Statistical Learning with Extreme Values
Master's program MVA, Université Paris-Saclay

Anne Sabourin[(1)*], Stephan Clémençon[(2)†]

[(1)] *Université Paris Cité, CNRS, MAP5, F-75006 Paris, France.*
[(2)] *LTCI, Télécom Paris, Institut polytechnique de Paris, France*

October 2, 2023

With the ubiquity of sensors, Big Data are now increasingly available in a wide variety of domains of human activity (science, industry, health, environment, commerce, security, ...) and rare/extreme phenomena are becoming observable in a significant manner. Before, such events were mainly 'out-of-sample' and Extreme Value Theory (EVT), the field of probability and statistics concerned with tails of distributions, tackled their study through (parametric) modelling essentially. With the need for analyzing extreme observations, carrying often the critical information to design solutions to applications (*e.g.* health monitoring of complex infrastructures) for which worst-case scenarios crucially matter, the most recent years have seen an increasing interest of the EVT research community towards novel machine learning algorithms and statistical learning theory, resonating with a continuing effort of the statistical community to address larger-dimensional problems with computationally feasible approaches (see e.g. the review Engelke and Ivanovs (2021)).

Let X be a random element (variable, vector, or function) of interest. One major goal of EVT is to provide probabilistic descriptions and statistical inference methods for the conditional distribution of $t^{-1}X$ given large $\|X\|$, where $\|\cdot\|$ is a semi-norm and t is a large threshold (see *e.g.* the monographs (De Haan and Ferreira (2007); Resnick (2008)). In applications, relevant thresholds t may be as high as the largest observation among n realizations of X. Probabilistic extrapolation is then needed to use the information brought by a subsample of size $k_n \ll n$ composed of the observations with the largest semi-norms. This requires sound theoretical assumptions pertaining to the theory of regular variation and maximum domains of attraction, ensuring that a limit distribution $\mu = \lim \mathrm{Law}(t^{-1}X \mid \|X\| > t)$ exists as $t \to \infty$, up to suitable standardization. This stylized setting encompasses a wide range of applications in various scientific displines and risk management where extremes have tremendous impact, such as climate science, insurance, industrial monitoring systems (Beirlant et al. (2004)).

*e-mail address: anne.sabourin@u-paris.fr
†e-mail address: stephan.clemencon@telecom-paris.fr

This course aims at introducing the students with the most recent development of statistical learning viewpoints on EVT. From a theoretical perspective they will be offered an overview of recent statistical learning theory for rare events, in addition to the necessary probabilistic background on extreme value theory and regular variation. Theoretical development will be motivated and illustrated by recent successful algorithms for handling extreme values, be it for anomaly detection, extreme event classification or dimension reduction in distributional tails.

Course structure: Each session is approximately divided into a 2h lecture and a 1h tutorial.

The second to last lecture is a Q&A session where students can in particular get help with their homework.

The last lecture will be organised as a seminar / working group where the students will present recent research articles.

Grading

- 50% Homework (theoretical and practical exercises): Each course comes with a list of exercises, partly coding, partly theory. These exercises should be handed out 2 weaks maximum after the day they are released. There is a Bonus rule allowing students to improve upon past exercises (maximum 4 of them) after the Q&A last course
- 50% Oral presentation (20 minutes, 10 slides) + written report ($\leq$ 10 pages).

Syllabus (9*3h)

1. **Basics of EVT: learning from block maxima.**
   Context and applications in risk management and anomaly detection. Fisher and Tipett's theorem withelements of proof. Method of block maxima
   Tutorial: derivation of norming constants, numerical illustration for the weak convergence of block maxima, case studies, choice of the block size.
2. **Peaks-Over-Tresholds (POT) and Regular variation.**
   Link between POT and block maxima. Generalized Pareto distributions. Basics of regular variation and vague convergence. Informal introduction to the Hill estimator.
   Tutorial: Elements of proof - POT modeling on case studies - threshold choice - Hill estimator in practice.
3. **Regular Variation II, tail measures and weak convergence.**
   More on regular variation - Karamata representation theorem - weak consistency of

the Hill estimator - Quantile estimation.
Tutorial: Visualization of weak convergence of tail measures - Elements of proof for Karamata.

4. **Multivariate extremes.**
   Reduction to the standard case - characterizing max-id distributions - characterizing simple max-stable distributions - Angular measure - Multivariate Peaks-over-threshold
   Tutorial: Elements of proofs - Simulation - Basics of non-parametric estimation in moderate dimension (kernel and histogram methods).

5. **Statistical learning guarantees for extremes.**
   Vapnik-type concentration inequalities for rare events - illustrations on estimation tasks in multivariate extremes (standard tail dependence function and angular measure)
   Tutorial: Elements of proof and numerical illustration of the error bounds

6. **High dimensional extremes.**
   Notions of sparsity in multivariate extremes - Applications to anomaly detection
   Tutorial: PCA - multiple subspace clustering - Anomaly detection

7. **Supervised learning with extreme values.**
   Learning with extreme covariates (classification) - Dimension reduction with extreme targets
   Tutorial: Demo and Elements of proof

8. **Q&A session.**
   Help with articles/homework

9. **Oral presentations**
   Mandatory attendance

References

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2004). *Statistics of extremes: theory and applications*, volume 558. John Wiley & Sons.

Chiapino, M., Clémençon, S., Feuillard, V., and Sabourin, A. (2020). A multivariate extreme value theory approach to anomaly clustering and visualization. *Computational Statistics*, 35(2):607–628.

De Haan, L. and Ferreira, A. (2007). *Extreme value theory: an introduction*. Springer Science & Business Media.

Engelke, S. and Hitz, A. S. (2020). Graphical models for extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(4):871–932.

Engelke, S. and Ivanovs, J. (2021). Sparse structures for multivariate extremes. *Annual Review of Statistics and Its Application*, 8:241–270.

Goix, N., Sabourin, A., and Clémençon, S. (2015). Learning the dependence structure of rare events: a non-asymptotic study. In *Conference on Learning Theory*, pages 843–860.

Goix, N., Sabourin, A., and Clémençon, S. (2016). Sparse representation of multivariate extremes with applications to anomaly ranking. In *Artificial Intelligence and Statistics*, pages 75–83.

Hitz, A. and Evans, R. (2016). One-component regular variation and graphical modeling of extremes. *Journal of Applied Probability*, 53(3):733–746.

Hult, H. and Lindskog, F. (2006). Regular variation for measures on metric spaces. *Publications de l'Institut Mathématique*, 80(94):121–140.

Lhaut, S., Sabourin, A., and Segers, J. (2022). Uniform concentration bounds for frequencies of rare events. *Statistics & Probability Letters*, 189:109610.

Resnick, S. I. (2008). *Extreme values, regular variation, and point processes*, volume 4. Springer Science & Business Media.

Thomas, A., Clémençon, S., Gramfort, A., and Sabourin, A. (2017). Anomaly Detection in Extreme Regions via Empirical MV-sets on the Sphere. In *Proceedings of AISTATS*, volume 54, pages 1011–1019. PMLR.