A moving window approach for nonparametric estimation of extreme level curves

Laurent GARDES

INRIA Rhône-Alpes, Team MISTIS
http://mistis.inrialpes.fr/people/gardes/

Joint work with Stéphane Girard and Alexandre Lekina

Let $Y \in \mathbb{R}$ be a real random value associated to a non-random covariate $x \in E$ where E is a metric space indowed by a metric d.

Goal: Estimate a conditional extreme quantile $q(\alpha, t)$ of order $1 - \alpha$ defined by

$$\mathbb{P}(Y \geq q(\alpha, x) | x) = \alpha$$

Difficulties:

- The quantile order $1 - \alpha$ can be very close to 1 (large quantile).
- The quantile is a function of the covariate x.
- The space E can be of infinite dimension.

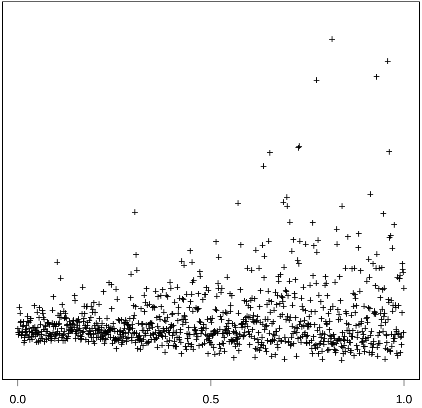
Main assumption: The conditional distribution of Y given $x \in E$ is a heavy tailed distribution *i.e.* for $\lambda > 0$,

$$\lim_{\alpha \to 0} \frac{q(\lambda\alpha, x)}{q(\alpha, x)} = \lambda^{-\gamma(x)},$$

- $\gamma(.)$ is an unknown positive function of the covariate x called the conditional tail index.
- The conditional extreme quantile $q(., x)$ decreases to infinity at a polynomial rate as $\alpha \to 0$.

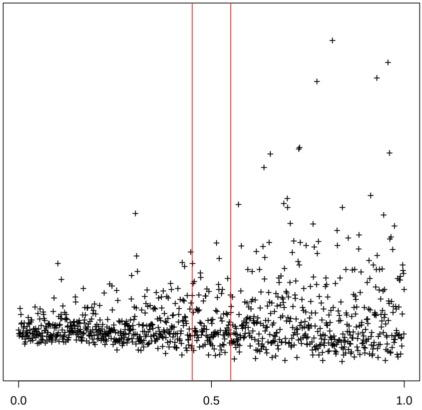
- Let $\{(Y_i, x_i), i = 1, \ldots, n\}$ be n independent copies of (Y, x). Using these observations, our aim is to estimate $q(\alpha, t)$ for a given $t \in E$.

- Let $\{(Y_i, x_i), i = 1, \ldots, n\}$ be n independent copies of (Y, x). Using these observations, our aim is to estimate $q(\alpha, t)$ for a given $t \in E$.
- Clearly, only the observations "close" to t are required.



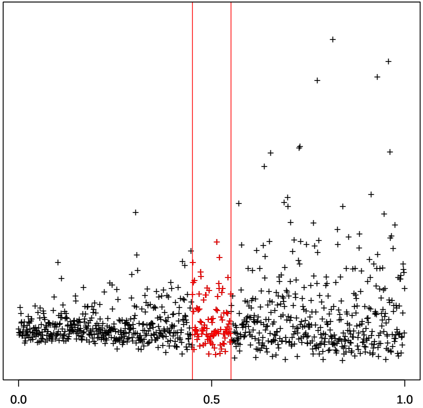
Example: Estimation at the point $t = 0.5$ using $n = 1000$ observations (Y_i, x_i), $i = 1, \ldots, n$ for $E = [0, 1]$.

- Let $\{(Y_i, x_i), i = 1, \ldots, n\}$ be n independent copies of (Y, x). Using these observations, our aim is to estimate $q(\alpha, t)$ for a given $t \in E$.
- Clearly, only the observations "close" to t are required.
- We thus choose a positive sequence $h_{n,t}$ tending to zero as $n \to \infty$ and we define the slice $S_t = (0, \infty) \times B(t, h_{n,t})$ where $B(t, h_{n,t})$ is the ball of center t and radius $h_{n,t}$.



Example: Estimation at the point $t = 0.5$ using $n = 1000$ observations $(Y_i, x_i), i = 1, \ldots, n$ for $E = [0, 1]$.

- Let $\{(Y_i, x_i), i = 1, \dots, n\}$ be n independent copies of (Y, x). Using these observations, our aim is to estimate $q(\alpha, t)$ for a given $t \in E$.
- Clearly, only the observations "close" to t are required.
- We thus choose a positive sequence $h_{n,t}$ tending to zero as $n \to \infty$ and we define the slice $S_t = (0, \infty) \times B(t, h_{n,t})$ where $B(t, h_{n,t})$ is the ball of center t and radius $h_{n,t}$.
- We select the observations Y_i's for which $x_i \in B(t, h_{n,t})$.



Example: Estimation at the point $t = 0.5$ using $n = 1000$ observations $(Y_i, x_i), i = 1, \ldots, n$ for $E = [0, 1]$.

- Let $\{(Y_i, x_i), i = 1, \ldots, n\}$ be n independent copies of (Y, x). Using these observations, our aim is to estimate $q(\alpha, t)$ for a given $t \in E$.
- Clearly, only the observations "close" to t are required.
- We thus choose a positive sequence $h_{n,t}$ tending to zero as $n \to \infty$ and we define the slice $S_t = (0, \infty) \times B(t, h_{n,t})$ where $B(t, h_{n,t})$ is the ball of center t and radius $h_{n,t}$.
- We select the observations Y_i's for which $x_i \in B(t, h_{n,t})$.
- These observations are denoted by $\{Z_i(t), i = 1, \ldots, m_{n,t}\}$ where $m_{n,t}$ is the number of x_i's in the ball $B(t, h_{n,t})$.
- The associated order statistics are denoted by

$$Z_{1,m_{n,t}}(t) \leq \ldots \leq Z_{m_{n,t},m_{n,t}}(t).$$

Influence of the rate of convergence of α

We consider three situations for the rate of convergence of α to zero:

- **(S.1)** Slow convergence of α to zero:

$$\alpha \to 0 \text{ and } m_{n,t}\alpha \to \infty.$$

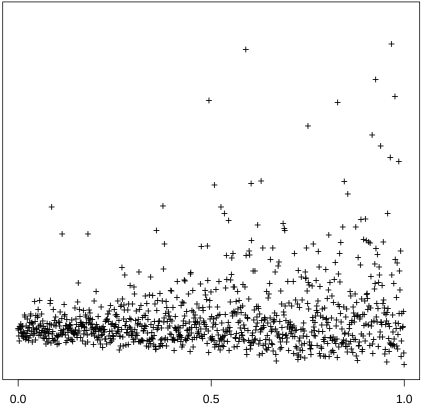
- **(S.2)** Fast convergence of α to zero:

$$\alpha \to 0 \text{ and } m_{n,t}\alpha \to c \in [1, \infty).$$

- **(S.3)** Very fast convergence of α to zero:

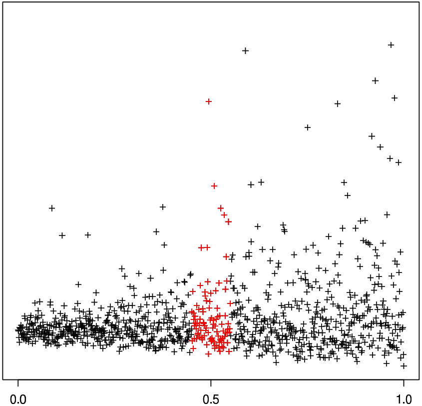
$$\alpha \to 0 \text{ and } m_{n,t}\alpha \to c \in [0, 1).$$

Example



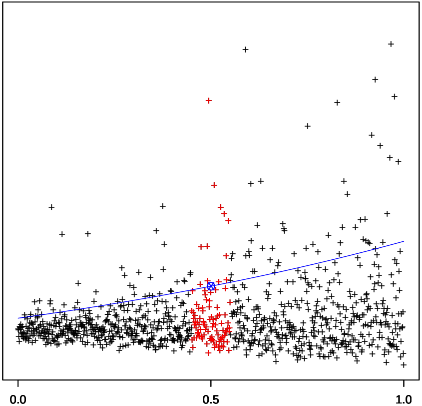
Using $n = 1000$ observations, we are interested in the estimation of the extreme quantile of order α at the point $t = 0.5$ ($E = [0, 1]$).

Example



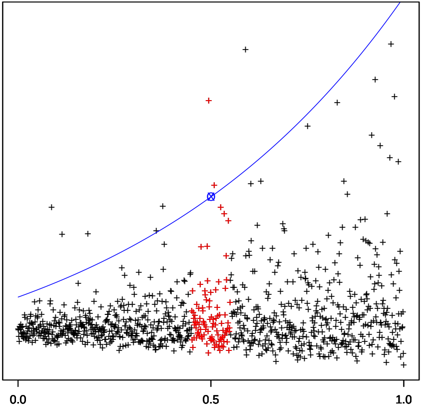
We select the observations in the slice S_t with $h_{n,t} = 0.05$ $(m_{n,t} = 100)$.

Example



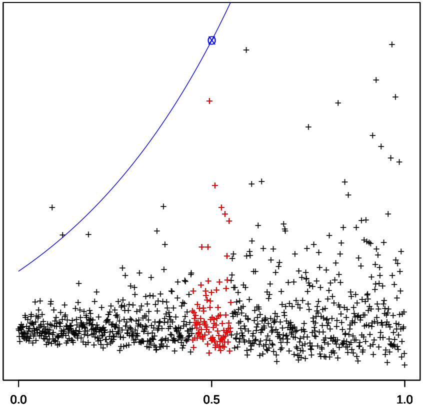
(S.1): Theoretical value of the quantile of order $\alpha = 10/100 = 0.1$

Example



(S.2): Theoretical value of the quantile of order $\alpha = 1/100 = 0.01$

Example



(S.3): Theoretical value of the quantile of order $\alpha = 0.1/100 = 0.001$

Situation **(S.1)**: The conditional quantile is in the range of the observations. We use the estimator:

$$\hat{q}_1(\alpha, t) = Z_{m_{n,t} - \lfloor m_{n,t}\alpha \rfloor + 1, m_{n,t}}(t).$$

Asymptotic distribution

If α satisfies **(S.1)**, under some assumptions on the conditional distribution,

$$(m_{n,t}\alpha)^{1/2} \left(\frac{\hat{q}_1(\alpha, t)}{q(\alpha, t)} - 1 \right) \xrightarrow{d} \mathcal{N}(0, \gamma^2(t))$$

Situation **(S.2)**: The conditional quantile is located near the boundary of the sample but still in the range of the data. We can also use the estimator $\hat{q}_1(\alpha, t)$.

Asymptotic distribution

If α satisfies **(S.2)**, under some assumptions on the conditional distribution,

$$\left(\frac{\hat{q}_1(\alpha, t)}{q(\alpha, t)} - 1\right) \xrightarrow{d} \mathcal{E}(c, \gamma(t)),$$

where $\mathcal{E}(c, \gamma(t))$ is a non-degenerated distribution (but not Gaussian !!)

Comments:

- The asymptotic distribution is not Gaussian and its expression is quite complicated.
- In this situation, estimator $\hat{q}_1$ is not consistent.

Situation **(S.3)**: The conditional quantile is beyond the range of the observations. Thus, we can not use the estimator $\hat{q}_1$. We propose to use the estimator:

$$\hat{q}_2(\alpha, t) = \hat{q}_1(\beta, t) \left(\frac{\beta}{\alpha}\right)^{\hat{\gamma}_n(t)},$$

where β satisfies **(S.1)** and $\hat{\gamma}_n(t)$ is a point wise estimator of the conditional tail index.

Situation **(S.3)**: The conditional quantile is beyond the range of the observations. Thus, we can not use the estimator $\hat{q}_1$. We propose to use the estimator:

$$\hat{q}_2(\alpha, t) = \hat{q}_1(\beta, t) \left(\frac{\beta}{\alpha} \right)^{\hat{\gamma}_n(t)},$$

where β satisfies **(S.1)** and $\hat{\gamma}_n(t)$ is a point wise estimator of the conditional tail index.
Estimator $\hat{q}_2(\alpha, t)$ can be decomposed in two parts:

- An estimator of a conditional quantile of order β satisfying **(S.1)** (*i.e.* an order statistics)

Situation **(S.3)**: The conditional quantile is beyond the range of the observations. Thus, we can not use the estimator $\hat{q}_1$. We propose to use the estimator:

$$\hat{q}_2(\alpha, t) = \hat{q}_1(\beta, t) \left(\frac{\beta}{\alpha} \right)^{\hat{\gamma}_n(t)},$$

where β satisfies **(S.1)** and $\hat{\gamma}_n(t)$ is a point wise estimator of the conditional tail index.
Estimator $\hat{q}_2(\alpha, t)$ can be decomposed in two parts:

- An estimator of a conditional quantile of order β satisfying **(S.1)** (*i.e.* an order statistics)
- An extrapolation term depending on $\hat{\gamma}_(t)$

Asymptotic distribution

If β satisfies **(S.1)**, if there exists a positive sequence $\upsilon_n(t)$ and a distribution $\mathcal{D}$ such that $\upsilon_n(t)(\hat{\gamma}_n(t) - \gamma(t)) \xrightarrow{d} \mathcal{D}$, then, under some assumptions on the conditional distribution, two situations arise:

i) The asymptotic distribution is driven by $\hat{q}_1(\beta, t)$ and then

$$(m_{n,t}\alpha)^{1/2} \left(\frac{\hat{q}_2(\alpha, t)}{q(\alpha, t)} - 1 \right) \xrightarrow{d} \mathcal{N}(0, \gamma^2(t)).$$

ii) The asymptotic distribution is driven by $\hat{\gamma}_n(t)$ and then

$$\frac{\upsilon_n(t)}{\log(\beta/\alpha)} \left(\frac{\hat{q}_2(\alpha, t)}{q(\alpha, t)} - 1 \right) \xrightarrow{d} \mathcal{D}.$$

- Estimator $\hat{q}_2$ can be used in the three situations.
- For the conditional tail index estimator, we can use for instance the Hill type estimator proposed by L. Gardes & S. Girard (2008):

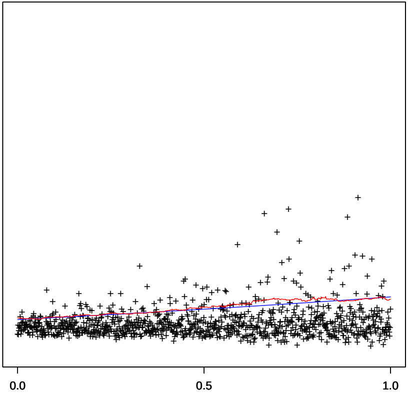
$$\hat{\gamma}_n^H(t) = \frac{1}{k_{n,t}} \sum_{i=1}^{k_{n,t}} i \log \left(\frac{Z_{m_{n,t}-i+1,m_{n,t}}}{Z_{m_{n,t}-i,m_{n,t}}} \right),$$

where $k_{n,t} = m_{n,t}\beta$.

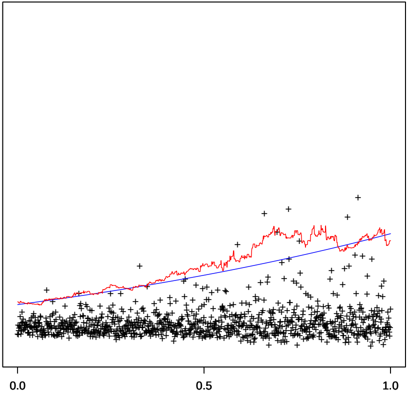
- We generate $n = 1000$ observations $\{(Y_i, x_i), i = 1, \ldots, n\}$ under the model: $x \in E = [0, 1]$, and the conditional extreme quantile of Y given x is defined by:

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

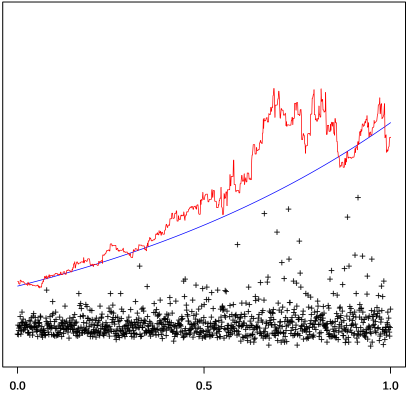
- To estimate $q(\alpha, t)$ we use the estimator $\hat{q}_2(\alpha, t) = \hat{q}_1(\beta, t)(\beta/\alpha)^{\hat{\gamma}_n^H(t)}$ where $\hat{\gamma}_n^H(t)$ is the conditional Hill type estimator, the observations are selected using a ball of radius $h_{n,t} = 0.1$ for all $t \in E$ (*i.e.* $m_{n,t} = 200$) and $\beta = 0.3$.



Estimation of the function $q(\alpha, .)$ with $\alpha = 20/200$ (**(S.1)**).



Estimation of the function $q(\alpha, .)$ with $\alpha = 2/200$ (**(S.2)**).



Estimation of the function $q(\alpha, .)$ with $\alpha = 0.2/200$ ((**S.3**)).

References

- L. Gardes & S. Girard. A moving window approach for nonparametric estimation of the conditional tail index, Journal of Multivariate Analysis, 99, 2368–2388, 2008
- L. Gardes, S. Girard & A. Lekina. Functional nonparametric estimation of conditional extreme quantiles, Journal of Multivariate Analysis, 101, 419–433, 2010.
- L. Gardes & S. Girard. Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels, Extremes, 13(2), 177–204, 2010.
- A. Daouia, L. Gardes, S. Girard & A. Lekina. Kernel estimators of extreme level curves, Test, 20(2), 311–333, 2011.
- L. Gardes & S. Girard Functional kernel estimators of conditional extreme quantiles, In F. Ferraty, editor, Recent advances in functional data analysis and related topics, pages 135-140, Springer, Physica-Verlag, 2011.

References

- J. Carreau & S. Girard. Spatial extreme quantile estimation using a weighted log-likelihood approach, Journal de la Société Française de Statistique, 152(3), 66–83 , 2011.
- A. Daouia, L. Gardes & S. Girard Nadaraya's estimates for large quantiles and free disposal support curves, In I. Van Keilegom and P. Wilson, editors, Exploring research frontiers in contemporary statistics and econometrics, pages 1–22, Springer, 2012.
- L. Gardes & S. Girard. Functional kernel estimators of large conditional quantile, Electronic Journal of Statistics, 6, 1715–1744, 2012.
- J. Carreau, D. Ceresetti, E. Ursu, S. Anquetin, J.D. Creutin, L. Gardes, S. Girard & G. Molinié. Evaluation of classical spatial-analysis schemes of extreme rainfall, Natural Hazards and Earth System Sciences, 12, 3229–3240, 2012.
- A. Daouia, L. Gardes & S. Girard. On kernel smoothing for extremal quantile regression, Bernoulli, 19, 2557–2589, 2013.