Nonparametric regression and extreme-value analysis

Stéphane Girard

Inria Grenoble Rhône-Alpes & LJK (team MISTIS).

655, avenue de l'Europe, Montbonnot. 38334 Saint-Ismier Cedex, France

Stephane.Girard@inria.fr

Abstract

This note summarizes my contributions to the estimation of extreme level curves. This problem is equivalent to estimating quantiles when covariate information is available and when their order converges to one as the sample size increases. Several estimators of these so-called "extreme conditional quantiles" are developped and the links with boundary or frontier estimation are emphasized.

1 Extreme-value analysis

Extreme value theory is a branch of statistics dealing with the extreme deviations from the bulk of probability distributions. More specifically, it focuses on the limiting distributions for the minimum or the maximum of a large collection of random observations from the same arbitrary (unknown) distribution. Let $x_1 < \cdots < x_n$ denote n ordered observations from a random variable X representing some quantity of interest. A p_n-quantile of X is the value q_{p_n} such that the probability that X is greater than q_{p_n} is p_n, i.e. $P(X > q_{p_n}) = p_n$. When $p_n < 1/n$, such a quantile is said to be extreme since it is usually greater than the maximum observation x_n. To estimate such extreme quantiles requires therefore specific methods to extrapolate information beyond the observed values of X. Those methods are based on Extreme value theory. This kind of issues appeared in hydrology. One objective was to assess risk for highly unusual events, such as 100-year floods, starting from flows measured over 50 years.

The decay of the survival function $P(X > x) = 1 - F(x)$, where F denotes the cumulative distribution function associated to X, is driven by a real parameter called the extreme-value index γ. When this parameter is positive, the survival function is said to be heavy-tailed, when this parameter is negative, the survival function vanishes above its right end point. If this parameter is zero, then the survival function decreases to zero at an exponential rate. An important part of our work is dedicated to the study of such distributions. For instance, in reliability, the distributions of interest are included in a semi-parametric family whose tails are decreasing exponentially fast.

These so-called Weibull tail-distributions encompass a variety of light-tailed distributions, such as Weibull, Gaussian, gamma and logistic distributions. Let us recall that a cumulative distribution function F has a Weibull tail if it satisfies the following property: There exists $\theta > 0$ such that for all $\lambda > 0$,

$$\lim_{y\to\infty} \frac{\log(1-F(\lambda y))}{\log(1-F(y))} = \lambda^{1/\theta}.$$

I also addressed the estimation of extreme level curves. This problem is equivalent to estimating quantiles when covariate information is available and when their order converges to one as the sample size increases. We show that, under some conditions, these so-called "extreme conditional quantiles" can still be estimated through a kernel estimator of the conditional survival function. Sufficient conditions on the rate of convergence of their order to one are provided to obtain asymptotically Gaussian distributed estimators. Making use of this result, some estimators of the extreme-value parameters are introduced and extreme conditional quantiles estimators are deduced [1, 2, 3, 4, 5, 6, 7, 8, 9]. Finally, the tail copula is widely used to describe the dependence in the tail of multivariate distributions. In some situations such as risk management, the dependence structure may be linked with some covariate. The tail copula thus depends on this covariate and is referred to as the conditional tail copula. The aim of [10] is to propose a nonparametric estimator of the conditional tail copula and to establish its asymptotic normality.

Applications are found in hydrology [11, 12] and more generally in risk estimation [13, 14, 15].

2 Boundary or frontier estimation

In image analysis, the boundary estimation problem arises in image segmentation as well as in supervised learning. In the extreme quantiles approach, the boundary bounding the set of points is viewed as the larger level set of the points distribution. Its estimation is thus an extreme quantile curve estimation problem. Estimators based on projections [16, 17] as well as on kernel regression methods are applied on the extreme values set [18, 19]. These two families are unified in [20, 21] and the asymptotic distribution of the L_1 error is investigated in [22, 23, 24]. Applications to econometrics are considered in [25, 26].

References

[1] L. Gardes and S. Girard. On the estimation of the functional Weibull tail-coefficient. *Journal of Multivariate Analysis*, 2015. to appear.

[2] A. Daouia, L. Gardes, and S. Girard. On kernel smoothing for extremal quantile regression. *Bernoulli*, 19:2557–2589, 2013.

[3] L. Gardes and S. Girard. Functional kernel estimators of large conditional quantiles. *Electronic Journal of Statistics*, 6:1715–1744, 2012.

[4] J. Carreau and S. Girard. Spatial extreme quantile estimation using a weighted log-likelihood approach. *Journal de la Société Française de Statistique*, 152(3):66–83, 2011.

[5] A. Daouia, L. Gardes, S. Girard, and A. Lekina. Kernel estimators of extreme level curves. *Test*, 20(14):311–333, 2011.

[6] L. Gardes, S. Girard, and A. Lekina. Functional nonparametric estimation of conditional extreme quantiles. *Journal of Multivariate Analysis*, 101:419–433, 2010.

[7] L. Gardes and S. Girard. A moving window approach for nonparametric estimation of the conditional tail index. *Journal of Multivariate Analysis*, 99:2368–2388, 2008.

[8] S. Girard and S. Louhichi. On the strong consistency of the kernel estimator of extreme conditional quantiles. In E. Ould-Said et al., editor, *Functional Statistics and Applications*, pages 59–77. Springer, 2015.

[9] L. Gardes and S. Girard. Functional kernel estimators of conditional extreme quantiles. In F. Ferraty, editor, *Recent advances in functional data analysis and related topics*, pages 135–140. Springer, Physica-Verlag, 2011.

[10] L. Gardes and S. Girard. Nonparametric estimation of the conditional tail copula. *Journal of Multivariate Analysis*, 137:1–16, 2015.

[11] J. Carreau, D. Ceresetti, E. Ursu, S. Anquetin, J.D. Creutin, L. Gardes, S. Girard, and G. Molinié. Evaluation of classical spatial-analysis schemes of extreme rainfall. *Natural Hazards and Earth System Sciences*, 12:3229–3240, 2012.

[12] L. Gardes and S. Girard. Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels. *Extremes*, 13(2):177–204, 2010.

[13] E. Deme, S. Girard, and A. Guillou. Reduced-bias estimators of the conditional tail expectation for heavy-tailed distributions. In M. Hallin et al., editor, *Mathematical Statistics and Limit Theorems*, pages 105–123. Springer, 2015.

[14] J. El Methni, L. Gardes, and S. Girard. Nonparametric estimation of extreme risks from conditional heavy-tailed distributions. *Scandinavian Journal of Statistics*, 41:988–1012, 2014.

[15] E. Deme, S. Girard, and A. Guillou. Reduced-bias estimator of the proportional hazard premium for heavy-tailed distributions. *Insurance: Mathematics and Economics*, 22:550–559, 2013.

[16] S. Girard and P. Jacob. Extreme values and Haar series estimates of point process boundaries. *Scandinavian Journal of Statistics*, 30(2):369–384, 2003.

[17] S. Girard and P. Jacob. Projection estimates of point processes boundaries. *Journal of Statistical Planning and Inference*, 116(1):1–15, 2003.

[18] S. Girard and P. Jacob. Extreme values and kernel estimates of point processes boundaries. *ESAIM: Probability and Statistics*, 8:150–168, 2004.

[19] S. Girard and P. Jacob. A note on extreme values and kernel estimators of sample boundaries. *Statistics and Probability Letters*, 78:1634–1638, 2008.

[20] S. Girard and L. Menneteau. Central limit theorems for smoothed extreme value estimates of point processes boundaries. *Journal of Statistical Planning and Inference*, 135(2):433–460, 2005.

[21] S. Girard and L. Menneteau. Smoothed extreme value estimators of non-uniform point processes boundaries with application to star-shaped supports estimation. *Communication in Statistics - Theory and Methods*, 37(6):881–897, 2008.

[22] J. Geffroy, S. Girard, and P. Jacob. Asymptotic normality of the L_1-error of a boundary estimator. *Nonparametric Statistics*, 18(1):21–31, 2006.

[23] S. Girard. On the asymptotic normality of the L_1- error for Haar series estimates of Poisson point processes boundaries. *Statistics and Probability Letters*, 66:81–90, 2004.

[24] S. Girard and P. Jacob. Asymptotic normality of the L_1-error for Geffroy's estimate of Poisson point process boundaries. *Publications de l'Institut de Statistique de l'Université de Paris*, XLIX:3–17, 2005.

[25] A. Daouia, S. Girard, and A. Guillou. A γ-moment approach to monotonic boundaries estimation: with applications in econometric and nuclear fields. *Journal of Econometrics*, 178:727–740, 2014.

[26] A. Daouia, L. Gardes, and S. Girard. Nadaraya's estimates for large quantiles and free disposal support curves. In I. Van Keilegom and P. Wilson, editors, *Exploring research frontiers in contemporary statistics and econometrics*, pages 1–22. Springer, 2012.