Estimation methods for conditional extremes

Stéphane Girard

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

December, 2008

Joint work with Laurent Gardes and Alexandre Lekina

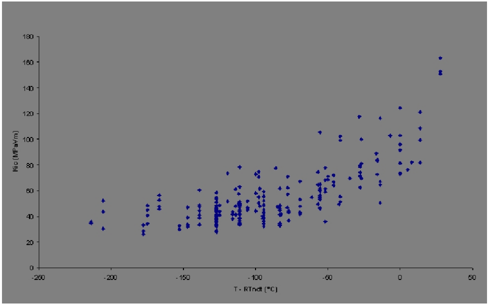
The problem

- Estimation of the tail index and extreme quantiles associated to a random variable Y.
- A covariate information x is recorded simultaneously with Y.
- The tail heaviness of Y given x depends on x, and thus the tail index is a conditional tail index $\gamma(x)$.
- Similarly, extreme quantiles are conditional extreme quantiles defined by

$$P(Y > q(\alpha, x)|x) = \alpha,$$

with order $\alpha \to 0$.

Illustration: Reliability of nuclear power plants



Y is the tenacity, X is the temperature.

Our approach

To combine nonparametric smoothing techniques with extreme-value methods in order to obtain efficient estimators of $\gamma(x)$ and $q(\alpha, x)$.

- Few assumptions are made on the nature of the covariate (asymptotic distributions are established for the proposed estimators, without assuming that x is finite dimensional).
- The estimators are easy to compute since they are closed-form and thus does not require optimization procedures.

Most recent related work. See for instance Beirlant *et al.* (2004) and Chavez *et al.* (2005).

Main assumptions

- The covariate x belongs to a metric space E associated to a distance d. It can be finite or infinite dimensional.
- The conditional distribution of Y given $x \in E$ is a heavy tailed distribution *i.e.* for $\lambda > 0$,

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

i.e. the conditional extreme quantile $q(., x)$ increases to ∞ at a polynomial rate as $\alpha \to 0$.

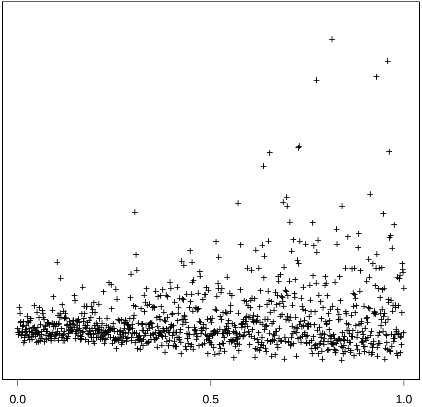
Selection of the observations

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

Selection of the observations

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

Selection of the observations

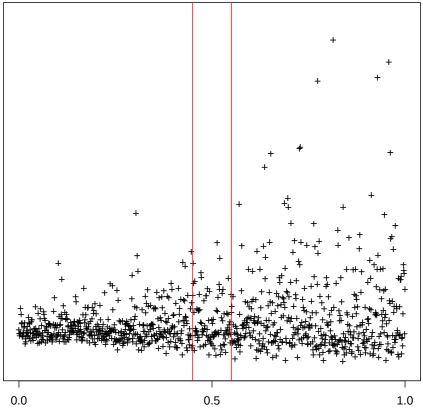


Example: Estimation at $t = 0.5$ using $n = 1000$ observations $(Y_i, x_i), i = 1, \ldots, n$ for $E = [0, 1]$.

Selection of the observations

- Let $\{(Y_i, x_i), i = 1, \ldots, n\}$ be n independent copies of (Y, x). Using these observations, our aim is to estimate $\gamma(t)$ and $q(\alpha, t)$ for a given $t \in E$.
- To this end, only the observations "close" to t are used.
- Let h be a positive sequence tending to zero as $n \to \infty$. Denote by $B(t, h)$ the ball of center t with radius h and by S_t the slice defined by $S_t = (0, \infty) \times B(t, h)$.

Selection of the observations

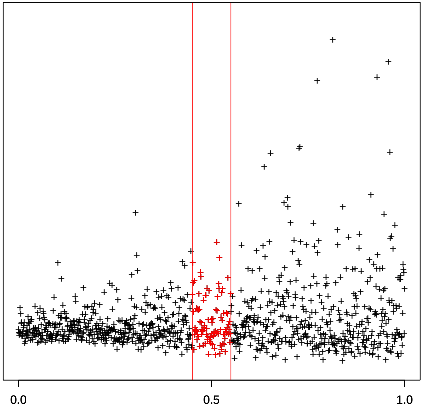


Example: Estimation at $t = 0.5$ using $n = 1000$ observations $(Y_i, x_i), i = 1, \ldots, n$ for $E = [0, 1]$.

Selection of the observations

- Let $\{(Y_i, x_i), i = 1, \ldots, n\}$ be n independent copies of (Y, x). Using these observations, our aim is to estimate $\gamma(t)$ and $q(\alpha, t)$ for a given $t \in E$.
- To this end, only the observations "close" to t are used.
- Let h be a positive sequence tending to zero as $n \to \infty$. Denote by $B(t, h)$ the ball of center t with radius h and by S_t the slice defined by $S_t = (0, \infty) \times B(t, h)$.
- The observations Y_i's for which $x_i \in B(t, h)$ are selected.

Selection of the observations



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

Selection of the observations

- Let $\{(Y_i, x_i), i = 1, \ldots, n\}$ be n independent copies of (Y, x). Using these observations, our aim is to estimate $\gamma(t)$ and $q(\alpha, t)$ for a given $t \in E$.
- To this end, only the observations "close" to t are used.
- Let h be a positive sequence tending to zero as $n \to \infty$. Denote by $B(t, h)$ the ball of center t with radius h and by S_t the slice defined by $S_t = (0, \infty) \times B(t, h)$.
- The observations Y_i's for which $x_i \in B(t, h)$ are selected.
- They are denoted $\{Z_i(t), i = 1, \ldots, m\}$ where m is the number of x_i's in the ball $B(t, h)$.
- The associated order statistics are denoted by $Z_{1,m}(t) \leq \ldots \leq Z_{m,m}(t)$.

Principle

A weighted sum of the rescaled log-spacings between the largest selected observations:

$$\hat{\gamma}_n(t, W) = \sum_{i=1}^{k} i \log\left(\frac{Z_{m-i+1,m}(t)}{Z_{m-i,m}(t)}\right) W\left(i/k, t\right) \Bigg/ \sum_{i=1}^{k} W\left(i/k, t\right),$$

with the following notations:

- Intermediate sequence: $k \to \infty$ and $k/m \to 0$,
- Weights: $W(., t)$ a function defined on $(0, 1)$ and integrating to one.

Further assumptions

- Lipschitz assumptions: There exists positive constants z_ℓ, c_ℓ, c_γ and $\alpha \leq 1$ such that for all $x \in B(t,1)$,

$$|\gamma(x) - \gamma(t)| \leq c_\gamma d^\alpha(x,t),$$

and

$$\sup_{z > z_\ell} \left| \log \left(\frac{\ell(z,x)}{\ell(z,t)} \right) \right| \leq c_\ell d(x,t),$$

- Second order condition: There exists a negative function $\rho(t)$ and a rate function $b(.,t)$ satisfying $b(y,t) \to 0$ as $y \to \infty$, such that for all $\lambda \geq 1$,

$$\log \left(\frac{\ell(\lambda y, t)}{\ell(y,t)} \right) = b(y,t) \frac{1}{\rho(t)} (\lambda^{\rho(t)} - 1)(1 + o(1)),$$

Asymptotic normality

Theorem

If, moreover, $k^{1/2} b(m/k, t) \to \lambda(t) \in \mathbb{R}$ and $k^{1/2} h^{\alpha} \to 0$ then

$$k^{1/2}\left(\hat{\gamma}_n(t, W) - \gamma(t) - b(m/k, t)\mathcal{AB}\right) \xrightarrow{d} \mathcal{N}\left(0, \gamma^2(t)\mathcal{AV}\right),$$

where we have defined

$$\mathcal{AB} = \int_0^1 W(s, t) s^{-\rho(t)} ds \text{ and } \mathcal{AV} = \int_0^1 W^2(s, t) ds.$$

Remarks

- The asymptotic bias involves two parts:
  - $b(m/k,t)$ which depends on the original distribution itself,
  - $\mathcal{AB}$ which can be made small by an appropriate choice of the weighting function W.
- Similarly, the asymptotic variance involves two parts:
  - $1/k$ which is inversely proportional to the number of observations used to build the estimator,
  - $\gamma^2(t)\mathcal{AV}$ which can also be adjusted.
- When $\lambda(t) \neq 0$, condition $k^{1/2}b(m/k,t) \to \lambda(t)$ forces the bias to be of the same order as the standard-deviation.
- Condition $k^{1/2}h^{\alpha} \to 0$ imposes to the fluctuations of $t \mapsto \gamma(t)$ to be negligible compared to the standard deviation of the estimator.

Two classical examples of weights

- Conditional Hill estimator: Constant weight functions $W^{\mathrm{H}}(s,t) = 1$ yield

$$\hat{\gamma}_n(t, W^{\mathrm{H}}) = \frac{1}{k} \sum_{i=1}^{k} i \log \left(\frac{Z_{m-i+1,m}(t)}{Z_{m-i,m}(t)} \right),$$

  which is formally the same expression as in Hill (1975). $\mathcal{AB} = 1/(1-\rho(t))$ and $\mathcal{AV} = 1$.

- Conditional Zipf estimator: Logarithmic weight functions $W^{\mathrm{Z}}(s,t) = -\log(s)$ yield an estimator $\hat{\gamma}_n(t, W^{\mathrm{Z}})$ similar to the Zipf estimator proposed by Kratz *et al.* (1996) and Schultze *et al.* (1996). $\mathcal{AB} = 1/(1-\rho(t))^2$ and $\mathcal{AV} = 2$.

Theoretical of weights

- Minimum variance estimator. The conditional Hill estimator is the unique minimum variance estimator in our family.
- Asymptotically unbiased estimator with minimum variance. The weight function associated to the unique asymptotically unbiased estimator with minimum variance is

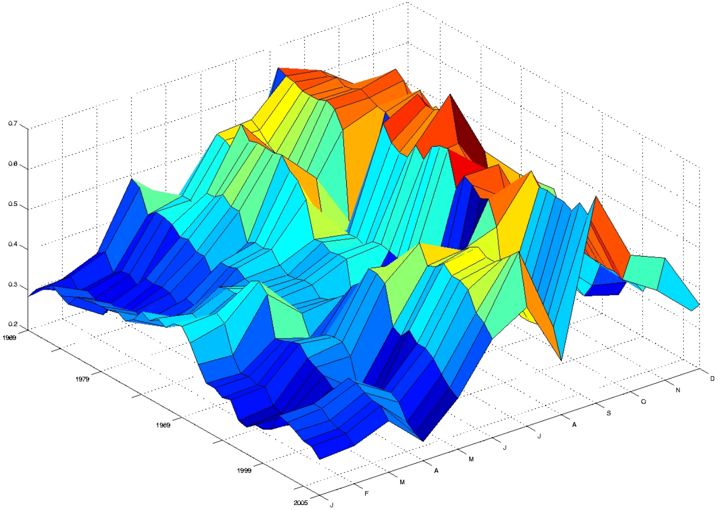
$$W^{\mathrm{opt}}(s,t) = \frac{\rho(t)-1}{\rho^2(t)} \left(\rho(t) - 1 + (1-2\rho(t)) s^{-\rho(t)} \right),$$

$\mathcal{AB} = 0$ and $\mathcal{AV} = (1 - 1/\rho(t))^2$.

Weights W^{HZ} and W^{opt} require the knowledge of the second order parameter $\rho(t)$.

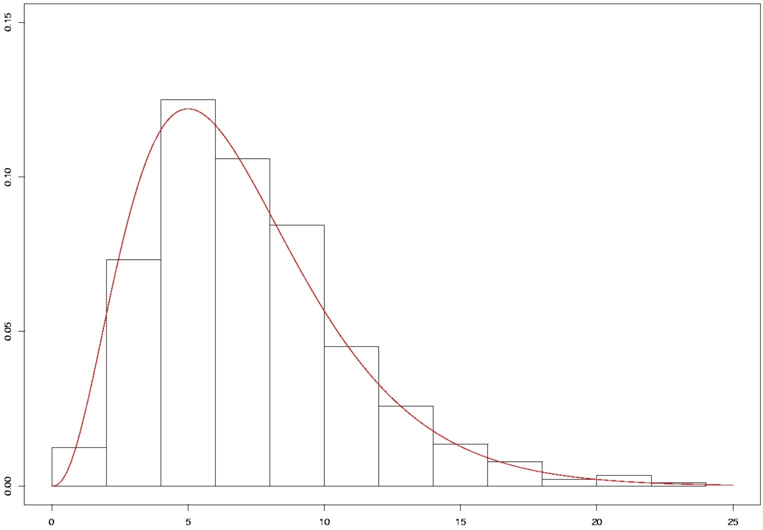
Illustration on real data

- $n = 13,505$ daily mean discharges (in m^3/s) of the Chelmer river collected by the Springfield gauging station, from 1969 to 2005.
- The data are provided by the Centre for Ecology and Hydrology (United Kingdom) and are available at `http://www.ceh.ac.uk/data/nrfa`.
- Y is the daily flow of the river,
- $x = (x_1, x_2)$ is a bi-dimensional covariate such that $x_1 \in \{1969, 1970, \dots, 2005\}$ is the year of measurement and $x_2 \in \{1, 2, \dots, 365\}$ is the day.
- The tail index $\gamma(t)$ is estimated on a grid: *i.e.* for $t \in \{1969, 1970, \dots, 2005\} \times \{15, 45, \dots, 345\}$.


0.7
0.6
0.5
0.4
0.3
0.2
1969
1979
1989
1999
2005
J
F
M
A
M
J
J
A
S
O
N
D

Implementation details

- h and k are assumed to be independent of t, they are selected by minimizing some distance between conditional Hill and Zipf estimators.
- The selected value of h corresponds to a smoothing over 4 years on x_1 and 2 months on x_2. Each ball $B(t, h)$ of the grid contains $m = 1089$ points and $k = 54$ rescaled log-spacings are used.
- The heuristics is validated by computing on each ball of the grid the χ^2 distance to the Exp(1) distribution. The histogram of these distances is superimposed to the theoretical density of the corresponding χ^2 distribution.

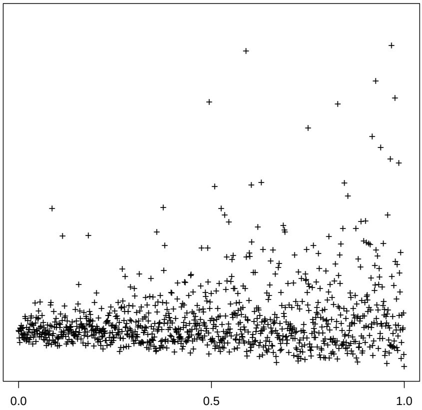

0.15
0.10
0.05
0.00
0
5
10
15
20
25

Discussion on the order of the extreme quantile

Three situations are considered, depending on the convergence rate of α to zero:

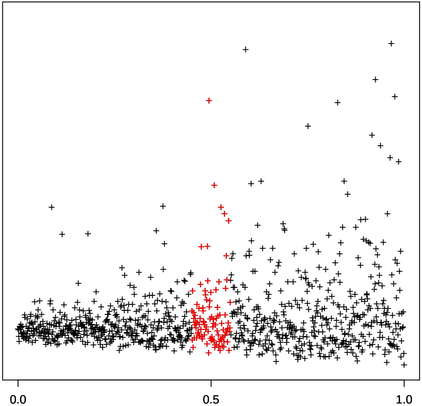
- **(S.1)** $\alpha \to 0$ and $m\alpha \to \infty$: The extreme quantile is located in the range of the data.
- **(S.2)** $\alpha \to 0$ and $m\alpha \to c \in [1, \infty)$: The extreme quantile is located near the boundary of the data.
- **(S.3)** $\alpha \to 0$ and $m\alpha \to c \in [0, 1)$: The extreme quantile is located beyond the boundary of the data.

Illustration



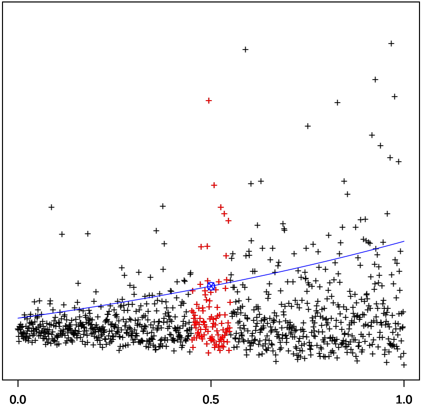
Using $n = 1000$ observations, estimation of the extreme quantile of order α at $t = 0.5$ ($E = [0, 1]$).

Illustration



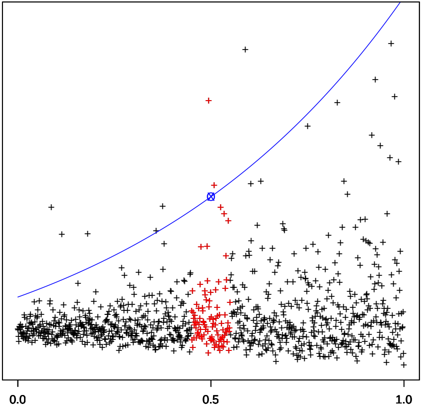
Selected observations in the slice S_t with $h = 0.05$. ($m = 100$).

Illustration



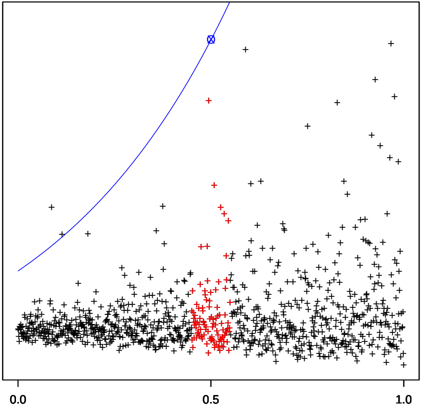
(S.1): Theoretical quantile of order $\alpha = 10/100 = 0.1$

Illustration



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

Illustration



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

Situation **(S.1)**

Since the conditional quantile is in the range of the observations, we can define the estimator:

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

Asymptotic distribution

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

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

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.

The asymptotic distribution is no more Gaussian and its expression is quite complicated. Note that, in this situation, $\hat{q}_1(\alpha, t)/q(\alpha, t)$ does not converge to one in probability.

Situation **(S.3)**

The conditional quantile is beyond the range of the observations. Thus, $\hat{q}_1$ cannot be used anymore. We propose a conditional Weissman estimator (Weissman, 1978):

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

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

Situation **(S.3)**

The conditional quantile is beyond the range of the observations. Thus, $\hat{q}_1$ cannot be used anymore. We propose a conditional Weissman estimator (Weissman, 1978):

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

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

Estimator $\hat{q}_2(\alpha, t)$ can be expanded into 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, $\hat{q}_1$ cannot be used anymore. We propose a conditional Weissman estimator (Weissman, 1978):

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

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

Estimator $\hat{q}_2(\alpha, t)$ can be expanded into two parts:

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

Situation **(S.3)**

Asymptotic distribution

If β satisfies **(S.1)** and there exist a positive sequence $\upsilon_n(t)$ and a distribution $\mathcal{D}$ such that $\upsilon_n(t)(\hat{\gamma}_n(t) - \gamma(t)) \stackrel{d}{\to} \mathcal{D}$, 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

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

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

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

Illustration on simulations

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

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

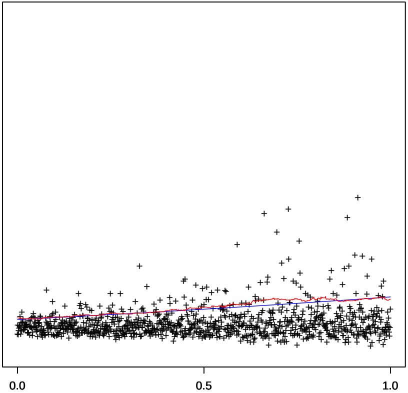
- The following estimator is considered:

$$\hat{q}_2(\alpha, t) = \hat{q}_1(\alpha, t)(\beta/\alpha)^{\hat{\gamma}_n(t, W^{\mathrm{H}})}$$

where $\hat{\gamma}_n(t, W^{\mathrm{H}})$ is the conditional Hill estimator.

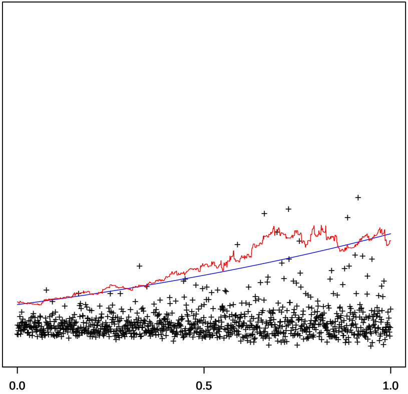
- We choose $h = 0.1$, $m = 200$ and $\beta = 60/200$.

Results



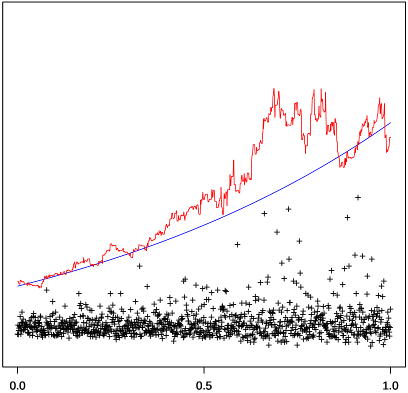
Estimation of $q(\alpha, .)$ with $\alpha = 20/200$ (situation **(S.1)**).

Results



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

Results



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

Bibliography

Non-conditional extremes

- Hill, B.M. (1975) A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, **3**, 1163–1174.
- Kratz, M. and Resnick, S. (1996) The QQ-estimator and heavy tails. *Stochastic Models*, **12**,699–724.
- Schultze, J. and Steinebach, J. (1996) On least squares estimates of an exponential tail coefficient. *Statistics and Decisions*, **14**, 353–372.
- Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *Journal of the American Statistical Association*, **73**, 812–815.

Bibliography

Conditional extremes

- Beirlant, J. and Goegebeur, Y. (2004) Local polynomial maximum likelihood estimation for Pareto-type distributions. *Journal of Multivariate Analysis*, **89**, 97–118.
- Chavez-Demoulin, V. and Davison, A.C. (2005) Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society, series C.*, **54**, 207–222.
- Gardes, L. and Girard, S. (2008) A moving window approach for nonparametric estimation of the conditional tail index. *Journal of Multivariate Analysis*, **99**, 2368–2388.
- Gardes, L., Girard, S. and Lekina, A. (2008) Functional nonparametric estimation of conditional extreme quantiles. *Technical report*, `http://hal.archives-ouvertes.fr/hal-00289996/fr/`