

PROPOSITION DE STAGE EN COURS D'ETUDES

Référence : **DTIS -2019-11**

(à rappeler dans toute correspondance)

Lieu : Toulouse

Département/Dir./Serv. : DTIS/RIME/M2CI

Tél. : 05 62 25 27 55 /26 63

Responsable(s) du stage : Sylvain Dubreuil,
Jérôme Morio

Email. : sylvain.dubreuil@onera.fr,
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

Intitulé : Estimation de probabilité d'évènements rares par échantillonnage préférentiel non paramétrique

Sujet : De nombreux systèmes physiques sont schématiquement décrits par une relation du type $Y = \phi(X)$, où l'entrée 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 paramètres de la simulation (paramètres matériaux, géométriques, conditions aux limites etc.). 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 .

Une question importante est la fiabilité d'un tel système. Cela correspond à la problématique d'estimation de la probabilité d'un évènement rare de la sortie Y : la défaillance d'un système est en effet un évènement de faible occurrence mais qu'il est indispensable d'estimer pour des raisons de fiabilité et de certification. Les techniques naïves de type Monte-Carlo sont mal adaptées à de tels cas, nécessitant par exemple 10^8 appels à la fonction ϕ pour estimer une probabilité de l'ordre de 10^{-6} avec une erreur relative de l'ordre de 10%. Des techniques plus efficaces ont été développées telles que les méthodes de Monte-Carlo séquentielles [1] ou l'échantillonnage préférentiel [2].

Le principe des techniques d'échantillonnage préférentiel est de proposer une distribution auxiliaire de l'entrée X capable de générer plus d'apparition d'évènements rares de la sortie Y . Le changement de densité de tirage des échantillons est pris en compte dans l'estimateur de probabilité de défaillance par la pondération des échantillons. Pour déterminer une densité efficace de tirage, plusieurs approches paramétriques et non paramétriques sont disponibles. Le stage se concentrera sur la méthode d'échantillonnage préférentiel adaptatif non paramétrique à noyaux pondérés car celle-ci est particulièrement robuste au non linéarité de la région de défaillance [3,4,5]. Toutefois cette approche n'est rapidement plus efficace lorsque la dimension de X est supérieure à 10.

L'objectif du stage est de proposer un nouvel algorithme d'échantillonnage préférentiel adaptatif semi paramétrique capable de gérer des vecteurs d'entrée X de grande dimension (>10). Dans ce but, une approche possible est de raffiner la modélisation de la dépendance entre échantillons à l'aide de copules tout en conservant la structure non paramétrique sur les marginales de la densité auxiliaire. Une copule est une fonction de répartition qui permet de modéliser la structure de dépendance entre échantillons [6]. Pour des échantillons de dimension importante, les modèles de copules en vignes sont particulièrement flexibles. Celles-ci se basent en effet sur la décomposition de la structure de dépendance totale du vecteur X en un produit de copule bidimensionnelle plus simple à identifier. La spécification complète d'une copule en vignes nécessite le choix d'une structure hiérarchique de l'arbre de dépendance en vignes, de la famille de copule pour chaque paire et de leurs paramètres respectifs. L'étude des différentes approches de construction de copule vigne pour la mise en œuvre d'une approche d'échantillonnage préférentiel semi paramétrique constituera un point important du stage.

Selon les avancées du projet, le stage pourra être décomposé de la façon suivante :

1) Etude de l'échantillonnage préférentiel non paramétrique et de ses limites

- 2) Modélisation semi paramétrique de la densité auxiliaire par une copule gaussienne globale
- 3) Extension de l'approche semi paramétrique par une description plus fine de la structure de dépendance (utilisation des copules vignes)
- 4) Mise en œuvre et test de l'approche proposée sur un cas test aéronautique (fiabilité d'une structure d'aile d'avion, simulation aéro-élastique).

[1] 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.

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

[3] P. Zhang. Nonparametric importance sampling. *Journal of the American Statistical Association*, 91(435):1245-1253, 1996

[4] J. Morio. Extreme quantile estimation with nonparametric adaptive importance sampling. *Simulation Modelling Practice and Theory*, 27:76-89, 2012

[5] M. Kohler, A. Krzyżak, R. Tent, and H. Walk, Nonparametric quantile estimation using importance sampling. *Annals of the Institute of Statistical Mathematics*, 70(2): 439-465, 2018

[6] H. Joe and D. Kurowicka (Eds.). Dependence modeling: vine copula handbook. World Scientific, 2011
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 mois Maximum : 6 mois

Période souhaitée : Mars – Septembre 2019

PROFIL DU STAGIAIRE

Connaissances et niveau requis : Mathématiques appliquées, probabilités / statistiques	Ecoles ou établissements souhaités : 3ème année Ecole d'Ingénieur généraliste ou M2R
---	---