



Échantillonnage préférentiel en grande dimension via des projections dans un sous-espace de petite dimension.

Soutenance de thèse – Maxime El Masri
Mercredi 16 Mars 2022 à 13h30
Salle des thèses de l'ISAE-Supaero, Toulouse

En raison de la crise sanitaire, la soutenance aura également lieu en visioconférence via zoom. Pour recevoir le lien de connexion, veuillez contacter : melmasri@onera.fr.

Devant le jury composé de :

M. Jean-Michel Marin	Université de Montpellier	Rapporteur
M. Bruno Tuffin	INRIA-Rennes	Rapporteur
M. Jean-Marc Bourinet	SIGMA-Clermont	Examinateur
Mme Gersende Fort	Institut de Mathématiques de Toulouse	Examinatrice
M. Olivier Zahm	INRIA-Grenoble	Examinateur
M. Jérôme Morio	ONERA, Toulouse	Directeur de thèse
M. Florian Simatos	ISAE Supaero	Directeur de thèse

Résumé

De nombreuses disciplines scientifiques s'intéressent à l'estimation d'espérances d'une fonction d'intérêt selon une certaine loi de probabilité. Cette fonction peut être considérée comme une boîte noire, potentiellement coûteuse à évaluer. Une méthode couramment utilisée pour estimer des espérances, tout en limitant le nombre d'appels à la boîte noire, est la méthode d'échantillonnage préférentiel (Importance Sampling, IS) qui consiste à échantillonner selon une loi de probabilité auxiliaire au lieu de la loi initiale. Pour annuler la variance de l'estimateur d'IS, il existe une densité optimale théorique, mais celle-ci est inutilisable en pratique. En revanche, des algorithmes adaptatifs (Adaptive Importance Sampling, AIS) ont été développés pour approcher cette densité théorique par des densités paramétriques, en mettant à jour des paramètres de manière itérative.

Mais lorsque la dimension de l'espace des paramètres augmente, l'estimation des paramètres se dégrade et les algorithmes d'AIS, et l'IS en général, deviennent inefficaces. L'estimation finale de l'espérance devient alors très imprécise, notamment du fait de l'accumulation des erreurs commises dans l'estimation de chaque paramètre.

L'objectif principal de cette thèse est ainsi d'améliorer la précision de l'IS en grande dimension, en réduisant le nombre de paramètres estimés à l'aide de projections dans un sous-espace de petite dimension. Nous nous concentrons particulièrement sur la recherche de directions de projection influentes pour l'estimation de la matrice de covariance dans un cadre gaussien unimodal (où l'on met à jour le vecteur moyenne et la covariance). La première piste explorée est la projection sur le sous-espace de dimension un engendré par la moyenne optimale.

Cette direction est particulièrement pertinente dans le cas d'estimation d'une probabilité d'événement rare, car la variance semble diminuer selon cette direction. La seconde proposition correspond à la projection optimale obtenue en minimisant la divergence de Kullback-Leibler avec la densité visée. Cette seconde proposition permet de projeter dans un espace de plusieurs dimensions contrairement à la première, et permet d'identifier les directions les plus influentes. Dans un premier temps, l'efficacité de ces projections est testée sur différents exemples d'estimation d'espérances en grande dimension, dans un cadre théorique n'impliquant pas d'algorithmes adaptatifs. Ensuite, nous proposons un couplage de ces projections avec l'algorithme d'Entropie Croisée (Cross Entropy, CE), un algorithme d'AIS destiné à l'estimation de probabilités d'événements rares. L'efficacité de ces algorithmes est vérifiée sur plusieurs cas-tests avec un faible budget de simulation. Dans tous les cas, les simulations montrent que les méthodes proposées sont plus précises que la CE classique en grande dimension avec un même budget.

Mots-clés

Estimation, échantillonnage préférentiel, grande dimension, projection