

PROPOSITION DE STAGE EN COURS D'ETUDES

Référence : **DTIS-2024-04**
(à rappeler dans toute correspondance)

Lieu : Toulouse

Département/Dir./Serv. : DTIS/RIME

Tél. : 05 62 25 27 55

Responsable(s) du stage : Jérôme Morio

Email : jerome.morio@onera.fr

DESCRIPTION DU STAGE

Thématique(s) : Mathématiques appliquées et leurs interactions, calcul scientifique

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

Intitulé : Estimation d'indices de Shapley à l'aide d'auto-encodeurs pour l'analyse de sensibilité fiabiliste

Sujet : De nombreux systèmes physiques sont schématiquement décrits par une relation du type $Y = \phi(X)$, où l'entrée multidimensionnelle X est supposée aléatoire et où la sortie Y est déterminée via la fonction déterministe ϕ . Un exemple proéminent d'application est l'analyse d'un code de calcul boîte noire : ϕ représente alors un code de calcul, tel que des calculs de contraintes sur des structures mécaniques complexes et X les conditions extérieures dans lesquelles ce calcul est effectué. On peut notamment penser à un code de type éléments finis, dont la complexité rend impossible toute étude analytique de la fonction ϕ et donc de la sortie Y .

L'analyse de fiabilité caractérise la probabilité de défaillance P_f du système étudié, souvent modélisée par la probabilité que Y dépasse un seuil S . Dans un cadre probabiliste, cette estimation de probabilité est aujourd'hui globalement bien maîtrisée et il existe de nombreuses méthodes permettant ce type d'estimation de probabilité (méthodes Monte-Carlo, échantillonnage préférentiel [1], subset simulation [2]). Par ailleurs, l'analyse de sensibilité détermine quelles sont les composantes de l'entrée X qui influent le plus sur la sortie Y afin de, par exemple, réduire la dimension du modèle ϕ . Un exemple relativement connu d'indice de sensibilité sont les indices de Sobol qui analysent la part de variance de Y issue des différentes composantes de X [3].

Le cadre de ce stage combine ces deux techniques afin d'effectuer une analyse de sensibilité fiabiliste qui consiste à déterminer quelles sont les dimensions de l'entrée les plus influentes sur la probabilité de défaillance. Cette question est particulièrement complexe car on cherche à obtenir les résultats de l'analyse de sensibilité fiabiliste en minimisant les coûts de calculs s'ajoutant à l'estimation de la probabilité de défaillance.

Une piste intéressante de recherche pour le stage concerne ainsi les indices de Shapley [4] qui sont utilisés en analyse de sensibilité classique. Si leur estimation était auparavant relativement coûteuse, de très récentes approches proposant de nouveaux algorithmes d'estimation ont largement permis de réduire le nombre d'échantillons nécessaires à leur estimation [5,6] et particulièrement dans un cadre fiabiliste [7,8]. Néanmoins, ces approches atteignent leurs limites quand la dimension de l'entrée X est trop importante ou quand la probabilité de défaillance est trop faible.

L'objectif du stage est d'estimer des indices de Shapley en analyse de sensibilité fiabiliste en minimisant le coût de calcul nécessaire dans un contexte de grande dimension. Dans ce contexte, de récents travaux soumis autour des auto-encodeurs variationnels [9] utilisés comme des échantillonneurs non paramétriques [10] sont une piste intéressante pour estimer les indices Shapley fiabiliste en grande dimension. Combinés aux approches par subset simulation [2] pour estimer la probabilité de défaillance, des cas-test où la dimension de X dépasse la centaine, pourraient devenir tout à fait envisageable pour effectuer une analyse de sensibilité fiabiliste.

L'algorithme d'estimation des indices de Shapley obtenu durant le stage sera appliqué sur différents cas-tests aéronautiques.

Le stage sera co-encadré par François Bachoc, enseignant-chercheur à l'Institut de Mathématiques de Toulouse.

Possibilité de thèse : une bourse de thèse de doctorat de l'Institut d'IA de Toulouse (ANITI) est prévue à la suite du stage avec les mêmes encadrants.

[1] J. A. Bucklew. Introduction to rare event simulation. Springer Series in Statistics. Springer-Verlag, New York, 2004.

[2] F. Cérou, P. Del Moral, T. Furon, and A. Guyader. Sequential Monte Carlo for Rare Event Estimation, Statistics and Computing, 22(3):795-808, 2012.

[3] B. Iooss, P. Lemaître. A review on global sensitivity analysis methods. In Uncertainty management in simulation-optimization of complex systems (pp. 101-122). Springer, Boston, 2015

[4] A. B. Owen, C. Prieur. On Shapley value for measuring importance of dependent inputs. SIAM/ASA Journal on Uncertainty Quantification, 5(1), 986-1002, 2017.

[5] F. Gamboa, P. Gremaud, T. Klein, A. Lagnoux. Global Sensitivity Analysis: a new generation of mighty estimators based on rank statistics. arXiv preprint arXiv:2003.01772, 2020.

[6] B. Broto, F. Bachoc, M. Depecker. Variance reduction for estimation of Shapley effects and adaptation to unknown input distribution. SIAM/ASA Journal on Uncertainty Quantification, 8(2), 693-716, 2020.

[7] M. Il Idrissi, V. Chabridon, B. Iooss, B. Developments and applications of Shapley effects to reliability-oriented sensitivity analysis with correlated inputs. Environmental Modelling & Software, 143, 105115, 2021.

[8] J. Demange-Chryst, F. Bachoc, J. Morio. Shapley effect estimation in reliability-oriented sensitivity analysis with correlated inputs by importance sampling. International Journal for Uncertainty Quantification, 13(3), 2023.

[9] D. P. Kingma, M. Welling. An introduction to variational autoencoders. Foundations and Trends® in Machine Learning, 12(4), 307-392, 2019.

[10] J. Demange-Chryst, F. Bachoc, J. Morio. Variational autoencoder with weighted sample for high-dimensional non-parametric adaptive importance sampling, to be submitted.

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

Méthodes à mettre en oeuvre :

- | | |
|---|--|
| <input checked="" type="checkbox"/> Recherche théorique | <input type="checkbox"/> Travail de synthèse |
| <input checked="" type="checkbox"/> Recherche appliquée | <input type="checkbox"/> Travail de documentation |
| <input type="checkbox"/> Recherche expérimentale | <input type="checkbox"/> Participation à une réalisation |

Possibilité de prolongation en thèse : **Oui**

Durée du stage : Minimum : 4 Maximum : 6

Période souhaitée : Début du stage au premier semestre 2024

PROFIL DU STAGIAIRE

Connaissances et niveau requis : Mathématiques appliquées, probabilités-statistiques	Ecoles ou établissements souhaités : 3ème année Ecole d'Ingénieur généraliste, Master 2
---	--