{q(\alpha_{n,j}|x)} - 1\right)\right\}_{\{j=1,\dots,J\}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0_{\mathbb{R}^J}, \gamma^2(x)\frac{\|K\|_2^2}{g(x)}\Sigma\right),$$

where $\Sigma_{j,j'}(x) = 1/\tau_{j\wedge j'}$ for $(j,j') \in \{1,\dots,J\}^2$ with (τ_j) a positive and strictly decreasing sequence.

IV. Applications

- Application 1 : A kernel version of the Hill estimator (see [3])

$$\hat{\gamma}_n^{\mathrm{H}}(x) = \sum_{j=1}^J \log\left(\hat{q}_n(\alpha_{n,j}|x)/\hat{q}_n(\alpha_{n,1}|x)\right) \Bigg/ \sum_{j=1}^J \log\left(\tau_1/\tau_j\right) \text{ with } J > 1.$$

- Application 2 : A kernel version of the Weissman estimator (see [4])

$$\hat{q}_n^{\mathrm{W}}(\beta_n|x) = \hat{q}_n(\alpha_n|x)\left(\alpha_n/\beta_n\right)^{\hat{\gamma}_n^{\mathrm{H}}(x)} \text{ where } \beta_n < \alpha_n.$$

- Second-order condition : the function $\varepsilon(y|x) \overset{def}{=} y\ell'(y|x)/\ell(y|x) \to 0$ as $y \to \infty$ and $|\varepsilon(.|x)|$ is continuous and ultimately non-increasing.
- This condition allows us to establish the asymptotic distribution of $\hat{\gamma}_n^{\mathrm{H}}(x)$ and $\hat{q}_n^{\mathrm{W}}(\beta_n|x)$.

V. Asymptotic distribution of $\hat{\gamma}_n^{\mathrm{H}}(x)$ and $\hat{q}_n^{\mathrm{W}}(\beta_n|x)$

1. If $\sqrt{nh^p\alpha_n}\varepsilon(q(\alpha_{n,1}|x)|x) \to 0$ as $n \to \infty$, then for all $x \in \mathbb{R}^p$,

$$\sqrt{nh^p\alpha_n}\left(\hat{\gamma}_n^{\mathrm{H}}(x) - \gamma(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{V_J\|K\|_2^2}{g(x)}\gamma^2(x)\right),$$

where $V_J = \left(\sum_{j=1}^J \frac{2(J-j)+1}{\tau_j} - J^2\right) \Bigg/ \left(\sum_{j=1}^J \log\left(\tau_1/\tau_j\right)\right)^2.$

2. If $\beta_n/\alpha_n \to 0$ as $n \to \infty$, then for all $x \in \mathbb{R}^p$,

$$\frac{\sqrt{nh^p\alpha_n}}{\log(\alpha_n/\beta_n)}\left(\frac{\hat{q}_n^{\mathrm{W}}(\beta_n|x)}{q(\beta_n|x)} - 1\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{V_J\|K\|_2^2}{g(x)}\gamma^2(x)\right).$$

VI. Numerical experiments on simulated data

- We have generated $m = 100$ replications of the sample $\{(X_i, Y_i), i = 1, \dots, n\}$ of size $n = 1000$ with the conditional quantile defined by

$$q(\alpha_n|x) = \left(-\log\left(1 - \alpha_n\right)\right)^{-\gamma(x)} \quad \text{(Fréchet distribution)},$$

and the following conditional tail-index function has been chosen

$$x \in [0,1] \mapsto \gamma(x) = \frac{1}{2}\left(\frac{1}{10} + \sin(\pi x)\right)\left(\frac{11}{10} - \frac{1}{2}\exp\left(-64\left(x - 1/2\right)^2\right)\right).$$

- We set $\tau_j = 1/j$ for each $j = 1, \dots, J$.To compute $\hat{\gamma}_n^{\mathrm{H}}(x)$ and $\hat{q}_n^{\mathrm{W}}(\beta_n|x)$ we have fixed $J = 9$ (since V_J is minimum for $J = 9$ with $V_9 \simeq 1.25$) and we have used a Biquadratic-kernel. We have represented the best estimators obtained in minimizing a distance between the estimated conditional extreme quantile and the true one.

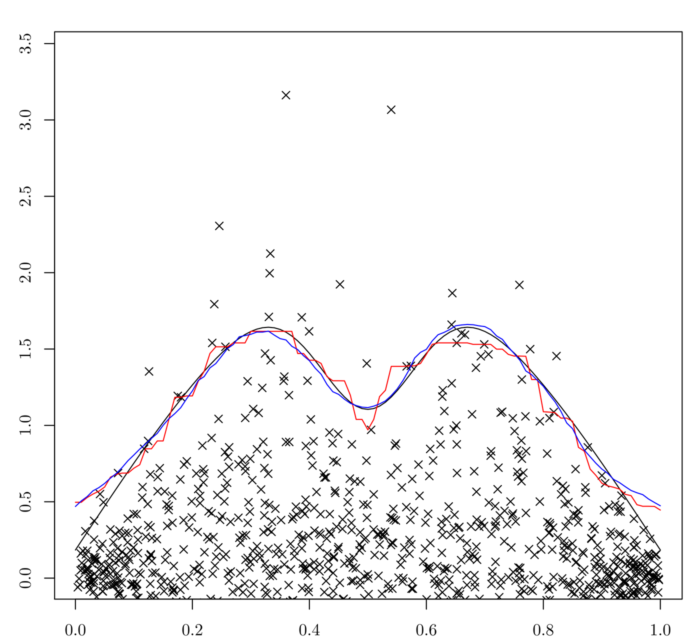

Figure 1: Comparison of the true conditional extreme quantile $q(5\log(n)/n|.)$ (black) with $\hat{q}_n(5\log(n)/n|.)$ (red) where $h_n = 0.111$ and the kernel Weissman estimator $\hat{q}_n^{\mathrm{W}}(5\log(n)/n|.)$ (blue) with $h_n = 0.111$ and $\alpha_n = 0.2$ (red). The vertical axis is in a logarithmic scale.

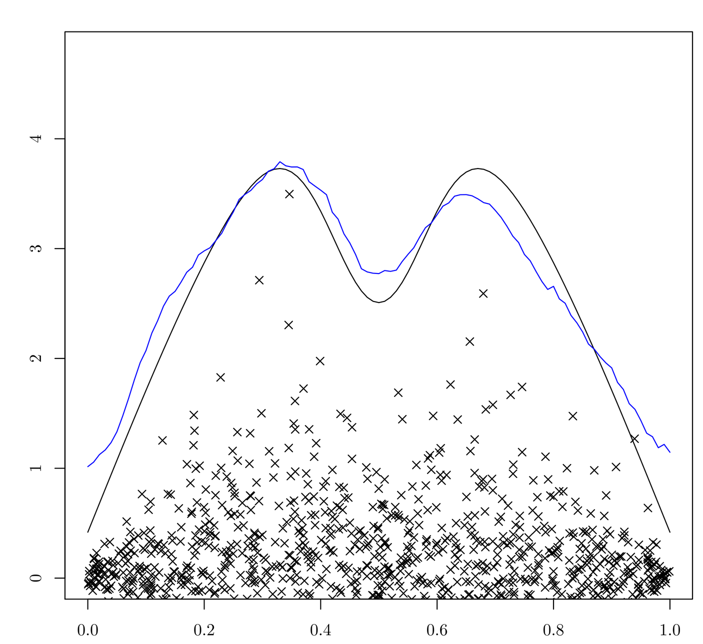

Figure 2: Comparison of the true conditional extreme quantile $q(1/2n|.)$ (black) with the kernel Weissman estimator $\hat{q}_n^{\mathrm{W}}(1/2n|.)$ (blue) with $h_n = 0.167$ and $\beta_n = 0.45$. The vertical axis is in a logarithmic scale.

References

[1] A. Daouia, L. Gardes, S. Girard and A. Lekina. Kernel estimators of extreme level curves. *Test*, to appear.

[2] G. Collomb. *Estimation non paramétrique de la régression par la méthode du noyau.* PhD thesis, Université Paul Sabatier de Toulouse, 1976.

[3] B.M. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3, 1163–1174, 1975.

[4] J. Weissman. Estimation of parameters and large quantiles based on the k-largest observations. *Journal of the American Statistical Association*, 73:812–815, 1978.

Contact: Alexandre.Lekina@inria.fr, http://mistis.inrialpes.fr/people/lekina