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

Intitulé : Théorie des valeurs extrêmes appliquée à l'estimateur LogSumExp

Sujet : Dans des domaines a priori très différents, l'estimation de quantités sensibles aux valeurs extrêmes d'un échantillon repose sur un même outil mathématique : le LogSumExp (LSE) défini comme le logarithme d'une somme d'exponentielles. En physique statistique, le LSE permet d'estimer des différences d'énergie libre à partir de simulations hors-équilibre avec l'estimateur de Jarzynski [1]. En optimisation et machine learning, le LSE sert de fonction de perte qui permet d'ajuster la sensibilité d'un modèle aux observations les plus difficiles [2]. Dans cette communauté, le LSE est aussi connu sous les noms de « RealSoftMax » ou de « SoftPlus multivarié », son gradient n'étant autre que la fonction softmax. En actuariat, le LSE estime le principe de prime exponentielle, utilisé pour la tarification sous aversion au risque [3].

Le LSE dépend d'un paramètre qui contrôle son degré de sensibilité aux valeurs extrêmes de l'échantillon : plus ce paramètre est élevé, plus l'estimateur se concentre sur les observations les plus atypiques. Ce réglage constitue un atout, puisqu'il permet d'adapter l'estimateur au problème considéré (aversion au risque, robustesse aux valeurs aberrantes). Il a néanmoins pour contrepartie une dégradation de la stabilité statistique de l'estimateur à mesure que sa sensibilité aux extrêmes augmente. Concrètement, lorsque ce paramètre augmente à taille d'échantillon fixée, la somme d'exponentielles est dominée par un nombre de plus en plus réduit d'observations, jusqu'à ce que la somme soit asymptotiquement dominée par la plus grande observation uniquement. La variance de l'estimateur croît alors rapidement, autrement dit sa valeur peut changer fortement d'un échantillon à l'autre, selon la présence de valeurs extrêmes.

La théorie des valeurs extrêmes (TVE) [4,5] constitue un cadre particulièrement pertinent pour étudier le LSE dans ce régime de forte sensibilité. Cette branche des statistiques, consacrée à l'étude des événements rares (crues centennales, krachs boursiers, catastrophes), établit notamment un résultat d'universalité central : indépendamment de la nature du phénomène étudié, le comportement des valeurs extrêmes appartient à une même famille de lois de probabilité (sous des conditions générales). Appliqué au LSE, ce résultat permet de décrire le comportement de l'estimateur à partir des seuls paramètres de queue de distribution, sans connaître la loi complète des données.

L'objectif du stage est d'étudier les propriétés statistiques du LSE dans le régime où celui-ci est fortement sensible aux valeurs extrêmes de l'échantillon. Le travail visera notamment à développer une approche fondée sur la TVE pour analyser l'influence du paramètre du LSE sur ses propriétés statistiques et identifier des choix de paramètre adaptés à ce régime, à en étudier les propriétés théoriques (biais, variance, convergence) et enfin, à évaluer ses performances par simulations numériques. Les résultats attendus sont une délimitation des régimes de sensibilité dans lesquels le LSE empirique reste fiable, des approximations de son biais et de sa variance au-delà de ces régimes, ainsi que des critères permettant de choisir le paramètre du LSE afin de minimiser son biais, sa variance ou son erreur quadratique moyenne (MSE) lorsque la sensibilité aux valeurs extrêmes est élevée.

Le stage s'organisera en deux temps : une étude bibliographique du LSE et de la TVE pour s'approprier les outils nécessaires, puis une phase de développements théoriques et de simulations numériques (Python/R)

menés de façon conjointe. L'estimateur LSE proposé sera ensuite appliqué à différents jeux de données réelles issues d'application aérospatiales (accrétion de givre sur une aile d'avion, compatibilité électromagnétique de câblages embarqués).

[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.