

PROPOSITION DE SUJET DE THESE

Intitulé : IA générative et théorie des valeurs extrêmes : modélisation et estimation des queues multivariées

Référence : **TIS-DTIS-2026-24**
(à rappeler dans toute correspondance)

Début de la thèse : Octobre 2026

Date limite de candidature : Juillet 2026

Mots clés

Estimation d'événements rares, Théorie des valeurs extrêmes, IA générative

Profil et compétences recherchées

Profil : Master 2 et/ou dernière année d'école d'ingénieur généraliste.

Compétences : Mathématiques appliquées, probabilités, statistique, maîtrise d'un langage de programmation.

Présentation du projet doctoral, contexte et objectif

Contexte

Les modèles génératifs modernes (normalizing flows, GAN, modèles de diffusion) ont récemment bouleversé le champ de l'apprentissage statistique en permettant de générer et d'approximer des distributions complexes de haute dimension. Ces méthodes ont trouvé de nombreuses applications en vision par ordinateur, en traitement du signal ou encore en simulation physique. Cependant, ces approches restent largement centrées sur la masse principale des données. Les régions de probabilité extrêmement faible, correspondant aux événements rares et intenses, demeurent mal représentées. Or, ces événements sont précisément l'objet de la théorie des valeurs extrêmes (TVE), branche des statistiques dédiée à l'étude des queues de distribution.

À l'inverse, la TVE dispose d'un formalisme mathématique robuste pour modéliser et extrapoler les comportements dans les queues. Mais elle souffre de limitations fortes : la rareté des observations extrêmes et la malédiction de la dimension rendent son application délicate dès que l'on considère plus de deux ou trois variables. Dans des domaines comme l'aéronautique, l'énergie ou la finance, la nécessité d'estimer correctement les risques extrêmes conjoints rend donc indispensable le développement de nouvelles méthodes hybrides.

Cette thèse s'inscrit dans cette dynamique, en cherchant à croiser la puissance des modèles génératifs modernes avec la rigueur de la TVE. Deux grandes questions motiveront les travaux :

- Comment rendre les modèles génératifs plus fiables dans les zones extrêmes ?
- Comment exploiter l'IA générative pour améliorer les estimations statistiques des queues multivariées ?

Problématique

L'une des tâches principales en TVE est l'extrapolation de quantiles ou de probabilités rares au-delà de l'échantillon observé. En pratique, les méthodes directes ne permettent pas d'extrapoler trop loin sans générer des erreurs d'estimation massives. La portée de l'extrapolation dépend directement du nombre d'extrêmes disponibles, souvent très faible. Dans ce cadre, l'utilisation d'un modèle génératif pourrait permettre d'augmenter artificiellement la quantité d'information disponible dans les queues, sous réserve de le concevoir pour bien capturer ces régions rares.

Deux axes de recherche structurent la problématique :

- **IA pour les extrêmes** : étudier comment modifier l'entraînement de modèles génératifs (normalizing flows, diffusion models) afin qu'ils respectent les contraintes asymptotiques de la TVE. Cela inclut l'intégration de connaissances théoriques (variations régulières, indices de queue, mesure angulaire) dans les architectures ou les fonctions de coût, et le

développement de schémas d'apprentissage pondérés qui concentrent l'attention sur les observations rares.

- **Extrêmes avec l'IA** : explorer l'utilisation des modèles génératifs comme outils pour l'estimation de quantités extrêmes. Une piste centrale est de les employer comme distributions de proposition dans un cadre d'importance sampling, ou comme générateurs de variables de contrôle exploitant leurs densités explicites. Ces approches visent à réduire drastiquement la variance des estimateurs de probabilités rares multivariées en grande dimension.

Contributions attendues

- **Améliorer l'IA dans les extrêmes** : une première étape sera de caractériser les performances des modèles génératifs existants dans l'apprentissage des queues. On cherchera ensuite à développer des variantes adaptées : normalizing flows entraînés avec des pertes pondérées par l'inverse de la probabilité, modèles de diffusion contraints par des conditions EVT, architectures hybrides combinant copules extrêmes et flows. L'objectif est de concevoir des modèles génératifs capables de reproduire fidèlement la structure des dépendances dans les queues multivariées.
- **Améliorer l'estimation extrême grâce à l'IA** : les modèles génératifs seront étudiés comme briques pour la simulation d'événements rares. L'idée est de les utiliser comme lois auxiliaires pour l'importance sampling ou comme base pour construire des variables de contrôle corrélées aux événements d'intérêt. On évaluera leur capacité à réduire la variance des estimateurs classiques de quantiles extrêmes et de probabilités conjointes rares, en comparaison avec les méthodes paramétriques traditionnelles (copules d'extrêmes, Hüsler–Reiss, Gumbel, etc.).
- **Applications en forte dimension** : une part importante des travaux consistera à appliquer ces méthodes à des cas réalistes, notamment dans le domaine aéronautique et spatial. On considérera par exemple l'estimation de la probabilité de conjonctions d'événements rares (turbulences sévères, pannes simultanées, trajectoires déviantes). Ces cas illustrent l'intérêt d'approches génératives capables de réduire les coûts de calcul et d'améliorer la fiabilité des estimations dans des contextes où les observations empiriques sont limitées.

Références

Allouche, M., Girard, S., Gobet, E. (2025). ExcessGAN: simulation above extreme thresholds using Generative Adversarial Networks.

Asghar, S., Pei, Q. X., Volpe, G., Ni, R. (2024). Efficient rare event sampling with unsupervised normalizing flows. *Nature Machine Intelligence*, 6(11), 1370-1381.

D. Rezende, S. Mohamed. Variational inference with normalizing flows. *ICML*, 2015.

J.L. Wadsworth. Recent perspectives on multivariate extremes. *JRSS-B*, 2024.

L. Rudi, A. Ghosal, A. Kroese. Generative models for rare event simulation. *Annals of Applied Probability*, 2023.

McDonald, A., Tan, P. N., Luo, L. (2022). Comet flows: Towards generative modeling of multivariate extremes and tail dependence. *arXiv preprint*.

Gao, Z., Zhang, D., Daniel, L., Boning, D. S. (2023). Rare event probability learning by normalizing flows. *arXiv preprint*.

Collaborations envisagées

ISAE-SUPAERO/DISC (Benjamin Bobbia)

Laboratoire d'accueil à l'ONERA

Département : Traitement de l'information et Systèmes

Lieu (centre ONERA) : Toulouse

Contact : Jérôme Morio

Tél. : 05 62 25 26 63 Email : jerome.morio@onera.fr

Directeur de thèse

Nom : Jérôme Morio

Laboratoire : ONERA/DTIS

Tél. : 05 62 25 26 63

Email : jerome.morio@onera.fr

Pour plus d'informations : <https://www.onera.fr/rejoindre-onera/la-formation-par-la-recherche>