Échantillonnage préférentiel dynamique informé par des graphes

Guillaume Chennetier [1,2] & Hassane Chraibi [1] & Anne Dutfoy [1] & Josselin Garnier [2]

[1] *EDF Lab Paris-Saclay, France.*
{guillaume.chennetier,hassane.chraibi,anne.dutfoy}@edf.fr
[2] *CMAP, CNRS, Ecole polytechnique, Institut Polytechnique de Paris, France.*
{guillaume.chennetier,josselin.garnier}@polytechnique.edu

Résumé. L'échantillonnage préférentiel est l'une des méthodes de réduction de variance les plus populaires pour l'estimation de quantités de la forme $\mathbb{E}_{X\sim\mathbf{p}}[\varphi(X)]$. Déterminer une distribution d'importance efficace est cependant notoirement délicat en grande dimension. Un cas typique de grande dimension, pourtant relativement peu traité, est celui où X ne représente pas un vecteur mais la trajectoire d'un processus stochastique. Nous avons en tête l'évaluation de la fiabilité de systèmes industriels complexes dont le fonctionnement est modélisé par des processus stochastiques. Nous proposons une nouvelle famille de distributions d'importance adaptées à la simulation d'événements rares pour des processus de Markov non-diffusifs, c'est-à-dire des processus de Markov déterministes par morceaux (abrégés PDMPs). Ces processus évoluent selon des équations différentielles déterministes dont les paramètres sont soumis à des sauts aléatoires. La distribution d'importance optimale pour ces PDMPs est caractérisée par la fonction dite “committor” du processus. Celle-ci associe à toute trajectoire partielle du PDMP, la probabilité que la trajectoire complète réalise l'événement d'intérêt sachant son passé. Notre méthodologie s'articule en trois phases. On approxime d'abord notre PDMP par une marche aléatoire homogène sur un graphe pour laquelle on peut calculer explicitement des temps moyens d'atteinte. On construit ensuite une famille d'approximations de la fonction committor à partir de ces temps d'atteinte. Enfin, on détermine séquentiellement un bon candidat au sein de cette famille (et par conséquent une densité d'importance efficace) en minimisant un critère d'entropie croisée.

Mots-clés. Monte-Carlo, Échantillonnage préférentiel adaptatif, Simulation d'événements rares, Processus de Markov déterministes par morceaux, Marche aléatoire sur graphes.

Abstract. Importance sampling is one of the most popular variance reduction methods for estimating quantities of the form $\mathbb{E}_{X\sim\mathbf{p}}[\varphi(X)]$. However, determining an effective importance distribution is notoriously challenging in high dimensions. A typical case of high dimensionality, yet relatively underexplored, is when X does not represent a vector but the trajectory of a stochastic process. We have in mind the reliability assessment of complex industrial systems whose operation is modeled by stochastic processes. We propose a new family of importance distributions tailored for rare event simulation with any non-diffusive Markov processes, i.e., any piecewise deterministic Markov processes (abbreviated PDMPs). These processes evolve according to deterministic differential equations whose parameters are subject to random jumps. The optimal importance distribution for these PDMPs is characterized by the so-called “committor” function of the process. This function associates with any partial trajectory of the PDMP the probability that the complete trajectory realizes the event of interest, given its past. Our methodology consists of three phases: first, we approximate our PDMP by a homogeneous random walk on a graph for which mean hitting times can be explicitly computed; then, we construct a family of committor function approximations based on these hitting times; finally, we sequentially determine a good candidate within this family (and consequently an effective importance density) by minimizing a cross-entropy criterion.

Keywords. Monte Carlo, Adaptive Importance Sampling, Rare event simulation, Piecewise deterministic Markov processes, Random walk on graphs.

1 Cadre et limites des méthodes de Monte-Carlo

On se place dans le cadre de l'inférence par simulation, où l'on tente d'estimer des quantités de la forme $\overline{\varphi} := \mathbb{E}_{X\sim\mathbf{p}}\left[\varphi(X)\right]$ avec $\mathbf{p}$ pour distribution de référence et φ une fonction d'intérêt. Une méthode de Monte-Carlo est alors l'approche la plus naturelle pour estimer une telle quantité : elle consiste à approcher $\overline{\varphi}$ par une moyenne empirique $\frac{1}{n}\sum_{k=1}^{n}\varphi(X_k)$ à partir d'un échantillon $X_1,\ldots,X_n$ i.i.d. que l'on a généré numériquement sous la distribution de référence $\mathbf{p}$. Cette méthode est simple à mettre en oeuvre à condition de pouvoir échantillonner $\mathbf{p}$ à coût raisonnable. Malheureusement, elle se révèle inefficace dès que la distribution $\mathbf{p}$ ne met pas assez de poids là où la fonction d'intérêt φ prend de grandes valeurs (en valeur absolue). La variance de l'estimateur Monte-Carlo devient alors trop grande pour estimer φ avec un budget restreint. Un cas typique qui nous intéresse est celui de la simulation d'événement rare où $\varphi(X) = \mathbb{1}_{X\in\mathbf{F}}$ avec $\mathbf{F}$ un ensemble rarement visité sous la distribution $\mathbf{p}$.

2 Échantillonnage préférentiel

L'échantillonnage préférentiel est l'une des méthodes de réduction de variance les plus connues et peut-être la plus directe. Elle consiste à générer un échantillon $X_1,\ldots,X_n$ sous une distribution alternative $\mathbf{g}$ (dite "distribution d'importance") et à retourner la même moyenne empirique mais cette fois-ci pondérée par le rapport de vraisemblance approprié. On suppose d'une part que $\mathbf{p}$ et $\mathbf{g}$ sont deux densités de probabilités sur un espace $\mathcal{X}$ dominées par une mesure μ, et d'autre part que le support de $\mathbf{g}$ est inclus dans le support de $\varphi\times\mathbf{p}$.

$$\begin{aligned}\overline{\varphi} := \mathbb{E}_{X\sim\mathbf{p}}\left[\varphi(X)\right] &= \int_{\mathcal{X}}\varphi(x)\mathbf{p}(x)\mu(\mathrm{d}x) = \int_{\mathcal{X}}\varphi(x)\frac{\mathbf{p}(x)}{\mathbf{g}(x)}\mathbf{g}(x)\mu(\mathrm{d}x)\\ &= \mathbb{E}_{X\sim\mathbf{g}}\left[\varphi(X)\frac{\mathbf{p}(X)}{\mathbf{g}(X)}\right] \approx \frac{1}{n}\sum_{k=1}^{n}\varphi(X_k)\frac{\mathbf{p}(X_k)}{\mathbf{g}(X_k)} \text{ avec } X_1,\ldots,X_n \overset{\text{i.i.d.}}{\sim} \mathbf{g}.\end{aligned} \tag{1}$$

La distribution d'importance optimale est caractérisée par la densité $\mathbf{g}_{\text{opt}} \propto |\varphi| \times \mathbf{p}$. Elle produit un estimateur de variance nulle lorsque la fonction d'intérêt φ est de signe constant. Cette distribution est bien entendu inaccessible mais elle guide l'objectif en pratique : générer un échantillon $X_1, \ldots, X_n$ sous une distribution $\mathbf{g}$ la plus proche possible de $\mathbf{g}_{\text{opt}}$ pour obtenir un estimateur de $\overline{\varphi}$ de faible variance.

En dehors de cas jouets simples, cet objectif est délicat. On a le plus souvent recours à des méthodes dites adaptatives, où la distribution d'importance est raffinée au cours des simulations. L'une des plus connues, appelée méthode d'entropie croisée [1], consiste à déterminer la distribution d'importance $\mathbf{g} \in \mathcal{G}$ la plus proche de la distribution optimale $\mathbf{g}_{\text{opt}}$ au sens de la divergence de Kullback-Leibler.

$$\begin{aligned}
\underset{\mathbf{g}\in\mathcal{G}}{\arg\min}\, \mathcal{D}_{\text{KL}}\left(\mathbf{g}_{\text{opt}} \| \mathbf{g}\right) &= \underset{\mathbf{g}\in\mathcal{G}}{\arg\min}\ \mathbb{E}_{\mathbf{g}_{\text{opt}}}\left[\log\left(\frac{\mathbf{g}_{\text{opt}}(X)}{\mathbf{g}(X)}\right)\right] \\
&= \underset{\mathbf{g}\in\mathcal{G}}{\arg\max}\ \mathbb{E}_{\mathbf{p}}\left[|\varphi(X)|\, \log\left(\mathbf{g}(X)\right)\right]. \qquad (2)
\end{aligned}$$

Le problème reste notoirement difficile en grande dimension. D'une part car les rapports de vraisemblance "dégénèrent" : ils tendent à prendre des valeurs de plus en plus proches de zéro lorsque la dimension augmente mais restent d'espérance $\mathbb{E}_{\mathbf{g}}\left[\frac{\mathbf{p}(X)}{\mathbf{g}(X)}\right] = 1$. Et d'autre part, la méthode adaptative est confronté au problème d'estimation de densité, très sensible à la dimension de l'espace. Différentes approches ont été proposées pour contourner le problème telles que la réduction de dimension par projection dans des sous-espaces bien choisis [2, 3], ou l'usage de méthodes d'apprentissage modernes que l'on retrouve derrière les modèles génératifs [4].

3 Processus de Markov déterministes par morceaux

La dimension de l'espace ne pose pas uniquement problème lorsque X est un vecteur de grande taille mais aussi lorsqu'il s'agit de la trajectoire d'un processus stochastique. Nous avons typiquement en tête l'évaluation de la fiabilité de systèmes industriels dynamiques dont le fonctionnement est modélisé par des processus stochastiques (voir notre article précédent [5]). On se place dans le cadre général des processus de Markov non-diffusifs qui forment la classe des processus de Markov déterministes par morceaux (abrégés PDMPs).

Un PDMP est un processus stochastique dit hybride car sa variable d'état $X_t = (Z_t, V_t) \in \mathcal{X} = (\mathcal{Z} \times \mathcal{V})$ est constituée d'une partie continue Z_t appelée position et d'une partie discrète V_t appelée régime. La position du PDMP évolue selon un système d'équations différentielles déterministes paramétré par le régime. Lorsque la position atteint la frontière de l'espace d'état, ainsi qu'à des instants aléatoires, le processus saute vers un nouvel état lui aussi choisi aléatoirement. La distribution des instants de sauts et de la destination des sauts dépend du régime et continûment de la position. Introduits par Mark HA Davis dans les années 1980 [6], le lecteur intéressé trouvera une introduction moderne aux PDMPs orientée fiabilité industrielle, dans [7].

On s'intéresse à la probabilité qu'une trajectoire de PDMP $(X_t)_{t\in[0,s_{\max}]}$ de durée $s_{\max}$ atteigne une région critique $\mathbf{F} \subset \mathcal{X}$ (que l'on suppose absorbante) de son espace d'état. Du point de vue applicatif, X représente l'état du système industriel et l'événement $\{X \in \mathbf{F}\}$ correspond à la défaillance critique du système. On peut montrer (voir [8]) que la distribution d'importance optimale dans ce cas de figure se déduit directement de la fonction dite "committor" du processus :

$$\xi^*(x,s) = \mathbb{P}_{\mathbf{p}}\left(\exists t \in [0, s_{\max} - s] \,:\, X_{s+t} \in \mathbf{F} \mid X_s = x\right) \tag{3}$$

Il s'agit simplement de la probabilité que la trajectoire réalise l'événement d'intérêt avant la date $s_{\max}$ sachant son état x à un instant donné s.

4 Approximation de la fonction committor

Cette fonction committor n'étant bien entendu pas connue, nous proposons de la remplacer par une approximation bien choisie. La fonction committor quantifie en un sens la proximité d'un état de l'espace $\mathcal{X}$ à la région $\mathbf{F}$. Une bonne approximation doit au moins pouvoir quantifier la proximité de chaque régime à la région $\mathcal{V}_{\mathbf{F}}$ définie par l'ensemble des régimes permettant d'accéder à $\mathbf{F}$.

On peut remarquer que l'évolution du régime d'un PDMP est une marche aléatoire, non nécessairement Markovienne, sur un graphe. On sait cependant (c'est une conséquence de la formule de Dynkin), que les temps moyens d'atteinte de n'importe quelle région d'un graphe peuvent être calculés explicitement pour une marche aléatoire homogène en résolvant un système linéaire. On se donne donc, avant toute simulation, une marche aléatoire Markovienne et homogène sur le graphe dont les sommets sont les régimes du PDMP, de préférence proche de la marche aléatoire non Markovienne décrite au-dessus. On détermine depuis chaque sommet $v \in \mathcal{V}$ le temps moyen d'atteinte ρ_v de la région $\mathcal{V}_{\mathbf{F}}$ pour cette marche aléatoire homogène.

On construit alors la famille d'approximations $(\xi_{\boldsymbol{\theta}})_{\theta\in\Theta}$ de la fonction committor ξ^*, indexée par un paramètre $\boldsymbol{\theta} \in \Theta$ de dimension arbitraire d_Θ, à partir des temps moyens d'atteinte $(\rho_v)_{v\in\mathcal{V}}$ selon la formule suivante :

$$\xi_{\boldsymbol{\theta}}(v) = \exp\left(-\sum_{i=1}^{d_\Theta} \theta_i \times \rho_v^i\right). \tag{4}$$

À chaque valeur de $\boldsymbol{\theta}$ correspond une approximation $\xi_{\boldsymbol{\theta}}$, à laquelle on peut associer une distribution d'importance possible $\mathbf{g}_{\boldsymbol{\theta}}$. La meilleure distribution d'importance au sein de la famille $(\mathbf{g}_{\boldsymbol{\theta}})_{\boldsymbol{\theta}\in\Theta}$ est déterminée séquentiellement par minimisation d'entropie croisée. Notre méthode permet une réduction de variance d'un facteur supérieur à 10 000 par rapport à une méthode de Monte Carlo standard sur un cas d'essai industriel complexe.

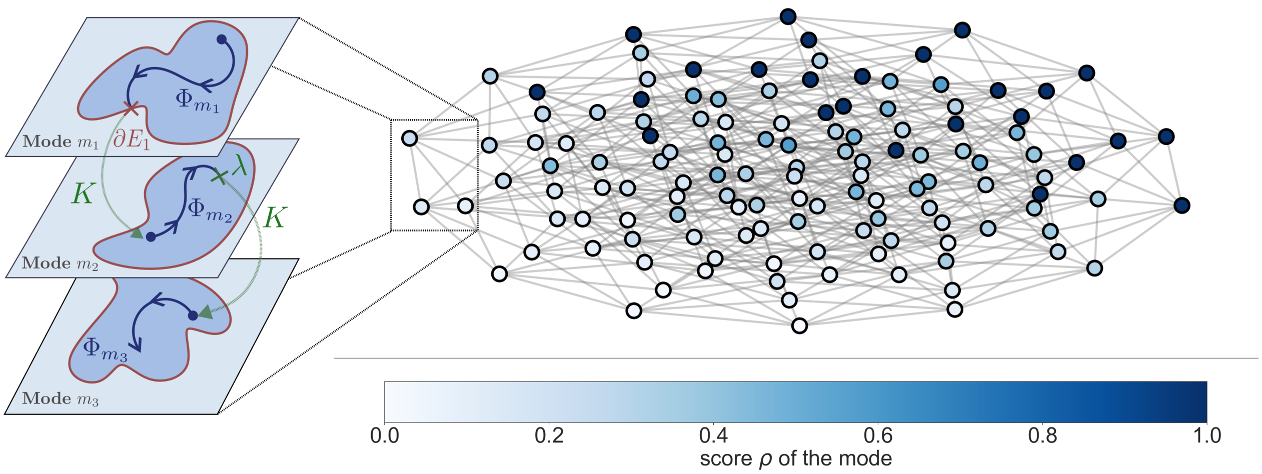

Figure 1: PDMP as a random walk on a graph

References

[1] Reuven Y Rubinstein and Dirk P Kroese. *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning*, volume 133. Springer, 2004.

[2] Maxime El Masri. *High dimensional importance sampling through projections on a low dimensional subspace.* PhD thesis, ISAE-Institut Supérieur de l'Aéronautique et de l'Espace, 2022.

[3] Felipe Uribe, Iason Papaioannou, Youssef M Marzouk, and Daniel Straub. Cross-entropy-based importance sampling with failure-informed dimension reduction for rare event simulation. *SIAM/ASA Journal on Uncertainty Quantification*, 9(2):818–847, 2021.

[4] Julien Demange-Chryst, François Bachoc, Jérôme Morio, and Timothé Krauth. Variational autoencoder with weighted samples for high-dimensional non-parametric adaptive importance sampling. *arXiv preprint arXiv:2310.09194*, 2023.

[5] Guillaume Chennetier, Hassane Chraibi, Anne Dutfoy, and Josselin Garnier. Adaptive importance sampling based on fault tree analysis for piecewise deterministic markov process. *arXiv preprint arXiv:2210.16185*, 2022.

[6] Mark HA Davis. Piecewise-deterministic markov processes: A general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(3):353–376, 1984.

[7] Benoite De Saporta, François Dufour, and Huilong Zhang. *Numerical methods for simulation and optimization of piecewise deterministic Markov processes: application to reliability.* John Wiley & Sons, 2015.

[8] Hassane Chraibi, Anne Dutfoy, Thomas Galtier, and Josselin Garnier. On the optimal importance process for piecewise deterministic markov process. *ESAIM: Probability and Statistics*, 23:893–921, 2019.