RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

ONERA
THE FRENCH AEROSPACE LAB

www.onera.fr

PROPOSITION DE STAGE EN COURS D'ETUDES

Référence : **DTIS-2027-15**
(à rappeler dans toute correspondance)

Lieu : Toulouse

Département/Dir./Serv. : DTIS/RIME

Tél. : 05 62 25 26 63

Responsable(s) du stage : Jérôme Morio — Email. : jerome.morio@onera.fr
Louison Bocquet-Nouaille — louison.bocquet_nouaille@onera.fr
Benjamin Bobbia (ISAE) — benjamin.bobbia@isae-supaero.fr

DESCRIPTION DU STAGE

Thématique(s) : 3S – Sureté et Sécurité des Systèmes cyber-physiques

Type de stage : ☒ Fin d'études bac+5 ☒ Master 2 ☐ Bac+2 à bac+4 ☐ Autres

Title: Extreme value theory applied to the LogSumExp estimator

Topic: In seemingly unrelated fields, the estimation of quantities that are sensitive to the extreme values of a sample relies on the same mathematical tool: the LogSumExp (LSE), defined as the logarithm of a sum of exponentials. In statistical physics, the LSE is used to estimate free energy differences from non-equilibrium simulations through the Jarzynski estimator [1]. In optimization and machine learning, the LSE serves as a loss function that makes it possible to tune a model's sensitivity to the hardest observations [2]. In this community, the LSE is also known as the "RealSoftMax" or the "multivariable softplus", its gradient being none other than the softmax function. In actuarial science, the LSE estimates the exponential premium principle, used for pricing under risk aversion [3].

The LSE depends on a parameter that controls its degree of sensitivity to the extreme values of the sample: the higher this parameter, the more the estimator focuses on the most atypical observations. This tuning is an asset, since it allows the estimator to be adapted to the problem at hand (risk aversion, robustness to outliers). However, it comes at the cost of a loss of statistical stability as the estimator's sensitivity to extremes increases. More specifically, as this parameter increases for a fixed sample size, the sum of exponentials is dominated by fewer and fewer observations, until it is asymptotically dominated by the largest observation alone. The variance of the estimator then grows rapidly. In other words, its value can change substantially from one sample to another, depending on whether extreme values are present.

Extreme value theory (EVT) [4,5] provides a particularly relevant framework for studying the LSE in this high-sensitivity regime. This branch of statistics, devoted to the study of rare events (hundred-year floods, stock market crashes, disasters), establishes in particular a central universality result: regardless of the nature of the phenomenon under study, the behavior of extreme values belongs to a single family of probability distributions (under general conditions). Applied to the LSE, this result makes it possible to describe the behavior of the estimator from the tail parameters of the distribution alone, without knowing the full distribution of the data.

The objective of this internship is to study the statistical properties of the LSE in the regime where it is highly sensitive to the extreme values of the sample. In particular, the work will aim to develop an EVT-based approach to analyze how the LSE parameter influences its statistical properties and to identify parameter choices suited to this regime, to study the theoretical properties of this approach (bias, variance, convergence), and finally to assess its performance through numerical simulations. The expected outcomes are a delineation of the sensitivity regimes in which the empirical LSE remains reliable, approximations of its bias and variance beyond these regimes, and criteria for choosing the LSE parameter so as to minimize its bias, variance or mean squared error (MSE) when sensitivity to extreme values is high.

The internship will be organized in two phases: a literature review of the LSE and EVT to become familiar with the necessary tools, followed by a phase of theoretical developments and numerical simulations (Python/R) carried out jointly. The proposed LSE estimator will then be applied to several real-world datasets

from aerospace applications (ice accretion on an aircraft wing, electromagnetic compatibility of on-board cabling).

[1] Jarzynski, C. (1997). Nonequilibrium Equality for Free Energy Differences. *Physical Review Letters*.

[2] Li, T., Beirami, A., Sanjabi, M., and Smith, V. (2023). On tilted losses in machine learning: Theory and applications. *Journal of Machine Learning Research*.

[3] Kaas, R., Goovaerts, M., Dhaene, J., and Denuit, M. (2008). Modern actuarial risk theory: using R.

[4] Bousquet, N. and Bernardara, P. (2021). Extreme Value Theory with Applications to Natural Hazards: From Statistical Theory to Industrial Practice.

[5] de Haan, L. and Ferreira, A. (2006). Extreme Value Theory.

Est-il possible d'envisager un travail en binôme ? Non

Méthodes à mettre en oeuvre :

☒ Recherche théorique	☐ Travail de synthèse
☒ Recherche appliquée	☒ Travail de documentation
☐ Recherche expérimentale	☐ Participation à une réalisation

Possibilité de prolongation en thèse : Oui

Durée du stage : Minimum : 4 mois Maximum : 6 mois

Période souhaitée : Février – Septembre 2027

PROFIL DU STAGIAIRE

Connaissances et niveau requis : Mathématiques appliquées, probabilités et statistique, maîtrise d'un langage de programmation.	Ecoles ou établissements souhaités : 3ème année d'école d'ingénieur généraliste, Master 2.