A Tail Index Regression Approach to Analyzing Extreme Event Data

Summer Research Fellowship (SRF) 2020 for Science Student
Poster No.: B6
Name: Xiong Jiaming
University No.: 3035533812
Student Major: Statistics

Abstract

The research focuses on proposing a nonparametric robust estimator for the tail index of a conditional Pareto-type distribution in the presence of censoring and random covariates. The estimator derived is based on conditional Pareto distribution since Pareto distribution is a heavy-tailed distribution. A tail index regression approach named minimum density power divergence (MDPD) is applied to derive the robust estimator. Compared to the existing approaches for extreme value data, namely the Block Maxima Method and Stupfler (2016) estimator, the estimator from MDPD approach can be applied in more general circumstances, is of high efficiency and can tolerate outliers and high proportions of censoring.

Introduction

Extreme value statistics is adopted to model extreme events that has a low frequency of occurrence but high impacts. In daily life, extreme value statistics are studied in various disciplines to predict that exceptional events will occur in the future. For example, in public health, the study of an unusual community epidemic (Figure 1) can facilitate the resource planning in hospitals. In environmental science, the extreme values of wave height is an important reference of safe ship design.

In previous studies, Block Maxima Method, one method from Extreme Value Theory (EVT), is widely used for modelling the extremes of a stationary time series. In Block Maxima Method, Generalized Extreme Value is expressed by three parameters, namely location parameter, scale parameter and shape parameter for the center, width and tail's heaviness of the given distributions respectively. This method groups maxima in each interval into a new sample so it is only useful for investigating events that occur at regular intervals and the efficiency of data is low. For example, the weekly number of Pneumonia and Influenza (P&I) mortality rate can be analyzed by Generalized Extreme Value using a moving time window.

Despite the constrain that the time series should be stationary, the study of extreme value statistics is concerned about the robustness of the estimator. For example, the estimator of Stupfler (2016) is not robust with respect to outliers. Estimators that are not robust can have severe instabilities in case of presence of outliers.

To study the extreme events in a more general and precise way, the new research uses minimum density power divergence method (MDPD) to find a robust estimator of a tail heaviness parameter under a random right censoring mechanism.

Methods and Materials

The minimum density power divergence (MDPD) is an approach which determines a parameter vector (θ) in the parametric model in order to minimize the density power divergence between the true (typically) unknown density of the data (denoted by f) and the parametric model (denoted by g). In the MDPD criterion, the estimating equations also consist of likelihood score functions with a relative-to-the model down-weighting for outlying observations (denoted by $\alpha \in [0,1]$).

$$\Delta_\alpha(f,g) := \begin{cases} \int_{\mathbb{R}} \left[g^{1+\alpha}(y) - \left(1+\frac{1}{\alpha}\right) g^\alpha(y) f(y) + \frac{1}{\alpha} f^{1+\alpha}(y)\right] dy, & \alpha > 0, \\ \int_{\mathbb{R}} \log \frac{f(y)}{g(y)} f(y) dy, & \alpha = 0. \end{cases}$$

For a conditional Pareto-type distribution, by means of the MDPD criterion, with adjustment of locally weighted estimation, minimize:

$$\widehat{\Delta}_\alpha(\gamma,\delta;\rho) := \frac{1}{n}\sum_{i=1}^{n} K_{h_n}(x_0 - X_i)\left\{\int_1^\infty g^{1+\alpha}(z;\gamma,\delta,\rho)dz - \left(1+\frac{1}{\alpha}\right) g^\alpha(Z_i;\gamma,\delta,\rho)\right\} \mathbb{1}_{\{T_i > t_n\}},$$

in case $\alpha > 0$ and minimize:

$$\widehat{\Delta}_0(\gamma,\delta;\rho) := -\frac{1}{n}\sum_{i=1}^{n} K_{h_n}(x_0 - X_i) \log g(Z_i;\gamma,\delta,\rho) \mathbb{1}_{\{T_i > t_n\}}$$

In case $\alpha = 0$.

α controls a trade-off between efficiency and robustness of the MDPD criterion. The estimator becomes more robust but less efficient when α increases. Figure 2 shows that If the dataset contains outliers which can deteriorate the estimation, $\alpha = 0$ is set to adverse outliers' effect on the estimation on the estimation of extreme value index.

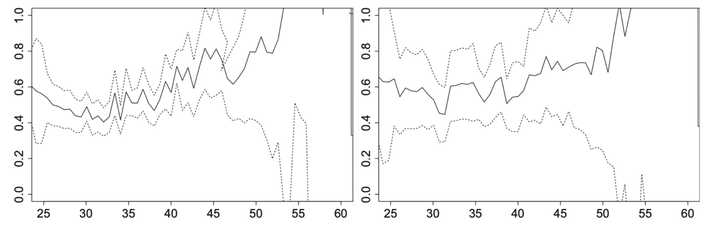

Figure 2. Dataset with contaminations: $\widehat{\gamma}_Y(x_0|-0.5)$, together with pointwise 95% confidence intervals for $\alpha = 0$ (left) and $\alpha = 0$ (right).

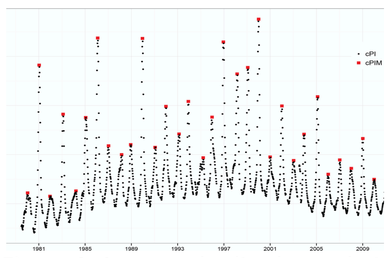

Figure 1. Age-standardized cumulative P&I mortality rates (1979–2011) in France. cPI rates correspond to black symbols and cPIM, the annual maxima, to the red symbols.

Results

In Figure 3, there is a substantial difference between the estimates $\widehat{\gamma}_{Y,n}(x_0|-0.5)$ and $\widehat{\gamma}_{T,n}(x_0|-0.5)$ due to the high percentage of censoring in the dataset. This emphasizes the value of MDPD approach to take censoring into account. The difference between $\widehat{\gamma}_{Y,n}(x_0|-0.5)$ with α = 0 and α = 0.5 shows the presents of contamination in the dataset and the control effect of α to ensure the robustness of the estimator from MDPD approach.

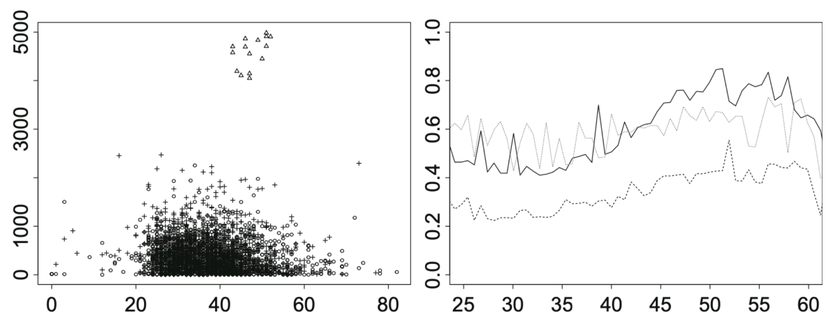

Figure 3. In left panel, Dataset with contamination (outliers represented with triangles; in right panel, $\widehat{\gamma}_{Y,n}(x_0|-0.5)$ with α = 0 (full line), α = 0.5 (dotted line) and $\widehat{\gamma}_{T,n}(x_0|-0.5)$ with a=0 as a function of age.

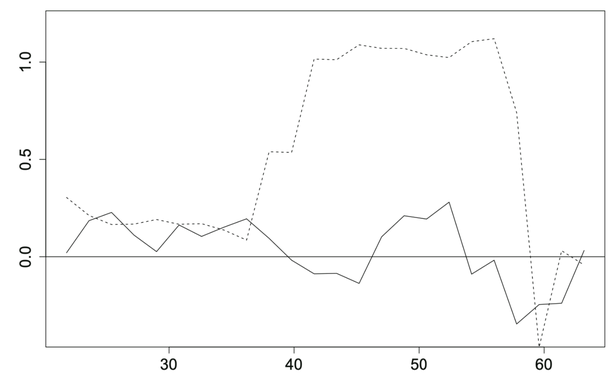

Figure 4. Same dataset as in Figure2 (right panel), Stupfler (2016) estimator for γ_Y as a function of x_0 for the uncontaminated data (full line) and the contaminated data (dotted line)

Compared with the estimator from the MDPD approach, figure 4 shows that there is a significant different between the the Stupfler (2016) estimates in two settings (uncontaminated data or contaminated data). This shows that the estimator of Stupfler (2016) is not robust with respect to outliers.

Conclusion

The extreme value statistics has wide applications in real-life examples. Compared to other estimators for extreme values, the estimator from the minimum density power divergence (MDPD) approach is robust in the sense that it can tolerate outliers and high proportions of censoring from the simulation study. Therefore, the estimator from MDPD approach can be applied to analyze extreme event data more precisely.

Discussion

Since the method for estimation is based on different estimators and on different data-driven procedures for selecting the tuning parameters, the direct comparisons with estimates are quite hard. For example, the estimator from MDPD is developed for $\gamma_Y > 0$ while the Stupfler (2016) estimator is designed for $\gamma_Y \in \mathbb{R}$. However, since in case of presence of outliers, estimators that are not robust can have severe instabilities, the robust estimator from MDPD approach is safer to be applied in practice.

References

1. Dierckx, G., Goegebeur, Y. & Guillou, A. Local Robust Estimation of Pareto-Type Tails with Random Right Censoring. Sankhya A (2019).
2. Richard Minkah, Tertius de Wet, Abhik Ghosh. Robust Estimation of Pareto-Type Tail Index through an Exponential Regression Model. 2019. ffhal-02116753v2f
3. Stupfler, G. (2016). Estimating the conditional extreme-value index under random right-censoring. *Journal of Multivariate Analysis, 144*, 1-24.
4. Quintela-del-Río A, Francisco-Fernández M. Nonparametric functional data estimation applied to ozone data: prediction and extreme value analysis. Chemosphere. 2011 Feb;82(6):800-8. doi: 10.1016/j.chemosphere.2010.11.025. Epub 2010 Dec 7. PMID: 21144549.
5. Thomas M, Lemaitre M, Wilson ML, Viboud C, Yordanov Y, Wackernagel H, et al. (2016) Applications of Extreme Value Theory in Public Health. PLoS ONE 11(7): e0159312. https://doi.org/10.1371/journal.pone.0159312
6. Wang, Hansheng and Tsai, Chih-Ling, Tail Index Regression (February 10, 2009). UC Davis Graduate School of Management Research Paper No. 08-09, Available at SSRN: https://ssrn.com/abstract=1340758