



A new location-scale model for conditional heavy-tailed distributions

Aboubacrène Ag Ahmad, Aliou Diop, El Hadji Deme, Stéphane Girard

► To cite this version:

Aboubacrène Ag Ahmad, Aliou Diop, El Hadji Deme, Stéphane Girard. A new location-scale model for conditional heavy-tailed distributions. IFSS 2018 - 2nd Italian-French Statistics Seminar, Sep 2018, Grenoble, France. hal-01942204

HAL Id: hal-01942204

<https://hal.archives-ouvertes.fr/hal-01942204>

Submitted on 10 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



A new location-scale model for conditional heavy-tailed distributions

Aboubacrène Ag Ahmad¹, Aliou Diop¹, El Hadji Deme² & Stéphane Girard³

¹ UGB - Université Gaston Berger de Saint-Louis, Sénegal.

² LERSTAD - Laboratoire d'Etudes et de Recherches en Statistiques et Développement, UGB Saint-Louis, Sénegal.

³ Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJL, 38000 Grenoble, France.

Bayesian learning theory for complex data modelling
2nd Italian-French Statistics Seminar - IFSS
Grenoble, 6-7 September 2018

1. Abstract

We are interested in a location-scale model for heavy-tailed distributions where the covariate is deterministic. We first address the nonparametric estimation of the location and scale functions and derive an estimator of the conditional extreme-value index. Second, new estimators of the extreme conditional quantiles are introduced. The asymptotic properties of the estimators are established under mild assumptions.

2. Model

$$Y = a(x) + b(x)Z,$$

where $x \in \mathbb{R}$ is a nonrandom covariate, $Y \in \mathbb{R}$ a random variable that depends on x and Z an independent random variable of x . The location function $a(\cdot)$ and the scale function $b(\cdot) > 0$ are unknown. We assume that the survival function of Z denoted by \bar{F}_Z belongs to the class of regularly varying functions at infinity

$$\bar{F}_Z(z) = z^{-1/\gamma} \ell(z), \quad \gamma > 0,$$

where ℓ is a slowly-varying function at infinity and γ is the conditional extreme-value index.

4. Asymptotic results

Assume n large enough so that $h = h_n < 1/2$, the following results are obtained under some regularity conditions on the probability density function of Z (f_Z) and \bar{F}_Z and a Lipschitzian condition on the kernel K .

Proposition 1 : Estimation of classical conditional quantiles

Let $(t_n)_n$ be a sequence in $[h_n, 1 - h_n]$ and $(\alpha_j)_{j=1,\dots,J}$ a strictly decreasing sequence in $(0, 1)$. If $nh_n \rightarrow +\infty$ and $nh_n^3 \rightarrow 0$ as $n \rightarrow +\infty$, then

$$\frac{\sqrt{nh_n}}{b(t_n)} \left[\hat{q}_{n,Y}(\alpha_j | t_n) - q_Y(\alpha_j | t_n) \right] \xrightarrow{\text{d}} \mathcal{N}(0_{\mathbb{R}^J}, \mathbf{K}^2(\mathbf{u}) \mathbf{d}\mathbf{u} \mathbf{A}),$$

where $A_{k,l} = \frac{\alpha_{k \vee l}(1 - \alpha_{k \wedge l})}{f_Z(q_Z(\alpha_k))f_Z(q_Z(\alpha_l))}$, for all $(k, l) \in \{1, \dots, J\}^2$ and \vee (resp. \wedge) denotes the maximum (resp. the minimum).

Proposition 2 : Estimation of location and scale functions

If $nh_n \rightarrow +\infty$ and $nh_n^3 \rightarrow 0$ as $n \rightarrow +\infty$, then for all sequence $(t_n)_n$ in $[h_n, 1 - h_n]$,

$$\frac{\sqrt{nh_n}}{b(t_n)} \left[\hat{a}_n(t_n) - a(t_n) \right] \xrightarrow{\text{d}} \mathcal{N}(0_{\mathbb{R}^2}, \mathbf{K}^2(\mathbf{u}) \mathbf{d}\mathbf{u} \mathbf{B}),$$

where B is a given symmetric matrix.

Proposition 3 : Estimation of extreme conditional quantiles

Let $(t_n)_n$ be a sequence in $[h_n, 1 - h_n]$, $(\tau_j)_{j=1,\dots,J}$ a positive and strictly decreasing sequence, $(\alpha_n)_n$ a sequence such that $\alpha_n \rightarrow 0$, $nh_n\alpha_n \rightarrow +\infty$ and $nh_n^3\alpha_n \rightarrow 0$ as $n \rightarrow +\infty$. Then,

$$\frac{\sqrt{nh_n\alpha_n}}{b(t_n)q_Z(\alpha_n)} \left[\hat{q}_{n,Y}(\alpha_n\tau_j | t_n) - q_Y(\alpha_n\tau_j | t_n) \right] \xrightarrow{\text{d}} \mathcal{N}(0_{\mathbb{R}^J}, \mathbf{K}^2(\mathbf{u}) \mathbf{d}\mathbf{u} \mathbf{C}),$$

where $C_{k,l} = \frac{1}{(\tau_{k \wedge l})^{1+\gamma} (\tau_{k \vee l})^\gamma}$, for all $(k, l) \in \{1, \dots, J\}^2$.

3. Estimators

Let $\{(x_i, Y_i), \dots, (x_n, Y_n)\}$ be a n -sample of observations such that $Y_i = a(x_i) + b(x_i)Z_i$, $i = 1, \dots, n$, where Z_i are independent and identically distributed. For simplicity, the design points are of the form $x_i = i/n$, $i = 1, \dots, n$ and $x_0 = 0$ by convention.

Notations :

- $\bar{F}_Y(y | x)$: the conditional survival function of Y given x .
For α fixed in $(0, 1)$,
- $q_Z(\alpha)$: the α -th quantile of Z ;
- $q_Y(\alpha | x)$: the conditional α -th quantile of Y given x ,

$$q_Y(\alpha | x) = \inf\{y / \bar{F}_Y(y | x) \leq \alpha\}.$$

Kernel estimator of $\bar{F}_Y(y | x)$:

$$\hat{F}_{n,Y}(y | x) = \sum_{i=1}^n \mathbb{1}_{\{Y_i > y\}} \frac{x_i}{x_{i-1}} K_h(x - t) dt, \quad (x, y) \in [0, 1] \times \mathbb{R},$$

where the bandwidth $h = h_n$ is a nonrandom sequence such that $h \rightarrow 0$ as $n \rightarrow \infty$, $\mathbb{1}(\cdot)$ is the indicator function and $K_h(\cdot) = K(\cdot/h)/h$, where K (kernel) is a probability density function on \mathbb{R} .

Kernel estimator of $q_Y(\alpha | x)$:

$$\hat{q}_{n,Y}(\alpha | x) \stackrel{\text{def}}{=} \hat{F}_{n,Y}^{\leftarrow}(\alpha | x) = \inf\{y / \hat{F}_{n,Y}(y | x) \leq \alpha\}, \quad \alpha \in (0, 1),$$

where $\hat{F}_{n,Y}^{\leftarrow}(\cdot | x)$ is the generalized inverse of $\hat{F}_{n,Y}(\cdot | x)$.

Estimators of $a(\cdot)$ and $b(\cdot)$:

Under this assumption : $\exists \rho_1, \rho_2, \rho_3 \in (0, 1)$, $\rho_1 > \rho_2 > \rho_3$ such that $q_Z(\rho_2) = 0$ and $q_Z(\rho_3) - q_Z(\rho_1) = 1$, we propose as :

• An estimator of $a(\cdot)$:

$$\hat{a}_n(x) = \hat{q}_{n,Y}(\rho_2 | x),$$

• An estimator of $b(\cdot)$:

$$\hat{b}_n(x) = \hat{q}_{n,Y}(\rho_3 | x) - \hat{q}_{n,Y}(\rho_1 | x).$$

Estimator of γ :

- Pseudo-observations \hat{Z}_i , associated with Z_i , $\lfloor nh \rfloor \leq i \leq n - \lfloor nh \rfloor$ defined by :

$$\hat{Z}_i = \frac{Y_i - \hat{a}_n(x_i)}{\hat{b}_n(x_i)}$$

• Number of pseudo-observations :

$$m = \#\{i / \lfloor nh \rfloor \leq i \leq n - \lfloor nh \rfloor\}$$

• Associated order-statistics : $\hat{Z}_{i,m}$, $i = 1, \dots, m$

• Hill-type estimator of γ :

$$\hat{\gamma}_n = \frac{1}{k} \sum_{i=1}^k \log \hat{Z}_{m-i+1,m} - \log \hat{Z}_{m-k,m},$$

where $k \in \{1, \dots, m\}$ is an intermediate sequence, i.e $k \rightarrow +\infty$, $\frac{k}{m} \rightarrow 0$.

6. References

- Antoch, J., Janssen, P., (1989). Nonparametric regression m -quantiles. *Statistics and Probability Letters*, 8(4) :355-362.
- Daouia, A., Gardes, L., Girard, S., (2013). On kernel smoothing for extremal quantile regression. *Bernoulli* 19, 2557-2589.
- Gardes, L., Girard, S., (2008). A moving window approach for nonparametric estimation of the conditional tail index. *J. Multivariate Anal.* 99(10), 2368-2388.
- Gardes, L., Girard, S., (2010). Conditional extremes from heavy-tailed distributions : an application to the estimation of extreme rainfall return levels. *Extremes* 13 (2), 177-204.
- Gardes, L., Girard, S., Lekina, A., (2010). Functional nonparametric estimation of conditional extreme quantiles. *J. Multivariate Anal.* 101 (2), 419-433.
- Goegebeur, Y., Guillou, A., Schorgen, A., (2014). Nonparametric regression estimation of conditional tails : the random covariate case. *Statistics*, 48 :732-755.
- Hill, B., (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3 :1163-1174.

5. Conclusion

The study of the problem in fixed design, which we have made, reveals interesting theoretical results. As a perspective, we are considering the conditional extreme-value index estimation and a validation of our results on simulations. We also plan to extend these results to the random framework.

7. Acknowledgements

Aboubacrène would like to thank CEA-MITIC (www.ceamitic.sn) and Inria of Grenoble Rhône-Alpes project team Mistis (<https://mistis.inrialpes.fr>) which hosted him for several research stays as part of the SIMERGE project (<http://mistis.inrialpes.fr/simerge/>).