NONPARAMETRIC ESTIMATION OF THE CONDITIONAL TAIL INDEX

Laurent Gardes and Stéphane Girard

INRIA Rhône-Alpes, Team Mistis,
655 avenue de l'Europe, Montbonnot,
38334 Saint-Ismier Cedex, France.
E-mail: {Laurent.Gardes, Stephane.Girard}@inrialpes.fr

Team Mistis: http://mistis.inrialpes.fr/

1. Introduction

The problem.

- Estimation of the tail index γ associated to a random variable Y.
- Some 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 function $\gamma(x)$ of the covariate.

Our approach: To combine nonparametric smoothing techniques with extreme-value methods in order to obtain efficient estimators of $\gamma(x)$.

- Few assumptions are made on the regularity of $\gamma(x)$ and on the nature of the covariate (a central limit theorem is established for the proposed estimator, without assuming that x is finite dimensional).
- The estimator is easy to compute since it is 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).

2. Estimators of the conditional tail index

Framework. E a metric space associated to a metric d.

- **Model**: Conditional tail quantile function of Y given $t \in E$ is, for all $y > 0$,

$$U(y,t) = \inf\{s; F(s,t) \geq 1 - 1/y\} = y^{\gamma(t)}\ell(y,t), \tag{1}$$

  where

  - $\gamma(t)$ is an unknown positive function of the covariate t and,
  - for t fixed, $\ell(.,t)$ is a slowly-varying function, *i.e.* for $\lambda > 0$,

$$\lim_{y\to\infty} \frac{\ell(\lambda y,t)}{\ell(y,t)} = 1.$$

- **Data**: A sample $(Y_1, x_1), \ldots, (Y_n, x_n)$ iid from (1), where the design points $x_1, \ldots, x_n$ are non random points in E.

Goal. For a given $t \in E$, estimate the conditional tail index $\gamma(t)$.

Nonparametric estimators

- **Window width**: $h_{n,t}$ a positive sequence tending to zero as $n \to \infty$,
- **Window**: Ball $B(t, h_{n,t}) = \{x \in E,\ d(x,t) \le h_{n,t}\}$,
- **Selected observations**: $\{Z_i(t),\ i = 1, \dots, m_{n,t}\}$ the response variables $Y_i's$ associated to the $m_{n,t}$ covariates $x_i's$ in the ball $B(t, h_{n,t})$.
- **Corresponding order statistics**: $Z_{1,m_{n,t}}(t) \le \dots \le Z_{m_{n,t},m_{n,t}}(t)$,
- **Intermediate sequence**: $k_{n,t} \to \infty$ and $k_{n,t}/m_{n,t} \to 0$,
- **Weights:** $W(.,t)$ a function defined on $(0,1)$ such that $\int_0^1 W(s,t)ds = 1$,
- **Moving-window estimators:** A weighted sum of the rescaled log-spacings between the largest selected observations:

$$\hat{\gamma}_n(t,W) = \sum_{i=1}^{k_{n,t}} i \log \left(\frac{Z_{m_{n,t}-i+1,m_{n,t}}(t)}{Z_{m_{n,t}-i,m_{n,t}}(t)} \right) W\left(i/k_{n,t}, t\right) \Big/ \sum_{i=1}^{k_{n,t}} W\left(i/k_{n,t}, t\right). \tag{2}$$

3. Main results

Assumptions on the conditional distribution

- **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 $\lambda \geq 1$,

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

where "o" is uniform in $\lambda \geq 1$ as $y \to \infty$.

Assumptions on the weights

- **Beirlant** *et al* **assumption**: (See Beirlant *et al.* (2002)).
- **Integrability condition**: There exists a constant $\delta > 0$ such that

$$\int_0^1 |W(s,t)|^{2+\delta} ds < \infty.$$

Asymptotic normality

Theorem 1 *If, moreover,* $k_{n,t}^{1/2} b_{n,t} \to \lambda(t) \in \mathbb{R}$ *and* $k_{n,t}^{1/2} h_{n,t}^{\alpha} \to 0$ *then*

$$k_{n,t}^{1/2}\left(\hat{\gamma}_n(t,a,\lambda) - \gamma(t) - \Delta\left(\frac{m_{n,t}}{k_{n,t}}, t\right)\mathcal{AB}(a,\lambda,\rho(t))\right) \quad (3)$$

where we have defined

$$b_{n,t} = b\left(\frac{m_{n,t}}{k_{n,t}}, t\right),$$

$$\mathcal{AB}(a,\lambda,\rho(t)) = (1-\lambda\rho(t))^{-a} \text{ and } \mathcal{AV}(a,\lambda) = \frac{\Gamma(2a-1)}{\lambda\Gamma^2(a)}(2-\lambda)^{1-2a}.$$

Remark 1.

- The asymptotic bias involves two parts:
  - $b_{n,t}$ which depends on the original distribution itself,
  - $\mathcal{AB}(t, W)$ which can be made small by an appropriate choice of the weighting function W.
- Similarly, the asymptotic variance involves two parts:
  - $1/k_{n,t}$ which is inversely proportional to the number of observations used to build the estimator,
  - $\gamma^2(t)\mathcal{AV}(t, W)$ which can also be adjusted.

Remark 2.

- When $\lambda(t) \neq 0$, condition $k_{n,t}^{1/2} b_{n,t} \to \lambda(t)$ forces the bias to be of the same order as the standard-deviation.
- Condition $k_{n,t}^{1/2} h_{n,t}^{\alpha} \to 0$ is due to the functional nature of the tail index to estimate. It imposes to the fluctuations of $t \to \gamma(t)$ to be negligible compared to the standard deviation of the estimator.

Corollary 1 *Suppose that* $E = \mathbb{R}^p$ *and that the slowly-varying function* ℓ *in (1) is such that* $\ell(y,t) = 1$ *for all* $(y,t) \in \mathbb{R}_+ \times \mathbb{R}^p$. *If*

$$\liminf_{n\to\infty} \frac{m_{n,t}}{nh_{n,t}^p} > 0, \tag{4}$$

then the convergence in distribution (3) holds with rate $n^{\frac{\alpha}{p+2\alpha}}\eta_n$, *where* $\eta_n \to 0$ *arbitrarily slowly.*

- Condition (4) is an assumption on the multidimensional design and on the distance d.
- Under the condition on the slowly-varying function $\ell(y,t) = 1$ for all $(y,t) \in \mathbb{R}_+ \times \mathbb{R}^p$, estimating $\gamma(t)$ is a nonparametric regression problem since $\gamma(t) = \mathbb{E}(\log Y | X = t)$. Let us highlight that the convergence rate is, up to the η_n factor, the optimal convergence rate for estimating α-Lipschitzian regression function in $\mathbb{R}^p$, see Stone (1982).

4. Two classical examples of weights

Conditional Hill estimator: Constant weight functions $W^{\text{H}}(s,t) = 1$ for all $s \in [0,1]$ yield

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

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

Conditional Zipf estimator: Considering in (2) the weight function $W^{\text{Z}}(s,t) = -\log(s)$ for all $s \in [0,1]$ yields an estimator $\hat{\gamma}_n(t, W^{\text{Z}})$ similar to the Zipf estimator proposed by Kratz *et al.* (1996) and Schultze *et al.* (1996). Convergence in distribution (3) holds with $\mathcal{AB}(t, W^{\text{Z}}) = 1/(1-\rho(t))^2$ and $\mathcal{AV}(t, W^{\text{Z}}) = 2$.

5. Theoretical choices of weights

Asymptotically unbiased estimators.

- Starting with two weight functions W_1 and W_2, it is possible to build a third one

$$W_{12} = \frac{\mathcal{AB}(t, W_2) W_1 - \mathcal{AB}(t, W_1) W_2}{\mathcal{AB}(t, W_2) - \mathcal{AB}(t, W_1)}$$

  such that $\mathcal{AB}(t, W_{12}) = 0$.

- Applying this principle to the conditional Hill and Zipf estimators yields

$$W^{\text{HZ}}(s,t) = \frac{1}{\rho(t)} - \left(1 - \frac{1}{\rho(t)}\right) \log(s).$$

  Convergence in distribution (3) holds with $\mathcal{AB}(t, W^{\text{HZ}}) = 0$ and $\mathcal{AV}(t, W^{\text{HZ}}) = 1 + (1 - 1/\rho(t))^2$.

Minimum variance estimator. The conditional Hill estimator is the unique minimum variance estimator in (2).

Asymptotically unbiased estimator with minimum variance. The weight function associated to the unique asymptotically unbiased estimator with minimum variance is

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

Convergence in distribution (3) holds with $\mathcal{AB}(t, W^{\text{opt}}) = 0$ and $\mathcal{AV}(t, W^{\text{opt}}) = (1 - 1/\rho(t))^2$.

Remark 3.

- Weights W^{HZ} and W^{opt} require the knowledge of the second order parameter $\rho(t)$.
- The estimation of the function $t \longrightarrow \rho(t)$ is not addressed here. See for instance Alves *et al.* (2003) for estimators when there is no covariate information. See also Gardes *et al.* (2007) for an illustration of the effect of using a arbitrary chosen value.

6. Illustration on real data

Description of the 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 $x_1 \in \{1969, 1970, \ldots, 2005\}$ is the year of measurement and $x_2 \in \{1, 2, \ldots, 365\}$ is the day.

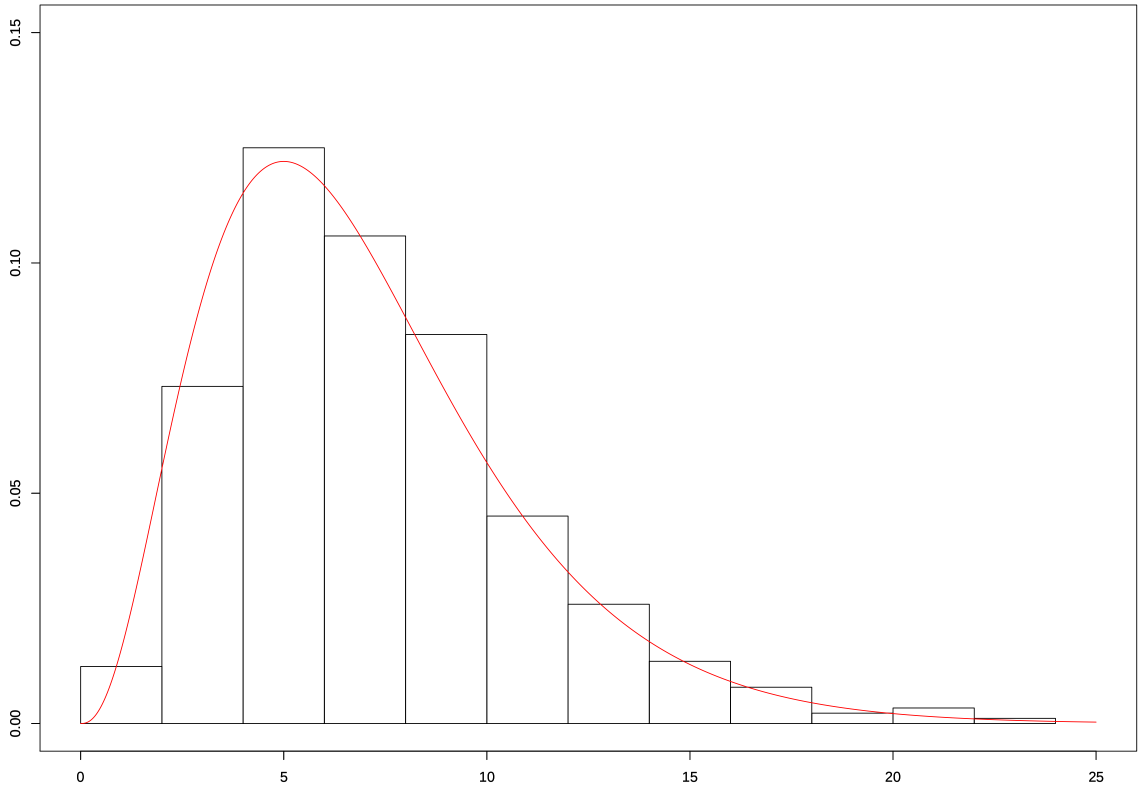
Selection of the hyperparameters.

- $h_{n,t}$ and $k_{n,t}$ are assumed to be independent of t, they are thus denoted by h_n and k_n respectively.
- They are selected by minimizing the following distance between conditional Hill and Zipf estimators:

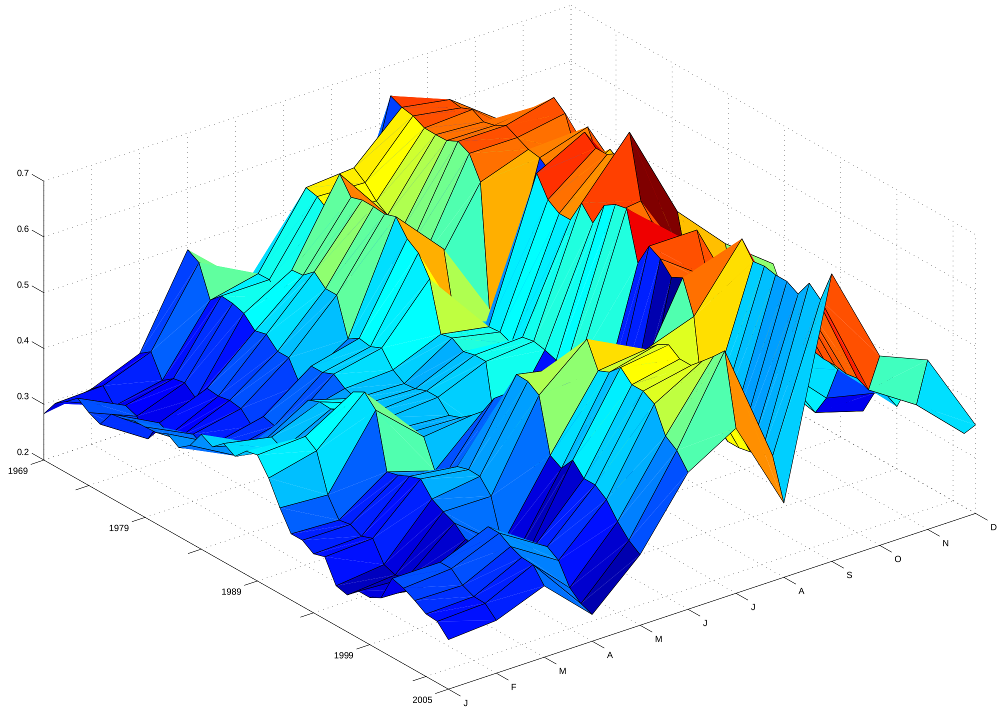
$$\min_{h_n,k_n} \max_{t \in T} \left| \hat{\gamma}_n(t, W^{\text{H}}) - \hat{\gamma}_n(t, W^{\text{Z}}) \right| ,$$

  where $T = \{1969, 1970, \ldots, 2005\} \times \{15, 45, \ldots, 345\}$.
- This heuristics is commonly used in functional estimation and relies on the idea that, for a properly chosen pair (h_n, k_n) we have $\hat{\gamma}_n(t, W^{\text{H}}) \simeq \hat{\gamma}_n(t, W^{\text{Z}})$.
- The selected value of h_n corresponds to a smoothing over 4 years on x_1 and 2 months on x_2. Each ball $B(t, h_n)$, $t \in T$ contains $m_n = 1089$ points and $k_n = 54$ rescaled log-spacings are used.

The heuristics is validated by computing on each ball $B(t, h_n)$, $t \in T$ the χ^2 distance to the standard exponential distribution. The histogram of these distances is superimposed to the theoretical density of the corresponding χ^2 distribution.



Conditional Zipf estimator



The results are located in the interval $[0.2, 0.7]$. The estimated tail index is almost independent of the year but dependent of the day. Heaviest tails are obtained in September: Extreme flows are more likely this month.

Bibliography

Alves, M.I.F., Gomes, M.I. and de Haan, L. (2003) A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica*, **60**, 193–214.

Beirlant, J., Dierckx, G., Guillou, A. and Stărică, C. (2002) On exponential representations of log-spacings of extreme order statistics. *Extremes*, **5**, 157–180.

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.

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.

Stone, C. (1982) Optimal global rates of convergence for nonparametric regression. *Annals of Statistics*, **10**, 1040–1053.

Bibliography

- L. Gardes, S. Girard and A. Lekina. Functional nonparametric estimation of conditional extreme quantiles, *Journal of Multivariate Analysis*, **101**, 419–433, 2010.
- L. Gardes and S. Girard. Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels, *Extremes*, **13**, 177–204, 2010.
- A. Daouia, L. Gardes, S. Girard and A. Lekina. Kernel estimators of extreme level curves, *Test*, **20**, 311–333, 2011.
- L. Gardes and S. Girard. Functional kernel estimators of large conditional quantiles, *Electronic Journal of Statistics*, **6**, 1715–1744, 2012.
- A. Daouia, L. Gardes, and S. Girard. On kernel smoothing for extremal quantile regression, *Bernoulli*, **19**, 2557–2589, 2013.
- J. El Methni, L. Gardes and S. Girard. Nonparametric estimation of extreme risks from conditional heavy-tailed distributions, *Scandinavian Journal of Statistics*, to appear, 2014.