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

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

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

Aboubacrène Ag Ahmad[1], Aliou Diop[1], El Hadji Deme[2] & Stéphane Girard[3]

[1] UGB - Université Gaston Berger de Saint-Louis, Sénégal.

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

[3] Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 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

$$\mathbf{Y} = \mathbf{a}(\mathbf{x}) + \mathbf{b}(\mathbf{x})\mathbf{Z},$$

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 $a(.)$ and the scale function $b(.) > 0$ are unknown. We assume that the survival function of Z denoted by $\bar{F}_Z$ belongs to the class of regulary varying functions at infinity

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

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

3.Estimators

Let $\{(x_1, Y_1), \ldots, (x_n, Y_n)\}$ be a n-sample of observations such that $Y_i = a(x_i) + b(x_i)Z_i$, $i = 1, \ldots, n$, where Z_i are independent and identically distributed. For simplicity, the design points are of the form $x_i = i/n, i = 1, \ldots, n$ and $x_0 = 0$ by convention.

Notations :

- $\bar{F}_Y(y \mid x)$: the conditional survival function of Y given x.

For α fixed in $(0, 1)$,

- $q_Z(\alpha)$: the $\alpha-$th quantile of Z;
- $q_Y(\alpha \mid x)$: the conditional $\alpha-$th quantile of Y given x,

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

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

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

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

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

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

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

Estimators of $a(.)$ and $b(.)$:

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 $a(.)$:

$$\widehat{\mathbf{a}}_{\mathbf{n}}(\mathbf{x}) = \widehat{\mathbf{q}}_{\mathbf{n},\mathbf{Y}}(\rho_2 \mid \mathbf{x}),$$

☛ An estimator of $b(.)$:

$$\widehat{\mathbf{b}}_{\mathbf{n}}(\mathbf{x}) = \widehat{\mathbf{q}}_{\mathbf{n},\mathbf{Y}}(\rho_3 \mid \mathbf{x}) - \widehat{\mathbf{q}}_{\mathbf{n},\mathbf{Y}}(\rho_1 \mid \mathbf{x}).$$

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 : $\hat{Z}_{i,m}, \ i = 1, \ldots, m$

☛ Hill-type estimator of γ :

$$\widehat{\gamma}_{\mathbf{n}} = \frac{\mathbf{1}}{\mathbf{k}} \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{k}} \log \widehat{\mathbf{Z}}_{\mathbf{m}-\mathbf{i}+\mathbf{1},\mathbf{m}} - \log \widehat{\mathbf{Z}}_{\mathbf{m}-\mathbf{k},\mathbf{m}},$$

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

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 $(\alpha_j)_{j=1,\ldots,J}$ a strictly decreasing sequence in $(0, 1)$. If $nh_n \to +\infty$ and $nh_n^3 \to 0$ as $n \to +\infty$, then

$$\left\{\frac{\sqrt{\mathbf{nh_n}}}{\mathbf{b}(\mathbf{t_n})}\left[\widehat{\mathbf{q}}_{\mathbf{n},\mathbf{Y}}(\alpha_{\mathbf{j}} \mid \mathbf{t_n}) - \mathbf{q}_{\mathbf{Y}}(\alpha_{\mathbf{j}} \mid \mathbf{t_n})\right]\right\}_{\{\mathbf{j}=\mathbf{1},\ldots,\mathbf{J}\}} \xrightarrow{\mathbf{d}} \mathcal{N}\left(\mathbf{0}_{\mathbb{R}^{\mathbf{J}}}, \int_{-\mathbf{1}}^{\mathbf{1}} \mathbf{K}^{\mathbf{2}}(\mathbf{u})\mathbf{du}\, \mathbf{A}\right),$$

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

Proposition 2 : Estimation of location and scale functions

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

$$\frac{\sqrt{\mathbf{nh_n}}}{\mathbf{b}(\mathbf{t_n})}\begin{pmatrix}\hat{a}_n(t_n) - a(t_n)\\ \hat{b}_n(t_n) - b(t_n)\end{pmatrix} \xrightarrow{\mathbf{d}} \mathcal{N}\left(\mathbf{0}_{\mathbb{R}^2}, \int_{-\mathbf{1}}^{\mathbf{1}} \mathbf{K}^{\mathbf{2}}(\mathbf{u})\mathbf{du}\, \mathbf{B}\right),$$

where B is a given symmetric matrix.

Proposition 3 : Estimation of extreme conditional quantiles

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

$$\left\{\frac{\sqrt{\mathbf{nh_n}\alpha_{\mathbf{n}}}}{\mathbf{b}(\mathbf{t_n})\mathbf{q}_{\mathbf{Z}}(\alpha_{\mathbf{n}})}\left[\widehat{\mathbf{q}}_{\mathbf{n},\mathbf{Y}}(\alpha_{\mathbf{n}}\tau_{\mathbf{j}} \mid \mathbf{t_n}) - \mathbf{q}_{\mathbf{Y}}(\alpha_{\mathbf{n}}\tau_{\mathbf{j}} \mid \mathbf{t_n})\right]\right\}_{\{\mathbf{j}=\mathbf{1},\ldots,\mathbf{J}\}} \xrightarrow{\mathbf{d}} \mathcal{N}\left(\mathbf{0}_{\mathbb{R}^{\mathbf{J}}}, \int_{-\mathbf{1}}^{\mathbf{1}} \mathbf{K}^{\mathbf{2}}(\mathbf{u})\mathbf{du}\, \mathbf{C}\right),$$

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

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.

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.

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/).