

Proposition de thèse en probabilités et statistique mathématique

Titre : Modélisation de la dépendance stochastique sous contraintes

Financement : EDF - R&D - Département Management des Risques Industriels.

Laboratoire d'accueil : Laboratoire de mathématiques et de leurs applications, université de Pau et des Pays de l'Adour, France (<http://lma-umr5142.univ-pau.fr>).

Durée de l'allocation : 3 ans.

Déroulement : La première année se déroulera probablement à Pau. Lors des deux années suivantes, le doctorant partagera son temps entre Clamart (EDF R&D) et Pau.

Pré-requis : Master recherche en mathématiques appliquées, probabilités, statistique ou similaire, grande école d'ingénieur.

Mots-clés : probabilités, modélisation de lois multivariées, dépendance stochastique, traitement des incertitudes, statistique mathématique, estimation, processus empiriques, probabilités et statistique computationnelles, python, R.

Encadrants : Anne Dutfoy, chercheur à EDF R&D, département Management des Risques Industriels, et Ivan Kojadinovic, professeur au laboratoire de mathématiques et de leurs applications, université de Pau et des Pays de l'Adour, France.

Contacts : Anne Dutfoy et Ivan Kojadinovic

Courriels : anne.dutfoy@edf.fr et ivan.kojadinovic@univ-pau.fr.

Tél : + 33 (0)5 59 40 71 33.

1 Problématique scientifique et industrielle

La simulation numérique met en oeuvre des modèles pour faire des prédictions, des analyses d'incertitudes et de sensibilité, ou encore pour résoudre des problèmes d'optimisation. Or ces modèles sont de plus en plus complexes car faisant appel à un nombre potentiellement très grand de variables d'entrée et parce qu'ils sont implémentés dans des codes de simulation souvent très coûteux en temps de calcul. Aussi, les techniques d'échantillonnage des données constituent un challenge certain : une exploration *optimale* du domaine des entrées permet de tirer des informations sur les variables de sortie du code considéré comme 'boîte noire'.

Le traitement des incertitudes [3, 6] des entrées d'un code de simulation est un domaine qui nécessite une première étape de modélisation probabiliste du vecteur aléatoire formé des variables d'entrée incertaines du code. Lorsque les variables ne sont pas supposées indépendantes, la structure de dépendance doit être modélisée : en plus des lois marginales, il est nécessaire de déterminer la *copule* du vecteur aléatoire.

2 Cadre mathématique

Les n variables incertaines d'entrée sont modélisées par un vecteur aléatoire de dimension n , noté \underline{X} . On s'intéresse dans ce projet à la définition de lois probabilistes multivariées garantissant le respect de contraintes telles que :

$$\underline{X} \in \mathcal{D} \text{ presque sûrement,} \quad (1)$$

ou de type :

$$Prob(\underline{X} \in \mathcal{D}) > p, \quad p < 1 \quad (2)$$

où \mathcal{D} est un domaine de \mathbb{R}^n .

Les contraintes peuvent être linéaires avec :

$$\mathcal{D} = \{\underline{X} \in \mathbb{R}^n \mid \underline{M}\underline{X} \leq \underline{B}\} \quad (3)$$

avec $\underline{M} \in \mathcal{M}_{pn}(\mathbb{R})$, $\underline{B} \in \mathbb{R}^p$.

Elles peuvent être aussi non linéaires avec :

$$\mathcal{D} = \{\underline{X} \in \mathbb{R}^n \mid G(\underline{X}) \leq \underline{0}\} \quad (4)$$

avec $G : \mathbb{R}^n \rightarrow \mathbb{R}^p$ non linéaire.

La loi multivariée de \underline{X} peut être représentée par sa fonction de répartition (f.d.r.) décomposée selon le théorème de Sklar [5]. Plus précisément, soit F la f.d.r. de \underline{X} et soient F_1, \dots, F_n les f.d.r.s marginales associées, supposées continues. Il existe alors une unique f.d.r. C sur $[0, 1]^n$ dont les marges sont uniformes, appelée *copule*, telle que

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)), \quad x_1, \dots, x_n \in \mathbb{R}.$$

La représentation précédente permet d'isoler la dépendance entre les composantes X_1, \dots, X_n de \underline{X} , caractérisée par la copule C , des comportements marginaux caractérisés par les f.d.r. marginales F_1, \dots, F_n .

On peut montrer que les contraintes (1) et (2) engendrent des contraintes de compatibilité sur les marginales. Elles engendrent de plus, à marginales imposées compatibles au sens précédent, des contraintes sur le support de la copule.

La référence [4] expose des premiers résultats mathématiques obtenus dans le cadre de contraintes de type $\phi_1(X_1) \leq \dots \leq \phi_n(X_n)$ avec les fonctions ϕ_i monotones : elle formule un théorème d'existence et de caractérisation de lois multivariées définies par une copule et des marginales vérifiant les contraintes d'inégalité, en explicitant les relations de compatibilité entre marginales et en caractérisant le support des copules admissibles.

Dans le cas particulier où la contrainte s'écrit $X_1 \leq \dots \leq X_n$, le problème revient à définir des lois multivariées de statistique d'ordre à marginales imposées. Dans ce cas, les contraintes sur les marginales sont déjà bien connues dans la littérature. En revanche, les contraintes sur les copules ne le sont pas.

Dans le cas de contraintes plus générales, même les contraintes de compatibilité entre les lois marginales restent un sujet de recherche.

3 Motivation industrielle du sujet

Le sujet de thèse adresse plusieurs domaines d'application :

- **Mécanismes d'assemblage avec jeu** : dans le cadre des assemblages de pièces, on doit prendre en compte des contraintes de non emboîtement des pièces entre elles. Les déplacements des pièces entre elles étant petits, les contraintes peuvent être linéarisées et représentées par une relation du type (3).
 - **Génie civil** : dans le cadre du génie civil, les assemblages sont définis avec des tolérances plus grandes qui ne permettent pas de linéarisation. Il est donc nécessaire de prendre en compte le terme du second degré, ce qui mène à des contraintes de type (4) quadratiques.
 - **Robotique** : dans le cadre de la robotique, dans les problèmes d'évitement d'obstacle au cours d'un mouvement incertain, on rencontre les contraintes de type (4). Par exemple, pour la modélisation de la trajectoire du bras d'un robot motorisé et dont l'asservissement n'est pas très précis.
 - **Stratification géologique** : la modélisation d'une couche géologique doit respecter certaines contraintes sur les paramètres d'environnement. Par exemple, les coefficients de perméabilité des couches modélisées doivent respecter des contraintes d'inégalité.
 - **Ségrégation des fluides d'un mélange** : dans un mélange de fluides non homogènes, la modélisation doit respecter le principe selon lequel le fluide le moins dense se retrouve au-dessus du mélange.
 - **Thermo-mécanique** : en fonction de la température, les propriétés mécaniques d'une structure augmentent ou diminuent, ce qui peut revenir à prendre en compte des contraintes de type (3) pour les cas de variations monotones du paramètre.
- Les références [1, 7] exposent d'autres exemples où de telles contraintes apparaissent.

4 Aspects de la thèse

Le sujet présente plusieurs thèmes, probabilistes et statistiques.

Aspects théoriques probabilistes :

- détermination des contraintes sur les marginales engendrées par les contraintes de type (1) ou (2),
- théorème d'existence de telles copules,
- expression des liens entre marginales et copules,
- existence et construction des copules maximales selon différents critères d'intérêt : maximum d'entropie, maximum de vraisemblance, ...

Aspects théoriques statistiques :

- construction d'estimateurs paramétriques et non paramétriques de modèles joints sous contraintes à l'aide de techniques de processus empiriques,
- étude de la qualité des estimateurs,
- étude de techniques de rééchantillonnage (*bootstrap*) pour la construction d'intervalles de confiance à l'aide de techniques de processus empiriques.

Aspects algorithmiques : construction d'algorithmes de simulation de modèles multivariés sous contraintes (n'exhibant pas nécessairement le modèle). Par exemple, par conditionnement successifs des variables marginales avec l'étude de l'influence de l'ordre du conditionnement, ...

Aspects Cas d'application : les recherches théoriques seront mises en application sur un cas industriel à enjeux pour EDF.

Références

- [1] O. Asserin and A. Loredó and M. Petelet and B. Iooss, *Global sensitivity analysis in welding simulations - What are the material data you really need?*, Finite Elements in Analysis and Design, 47(2011), pp 1004-1016.
- [2] E. Borgonovo, *Sensitivity analysis of model output with input constraints : A generalized rationale for local methods* Risk Analysis, 28(2008), pp 667-680.
- [3] www.openturns.org, *Reference guide for Uncertainty treatment*
- [4] R. Lebrun, A. Dutfoy, *Copulas for order statistics with prescribed marginal distribution functions*, Journal of Multivariate Analysis, (2012), under review.
- [5] R.B. Nelsen, *An introduction to copulas*, Springer, (2006).
- [6] A. Pasanisi, A. Dutfoy, *An Industrial Viewpoint on Uncertainty Quantification In Simulation : Stakes, Methods, Tools, Examples* Lecture Notes in Computer Science (in press), (2012).
- [7] E. Volkova and B. Iooss and F. Van Dorpe, *Global sensitivity analysis for a numerical model of radionuclide migration from the RRC "Kurchatov Institute" radwaste disposal site*, Stochastic Environmental Research and Risk Assessment, 22(2008), pp 17-31.