

# Estimation des indices de Shapley en analyse de sensibilité fiabiliste en présence d'entrées dépendantes par échantillonnage préférentiel

Julien DEMANGE-CHRYST<sup>a,b</sup>, François BACHOC<sup>b</sup>, Jérôme MORIO<sup>a</sup>

<sup>a</sup>ONERA/DTIS, Université de Toulouse, F-31055 Toulouse, France

<sup>b</sup>Institut de Mathématiques de Toulouse, UMR5219 CNRS, 31062 Toulouse, France

---

*Mots clés* : Analyse de sensibilité, indices de Shapley, estimation d'évènements rares, échantillonnage préférentiel.

---

De nombreux systèmes physiques sont schématiquement décrits par une relation du type  $Y = \phi(\mathbf{X})$ , où l'entrée  $\mathbf{X}$  de dimension  $d$  est supposée aléatoire et où la sortie  $Y$  est déterminée via la fonction déterministe  $\phi$ . Un exemple d'application est l'analyse d'un code de calcul de type boîte noire, dont la complexité rend impossible toute étude analytique de la fonction  $\phi$ . De plus, les appels au code  $\phi$  sont généralement extrêmement coûteux et ne peuvent donc être fait qu'en nombre limité.

Pour des raisons de sécurité et/ou de certification, l'**étude de la fiabilité** d'un tel système est primordiale. Sans perte de généralité, la défaillance peut être décrite comme un dépassement de seuil de la sortie  $\{Y > t\}$ . Il s'agit d'un évènement rare, et il est crucial d'estimer précisément sa probabilité. Les techniques naïves par échantillonnage de Monte-Carlo sont alors mal adaptées à une telle estimation lorsque cette probabilité est très faible car leur convergence nécessite des budgets de simulations souvent trop importants. Des techniques plus efficaces telles que les méthodes de simulations multi-niveaux [3] ou l'échantillonnage préférentiel [2] ont été développées et permettent aujourd'hui de globalement bien maîtriser cette question.

Par ailleurs, une autre problématique importante de l'étude d'un code de calcul est l'**analyse de sensibilité**, qui consiste à étudier l'influence des composantes du vecteur d'entrée  $\mathbf{X}$  sur la sortie  $Y$ , afin par exemple de réduire la dimension du problème. Il existe de nombreuses méthodes locales et globales d'analyse de sensibilité. Le calcul des indices de Sobol est l'une des plus connue. Les indices de Sobol [6] sont des indices de sensibilité globaux qui permettent d'évaluer quantitativement l'influence des différentes composantes de  $\mathbf{X}$  sur  $Y$ . Cependant, ces indices atteignent leurs limites lorsque les entrées sont dépendantes, car bien qu'ils soient toujours calculables, ils ne permettent plus d'identifier clairement l'origine de la variabilité de la sortie  $Y$ . C'est pourquoi, récemment, en effectuant une analogie avec la théorie des jeux, il a été proposé de calculer des indices de Shapley pour l'analyse de sensibilité [5]. En revanche, leur estimation est difficile car les approches naïves requièrent un budget de simulation souvent rédhibitoire. Des approches récentes ont néanmoins permis une réduction drastique du coût de calcul par des méthodes ne nécessitant plus que la connaissance d'un échantillon entrée-sortie [1].

Nous nous proposons de combiner ces deux méthodes afin d'effectuer une **analyse de sensibilité fiabiliste** du système, qui consiste à étudier l'influence de chaque composante de  $\mathbf{X}$  sur la variable aléatoire  $\mathbf{1}(\phi(\mathbf{X}) > t)$ , sous l'hypothèse de dépendance entre les entrées. 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. L'estimation des indices de Shapley sur la variable aléatoire  $\mathbf{1}(\phi(\mathbf{X}) > t)$  est particulièrement intéressante car elle permet de cibler les composantes de  $\mathbf{X}$  dont la réduction de variance réduira l'incertitude sur  $\mathbf{1}(\phi(\mathbf{X}) > t)$ . Ce type d'indice et leur estimation ont été introduits très récemment par [4]. Des estimateurs ne nécessitant qu'un échantillon entrée-sortie  $y$  sont proposés et permettent d'estimer les indices de Shapley fiabilistes sans appels supplémentaires à la fonction  $\phi$  après l'estimation de la probabilité de défaillance par méthode de Monte-Carlo classique. Cependant, ces estimateurs atteignent leurs limites lorsque cette dernière est très faible car leur convergence nécessite un échantillon de taille trop importante. Dans cette communication, nous présentons de nouveaux estimateurs à plus faible variance des indices de Shapley fiabilistes lorsque la probabilité de défaillance est estimée par échantillonnage préférentiel. Ces estimateurs ne nécessitent qu'un échantillon entrée-sortie et ne requièrent donc toujours pas d'appels supplémentaires au code  $\phi$  après l'estimation de la probabilité de défaillance.

## Références

- [1] Baptiste Broto, François Bachoc, and Marine 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.
- [2] James Bucklew. *Introduction to rare event simulation*. Springer Science & Business Media, 2004.
- [3] Frédéric Cérou, Pierre Del Moral, Teddy Furon, and Arnaud Guyader. Sequential monte carlo for rare event estimation. *Statistics and computing*, 22(3) :795–808, 2012.
- [4] Marouane El Idrissi, Vincent Chabridon, and Bertrand Iooss. Developments and applications of shapley effects to reliability-oriented sensitivity analysis with correlated inputs. *Environmental Modelling & Software*, 143 :105115, 2021.
- [5] Art B Owen. Sobol'indices and shapley value. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1) :245–251, 2014.
- [6] Ilya M Sobol. Sensitivity analysis for non-linear mathematical models. *Mathematical modelling and computational experiment*, 1 :407–414, 1993.