

Thèse en Statistique

Clustering de valeurs extrêmes sur des séries chronologiques

Gilles DURRIEU et Ion GRAMA

Laboratoire de Mathématiques de Bretagne Atlantique UMRS CNRS 6205
Faculté des Sciences et Sciences de l'Ingénieur (UFR SSI), Université Bretagne Sud
gilles.durrieu@univ-ubs.fr et ion.grama@univ-ubs.fr

Arthur FETIVEAU

ALDECIS, Paris

afu5600@aldecis.com

Mots clefs : statistique non paramétrique, valeurs extrêmes, Deep Learning, Machine Learning.

Financement : Allocation de thèse CIFRE d'une durée de 3 ans.

Salaire : 2400€ brut mensuel + frais de transport (dont ceux vers les bureaux d'Aldecis)
+ avantages annexes de l'entreprise (titres restaurants, chèques cadeaux ...).

Laboratoire et entreprise d'accueil :

- Laboratoire de Mathématiques de Bretagne Atlantique UMRS CNRS 6205, Université Bretagne Sud, Campus de Tohannic, BP573, 56017 Vannes.
- Entreprise ALDECIS, 22 rue Bréguet, 75011 Paris.

Présentation synthétique du sujet de thèse : La généralisation de la digitalisation (doublement tous les trois mois des informations accessibles en entreprise), couplée à l'essor rapide des intelligences artificielles, entraîne une croissance exponentielle des Tableaux de Bord. Paradoxalement, le temps réellement consacré à l'analyse de ces données tend à se réduire, en grande partie en raison de la complexification par la normalisation de sources de données toujours plus riches et variées dans le cadre du Reporting. Pris dans cette double contrainte, les analystes sont bien souvent davantage sollicités pour produire des Rapports que pour mener une analyse approfondie des données et en extraire des insights pertinents, au service de la résolution des problématiques sous-jacentes.

Dans le cadre d'une première thèse CIFRE, nous avons engagé une modélisation des données de performance des organisations, dans le but d'y détecter des anomalies structurelles et conjoncturelles. Cette démarche a permis de développer des modèles stochastiques de priorisation des signaux atypiques, afin de se focaliser sur les cas les plus critiques. Aujourd'hui l'objectif est d'identifier et d'anticiper les causes profondes de ces anomalies rares, dans une perspective d'analyse prédictive et de prévention.

L'estimation des probabilités et des risques associés à des événements rares joue un rôle essentiel dans des domaines tels que la climatologie, la santé publique, l'assurance ou la mesure de performance. Nous proposerons une méthode d'estimation des probabilités rares, ainsi que des quantiles extrêmes, basée sur un ajustement automatique du seuil de localisation d'un modèle statistique en adéquation avec les données observées.

Nous développerons également une procédure statistique automatique de détection des anomalies extrêmes, s'inscrivant dans une approche combinant apprentissage automatique (Machine Learning) et apprentissage profond (Deep Learning). Cette méthode aura pour objectif d'identifier les observations atypiques et de les regrouper en clusters selon leurs similarités structurelles, en vue de mutualiser l'application cohérente de traitements correctifs

ciblés et de proposer des mesures préventives (soit par risques de propagation, soit par faiblesses structurelles identiques).

Les propriétés asymptotiques des procédures proposées seront rigoureusement analysées, tant sur le plan théorique que par le biais de simulations numériques, avant leur mise en œuvre sur les données réelles issues d'ALDECIS, une organisation constituée d'experts en IA, chargée d'aider ses clients à améliorer leurs performances.

Compétences requises : Master 2 Mathématiques Appliquées, spécialisé en Statistique, ou formation équivalente de type école d'ingénieurs. Des compétences en informatique et logiciel de statistique sont souhaitées (Python et/ou R).

Date limite de dépôt des candidatures (CV détaillé et relevés de notes depuis le BAC) : 8 décembre 2025 à 16h par mail à l'adresse mail gilles.durrieu@univ-ubs.fr

Date souhaitée pour le début de la thèse : janvier 2026.

Références :

- Ion Grama and Vladimir Spokoiny (2008) Statistics of extremes by oracle estimation. *Annals of Statistics*, Vol. 36, No. 4, 1619-1648.
- Gilles Durrieu, Ion Grama, Quang Khoai Pham and Jean-Marie Tricot (2015) Nonparametric adaptive estimation of conditional probabilities of rare events and extreme quantiles. *Extremes*. Vol. 18, Issue 3, 437-478.
- Gilles Durrieu, Quang-Khoai Pham, Anne-Sophie Foltête, Valérie Maxime, Ion Grama, Véronique Le Tilly, Helene Duval, Jean-Marie Tricot, Chiraz Ben Naceur, Olivier Sire (2016) Dynamic extreme values modeling and monitoring by means of sea shores water quality biomarkers and valvometry. *Environmental Monitoring and Assessment*, Vol. 188, 401-409.
- Gilles Durrieu, Ion Grama, Kevin Jaunâtre, Quang-Khoai Pham, Jean-Marie Tricot (2018) Extremefit: A Package for Extreme Quantiles. *Journal of Statistical Software*, Vol. 87, Issue 12, 1-20.
- Grama I., Jaunâtre K. (2019) Estimation of extreme survival probabilities with Cox model, *Statistics*, 53:4, 807-838.
- Arthur Fétiveau, Gilles Durrieu, Emmanuel Frénod, Claude-Henri Meledo and Benoit Prat (2025) Permutations based model for business performance, *Discrete and Continuous Dynamical Systems Series S, American Institute of Mathematical Science*, 18(7), 1699-1722.
- Arthur Fétiveau, Gilles Durrieu and Emmanuel Frénod (2025) Probabilistics models for permutations and dependency, *Communication in Statistics - Simulation and Computation*, 1-19.
- Tristan Senga Kiessé and Gilles Durrieu (2024) On a discrete symmetric optimal associated kernel for estimating count data distributions, *Statistics and Probability Letters*, vol 208, <https://www.sciencedirect.com/science/article/pii/S0167715224000476>

Thèses soutenues sur le sujet :

- Khoai Quang Pham (2011-2014). Thèse intitulée « Estimation de quantiles extrêmes et probabilités rares d'un processus stochastique : application en environnement » et soutenue le 9 janvier 2015. Direction : Gilles Durrieu (50%) et Ion Grama (50%).
- Kevin Jaunâtre (2015-2018). Thèse intitulée « Analyse et modélisation statistique de données de consommation électrique » et soutenue le 18/01/2019. Direction : Gilles Durrieu (50%) et Ion Grama (50%).
- Arthur Fétiveau (2021-2024). Thèse intitulée « Modélisation et prédiction des indicateurs de performance d'entreprises : analyse des atypiques » et soutenue le 8 novembre 2024. Direction : Gilles Durrieu (50%) et Emmanuel Frénod (50%).