



HAL
open science

Rare Event Simulation for Piecewise Deterministic Markov Processes, application in reliability assessment with PyCATSHOO tool.

Guillaume Chennetier

► To cite this version:

Guillaume Chennetier. Rare Event Simulation for Piecewise Deterministic Markov Processes, application in reliability assessment with PyCATSHOO tool.. Methodology [stat.ME]. Institut Polytechnique de Paris, 2024. English. NNT : 2024IPPAX087 . tel-04904282

HAL Id: tel-04904282

<https://theses.hal.science/tel-04904282v1>

Submitted on 21 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT
POLYTECHNIQUE
DE PARIS

NNT : 2024IPPAX087

Thèse de doctorat



Rare event simulation with piecewise deterministic Markov processes.

A reliability application for PyCATSHOO tool

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à l'École polytechnique

École doctorale n°574 École Doctorale de Mathématiques Hadamard (EDMH)
Spécialité de doctorat : Mathématiques appliquées

Thèse présentée et soutenue à Palaiseau, le 24 septembre 2024, par

GUILLAUME CHENNETIER

Composition du Jury :

Randal Douc Professeur, Télécom SudParis	Président
Arnaud Guyader Professeur, Sorbonne Université	Rapporteur
Pierre-Étienne Labeau Professeur, École polytechnique de Bruxelles	Rapporteur
Christophe Berenguer Professeur, Université de Grenoble	Examinateur
Marylou Gabrié Professeure assistante, École polytechnique	Examinatrice
Hassane Chraïbi Ingénieur-chercheur, EDF R&D	Encadrant industriel
Josselin Garnier Professeur, École polytechnique	Directeur de thèse
Anne Dutfoy Ingénieure-Chercheuse, EDF R&D	Invitée

À Jean-Charles Chennetier.

Table of contents

<i>Table of contents</i>	v
--------------------------	---

FOREWORD

<i>Acknowledgments</i>	xi
<i>French summary</i>	xiii
<i>Notations and abbreviations</i>	xxv

STATE OF THE ART

I Monte Carlo simulation and rare events	3
1.1 Introduction to Monte Carlo methods	4
1.1.1 Generating random variables	4
1.1.2 Monte Carlo methods	9
1.2 Particle-based methods	12
1.2.1 Subset simulation	13
1.2.2 Adaptive multilevel splitting	15
1.3 Importance sampling	15
1.3.1 The essentials	17
1.3.2 Adaptive procedures	23
1.3.3 Curse of dimensionality	29
2 Piecewise deterministic Markov processes	33
2.1 Poisson process	34
2.1.1 Introduction to Markovian jump processes	34
2.1.2 The simple Poisson process	39
2.1.3 Inhomogeneous Poisson process	44
2.2 PDMP essentials	46

2.2.1	Local characteristics of the PDMP	47
2.2.2	Strong Markov PDP	52
2.3	Modeling and simulation with PDMPs	56
2.3.1	PDMP superposition	57
2.3.2	PDMP thinning	61
2.3.3	The inversion method with a twist	62
3	Rare event simulation for PDMP	65
3.1	Distribution of a PDMP trajectory	66
3.1.1	Skeleton of a trajectory	66
3.1.2	Space of admissible trajectories	68
3.1.3	Likelihood of a PDMP trajectory	69
3.2	Importance sampling of PDMP	72
3.2.1	Admissible importance process	72
3.2.2	Optimal importance process	75
3.3	Rare event application	77
3.3.1	Committer function	77
3.3.2	Practical implementation	79

CONTRIBUTIONS

4	Adaptive importance sampling strategy	87
4.1	Towards committor function approximation	88
4.1.1	Reminder on importance sampling of PDMP	88
4.1.2	Problem parameterization	89
4.2	General parametric form of the importance function	91
4.2.1	Parametric form of the IF	91
4.2.2	Proximity scores	93
4.2.3	Basis functions	94
4.3	Adaptive importance sampling with recycling scheme	95
4.3.1	Cross entropy with recycling scheme	95
4.3.2	Asymptotic properties	97
4.3.3	Implementation guidelines	103
5	Importance function for multicomponent systems	107
5.1	A first approach for multicomponent systems	108

5.1.1	Multicomponent systems in reliability analysis	108
5.1.2	Coherent systems	110
5.1.3	Broken Components Rate approach	112
5.2	Proximity scores based on fault tree analysis	114
5.2.1	The limits of structure blindness	114
5.2.2	Minimal cut sets and minimal path sets	118
5.2.3	Proximity score and implementation	120
5.3	Numerical experiments	123
5.3.1	Series and parallel systems	123
5.3.2	The spent fuel pool	125
6	Graph-informed importance function	131
6.1	Motivations	132
6.1.1	Broadening the scope of applications	132
6.1.2	Regime as a random walk on a graph	133
6.2	Graph-based mean hitting times	136
6.2.1	Explicit computation of mean hitting times	137
6.2.2	Mean hitting times as proximity scores	142
6.3	Numerical experiments	144
6.3.1	Computation of mean hitting times	145
6.3.2	Application on the spent fuel pool	146
CONCLUSION		
7	Conclusion and perspectives	153
7.1	Conclusion	154
7.1.1	Summary of contributions	154
7.1.2	Discussion	157
7.2	Perspectives	158
7.2.1	Refinements to the AIS Method	158
7.2.2	Bandits: Best Arm Identification	161

APPENDIX

<i>References</i>	165
<i>English summary</i>	185
<i>French conclusion</i>	197

FOREWORD

Acknowledgments

Les remerciements occupent, paraît-il, les lignes les plus lues (les seules lues ?) d'une thèse. Je vais vous décevoir tout de suite : les miens ne sont pas prêts.

Jury, encadrants, collègues, proches et moins proches, je dois cette thèse à votre gentillesse, votre patience, vos bons conseils et votre bonne compagnie.

Je voudrais pouvoir remercier chacun et chacune, dans le détail, mais je dois me contenter ici de remercier tout le monde.

Si vous lisez cette page, c'est sans doute que vous devriez y figurer.

French summary

Disclaimer

In accordance with the regulations of the doctoral school, this summary chapter is written in French. An English translation is available in appendix.

Résumé de la thèse

L'objet de cette thèse est de fournir de nouvelles méthodes d'estimation de probabilités d'événements rares reposant sur la simulation de *processus de Markov déterministes par morceaux* (PDMPs). Cette classe très générale de processus offre la flexibilité nécessaire pour représenter fidèlement des systèmes industriels dynamiques complexes. Elle permet en particulier de modéliser conjointement la dynamique déterministe et continue des variables physiques du système (température, pression, niveaux de liquide, etc.), et la dynamique de saut aléatoire qui régit le changement de statut de ses composants (pannes, réparations, mécanismes de contrôle, etc.). L'enjeu industriel est de permettre à l'outil PyCATSHOO, utilisé par l'entreprise Électricité de France pour ses études probabilistes de sûreté, d'estimer la probabilité de défaillance de tels systèmes efficacement, et avec une précision garantie. Une approche Monte-Carlo classique réclame, à niveau de précision fixé, une quantité de simulations inversement proportionnelle à la probabilité recherchée. Elle n'est donc pas adaptée au cas de systèmes hautement fiables et dont le coût de simulation est élevé. L'*échantillonnage d'importance* est une méthode de réduction de variance populaire en situation d'événements rares. Elle consiste à générer les simulations sous une distribution biaisée favorisant la réalisation de l'événement, et à rectifier le biais a posteriori. Des travaux récents ont proposé un cadre théorique d'implémentation de l'échantillonnage d'importance pour des PDMPs, et mis en évidence le lien existant entre la distribution biaisée optimale et ce qu'on appelle la *fonction committor* du processus.

À l'aide d'outils issus de l'analyse fiabiliste et de la théorie des marches aléatoires sur graphes, de nouvelles familles d'approximations de la fonction committor sont introduites dans cette thèse. La méthodologie proposée est adaptative : une approximation de la fonction committor est construite a priori puis raffinée au cours des simulations d'une procédure d'entropie croisée. Les simulations sont ensuite recyclées pour produire un estimateur par échantillonnage d'importance de la probabilité cible. Les résultats de convergence obtenus permettent la construction d'inter-

valles de confiance asymptotiques malgré la dépendance entre les simulations. Cette méthode produit d'excellents résultats en pratique sur les systèmes industriels testés.

À propos de l'auteur

Je m'appelle Guillaume Chennetier et suis né en 1995 à Nice. Je reste et effectue ma scolarité dans les environs de Nice jusqu'au baccalauréat puis entame mes études supérieures à Toulouse en licence d'économie et mathématiques¹. En troisième année de licence, je bifurque en mathématiques appliquées toujours à Toulouse². Je "monte" ensuite à Paris pour suivre le Master 1 *Mathématiques et applications* à Sorbonne Université³. Je me spécialise en mathématiques aléatoires et m'inscris en Master 2 dans la mention *Statistique* pour l'année universitaire 2019-2020. J'effectue mon stage de fin d'études au sein de la division R&D de l'entreprise EDF (Électricité de France), déjà sur le même sujet et avec le même encadrement que pour la thèse qui suivra.

Cadre de la thèse

Cette thèse est une collaboration industrielle sur le format du dispositif CIFRE⁴. Mon temps de travail, bien qu'entièrement dédié à cette thèse, était donc partagé entre un laboratoire académique et une entreprise.

Environnement industriel

L'entreprise EDF (Électricité de France) est le premier producteur et fournisseur d'électricité en France et en Europe. Elle possède trois centres de R&D (Recherche et Développement) sur le territoire national : les Renardières, Chatou et Paris-Saclay. Mon stage de fin d'études puis ma thèse, se sont déroulés au centre EDF Lab Paris-Saclay au sein du département PERICLES (PERformance et prévention des Risques Industriels du parC par la simuLation et les EtudeS).



J'ai été intégré au groupe i2i *Etudes probabilistes de sûreté et de disponibilité des systèmes*, l'un des huit groupes du département PERICLES. Cette équipe est constituée d'une quinzaine d'agents EDF auxquels s'ajoutent chaque année quelques doctorants et stagiaires. J'étais encadré au sein de

¹À l'époque licence *MASS*, à présent *MLASHS*, à l'Université Toulouse I Capitole.

²Licence *MAP*³ à l'Université Toulouse III Paul Sabatier.

³À l'époque Université Pierre et Marie Curie, toujours Paris VI et surnommée Jussieu.

⁴Conventions Industrielles de Formation par la REcherche.

l'équipe par Hassane CHRAIBI⁵, ingénieur-chercheur et chef de projet, et Anne DUTFOY⁶, chercheuse senior.

Environnement académique

Ma thèse était dirigée par Josselin GARNIER, professeur à l'École polytechnique (école membre de l'Institut polytechnique de Paris).

J'étais doctorant au laboratoire CMAP (Centre de mathématiques appliquées) de l'École polytechnique et membre de l'équipe de statistiques CMAP-SIMPAS et de l'équipe de probabilités INRIA-ASCII. Enfin, j'étais inscrit à l'école doctorale de mathématiques Hadamard (EDMH).



Au-delà de mon affiliation académique, mon travail de thèse s'inscrit dans les thématiques de recherche du réseau thématique en quantification d'incertitude RT-UQ (anciennement MASCOT-NUM). J'ai beaucoup profité de la possibilité d'échanger et de diffuser mon travail auprès de cette communauté. Enfin, je suis secrétaire du groupe Jeunes Statisticien-ne-s de la Société Française de Statistique (SFdS) depuis juin 2023.

Motivations

Cette thèse est d'abord motivée par un problème industriel situé à la croisée de deux domaines : l'analyse fiabiliste et la simulation d'événements rares.

Fiabilité des systèmes

En 2023, 77.7% de l'électricité produite par le groupe EDF dans le monde était d'origine nucléaire; et 9.1% d'origine hydraulique. La défaillance critique d'une centrale ou d'un barrage ayant des conséquences majeures, EDF est tenue de garantir la sûreté de ses installations nucléaires et hydrauliques.

Études probabilistes de sûreté

Le groupe i2i est spécialistes d'études probabilistes de sûreté (EPS, ou PSA en anglais pour probabilistic safety assessment). D'après la Règle Fondamentale de Sûreté (RFS n° 2002-01)

⁵hassane.chraibi@edf.fr

⁶anne.dutfoy@edf.fr

rédigée par EDF, l'IRSN (Institut de radioprotection et de sûreté nucléaire) et l'ASN (Autorité de sûreté nucléaire) :

“Les EPS sont une méthode d'évaluation des risques fondée sur une investigation systématique des scénarios accidentels. Elles se composent d'un ensemble d'analyses techniques permettant d'apprécier les fréquences d'événements redoutés et leurs conséquences. Elles permettent d'obtenir une appréciation globale du niveau de sûreté, intégrant aussi bien la fiabilité des équipements que le comportement des opérateurs.”

La sûreté de fonctionnement est pilotée selon quatre aspects que l'on rassemble sous l'acronyme FMDS⁷ pour :

- (i) *Fiabilité* : l'aptitude d'un dispositif à accomplir une fonction requise, dans des conditions d'utilisation et pour une période de temps déterminées.
- (ii) *Maintenabilité* : dans des conditions données d'utilisation, l'aptitude d'un dispositif à être maintenu ou rétabli dans un état dans lequel il peut accomplir sa fonction requise, avec des moyens de maintenance prescrits.
- (iii) *Disponibilité* : l'aptitude d'un système à être en état de remplir une fonction dans des conditions et à un instant donnés.
- (iv) *Sécurité* : l'aptitude d'un système à ne pas générer d'événements critiques ou catastrophiques dans des conditions données et pendant une durée donnée.

Dans le cadre de cette thèse, on s'intéresse à la fiabilité de systèmes industriels multicomposants intervenant dans le fonctionnement des centrales hydrauliques et nucléaires. En analyse de fiabilité, un système multicomposant désigne un système complexe composé de multiples éléments ou sous-systèmes interconnectés, appelés composants. Chaque composant peut être un équipement, une machine, un dispositif électronique, etc. Il contribue d'une manière ou d'une autre, et le plus souvent en interaction avec les autres composants, à accomplir la fonction globale du système. Les interactions peuvent inclure des flux de matière, d'énergie, d'information ou de contrôle entre les composants. Par exemple, un composant peut recevoir des entrées d'un autre composant et produire des sorties qui sont ensuite utilisées par d'autres composants. Un composant peut être dans différents modes de fonctionnement, par exemple : actif, inactif, dégradé, en panne, etc. Ce mode de fonctionnement est appelé le statut du composant. Le statut peut être automatiquement modifié via des mécanismes de contrôle ou suite à des événements aléatoires tels que des pannes ou des réparations.

⁷L'acronyme anglais est RAMS pour Reliability – Availability – Maintainability – Safety.

Modélisation de PDMPs avec l'outil PyCATSHOO

La fiabilité d'un système industriel est sa capacité à accomplir sa fonction globale pendant une période donnée dans des conditions spécifiées. Une défaillance critique est déclarée lorsque cette fonction n'est plus remplie. Elle correspond généralement au dépassement de seuils critiques par des variables physiques continues du système (par exemple, une température, une pression ou un niveau de liquide dans un réservoir). Si le système est bien conçu, cela ne peut se produire que si le système reste suffisamment longtemps dans un état très dégradé, c'est-à-dire après la défaillance de groupes vitaux de composants et avant leur réparation. Dans l'industrie nucléaire et hydraulique, ces systèmes ont en général un très haut niveau de fiabilité. Cela s'explique d'une part par leur haut niveau de redondance : le système peut être reconfiguré en utilisant plusieurs composants identiques pour assurer son fonctionnement en attendant la réparation ou le remplacement des composants défectueux. D'autre part, le temps d'attente moyen avant une panne ou un incident est généralement considérablement plus long que le temps d'attente moyen avant la réparation du problème.

La défaillance critique est donc très rare. Sa probabilité d'occurrence ne peut en pratique pas être estimée à partir de données réelles mais les systèmes impliqués se prêtent en revanche très bien à la simulation informatique. Une représentation dynamique du système doit modéliser à la fois les trajectoires des variables physiques à surveiller et le statut de chaque composant au cours du temps. C'est le rôle d'un outil de calcul développé à EDF nommé PyCATSHOO [CHS16; Des+21].



FIGURE 1 **Logo de l'outil PyCATSHOO.**
Plus d'informations sur le site :
<http://www.pycatshoo.org/>.

Une fonctionnalité essentielle de PyCATSHOO est la simulation, via une API Python et une API C++, de systèmes industriels dynamiques sous la forme de processus de Markov déterministes par morceaux ou PDMP (pour "Piecewise Deterministic Markov Processes") [Dav84; DDZ15]. Les PDMPs forment une classe très large de processus stochastiques, celle de tous les processus markoviens sans partie diffusive. Ce sont des processus dit hybrides, car leur variable d'état est constituée d'une partie continue (appelée "position") et d'une partie discrète (appelée "régime"). La position évolue selon des trajectoires déterministes paramétrées par le régime. À des instants

aléatoires (dont la distribution dépend continuellement de l'état du processus) ainsi qu'à des instants déterministes (lorsque la position atteint la frontière de l'espace d'états), le processus saute vers une nouvelle position et un nouveau régime aléatoires, puis reprend une nouvelle trajectoire déterministe.

Les pannes et les réparations des composants sont considérées comme des événements aléatoires ponctuels, tandis que l'évolution des variables continues est régie par des équations différentielles déterministes dérivées de lois physiques. Chaque composant a ses propres caractéristiques de fiabilité qui peuvent dépendre continuellement de variables physiques continues (par exemple, un composant immergé peut être affaibli par une pression ou une température excessive d'un liquide). Les équations différentielles régissant l'évolution des variables physiques sont, à leur tour, paramétrées par le statut des composants (l'eau, par exemple, circulera dans le système à un débit différent si une vanne reste bloquée en position ouverte ou fermée). Un tel comportement se prête donc bien à la modélisation et à la simulation de PDMPs. Ces simulations peuvent cependant être coûteuses, essentiellement à cause de la résolution des équations différentielles.

Simulation d'événements rares

Notre objectif est d'estimer la probabilité qu'un système industriel donné entre en défaillance critique avant une date fixée. Mathématiquement, cela se représente par la probabilité qu'une trajectoire de PDMP de distribution donnée atteigne une région cible de son espace d'états avant une date fixée.

Le problème avec l'approche Monte-Carlo classique

Puisque l'on dispose d'un simulateur de PDMPs (PyCATSHOO), l'idée la plus naturelle est de recourir à une estimation Monte-Carlo classique. Le principe est très simple : générer un grand nombre de trajectoires possibles du PDMP et retourner la proportion de ces trajectoires qui réalisent l'événement d'intérêt (la défaillance du système). Cet estimateur a, sur le papier, toutes les bonnes propriétés : il n'est pas biaisé et converge vers la quantité d'intérêt. De plus, sa vitesse de convergence, donnée par le théorème central limite, est indépendante de la dimension de l'espace. Il n'est cependant pas adapté à l'estimation de probabilités d'événements rares. C'est un problème bien connu. Lorsque la probabilité cible est très inférieure à l'inverse du nombre de simulations générées n , deux cas de figures sont possibles :

- (i) avec grande probabilité, n'observer aucune réalisation de l'événement d'intérêt, et retourner 0 à tort,
- (ii) observer "par chance" $k \geq 1$ réalisations de l'événement d'intérêt, et retourner k/n qui est bien plus grand que la probabilité cible.

En clair, l'estimateur est toujours loin (relativement) de la quantité à estimer. C'est donc un problème de variance. On considère généralement qu'il faut générer un échantillon de taille supérieure à 100 fois l'inverse de la probabilité cible pour estimer correctement cette dernière (le coefficient de variation est alors légèrement inférieur à 0.1).

Comme on l'a dit, la défaillance critique des systèmes industriels d'intérêt est un événement rare. Les probabilités de défaillance typiques sont comprises entre 10^{-5} et 10^{-7} . Leur estimation Monte Carlo à l'aide de l'outil PyCATSHOO requiert donc des centaines de millions, si ce n'est des milliards de simulations. En raison de la physique complexe des systèmes considérés, le coût de simulation d'une seule trajectoire de PDMP peut déjà être très élevé. Il faut en effet résoudre numériquement de nouvelles équations différentielles après chaque saut, parfois avec un code industriel dédié, pour déterminer les portions de trajectoire déterministe. L'estimation de la probabilité de défaillance du système est donc extrêmement coûteuse.

Travaux de Thomas Galtier

Des alternatives aux méthodes Monte Carlo adaptées à l'estimation de probabilités d'événements rares existent et forment un champ riche et dynamique des mathématiques appliquées ([Buc04; RT+09]). Dans sa thèse soutenue en 2019 ([Gal19]), Thomas Galtier a étudié l'application de deux grandes familles de méthodes aux PDMPs. D'abord les méthodes d'échantillonnage d'importance [Chr+19] et ensuite les méthodes dites de "splitting" basées sur des systèmes de particules en interaction [Chr+21]. Ma thèse fait suite à celle de Thomas Galtier auprès de la même équipe.

Des résultats assez spectaculaires ont pu être obtenus sur de petits systèmes industriels à l'aide de l'échantillonnage d'importance. Le principe est de simuler une trajectoire de PDMP, non pas selon sa véritable distribution mais selon une distribution biaisée favorisant la réalisation de l'événement d'intérêt (la défaillance du système), et de corriger le biais à l'aide du rapport de vraisemblance approprié. Thomas Galtier et ses co-auteurs ont défini à la fois

- ◊ la mesure dominante selon laquelle on pouvait expliciter la densité de probabilité d'une trajectoire de PDMP de durée fixée (essentielle pour évaluer les rapports de vraisemblance intervenant en échantillonnage d'importance),
- ◊ la distribution d'importance optimale à utiliser dans le cadre des PDMPs et permettant de produire un estimateur de variance nulle de la quantité d'intérêt.

Le processus de distribution d'importance optimale est un PDP (processus déterministe par morceaux), que l'on peut rendre markovien, vivant dans le même espace d'états et suivant les mêmes trajectoires déterministes que le PDMP originel. La distribution de ses instants de sauts et de la destination de ses sauts est complètement caractérisée par ce qu'on appelle la "fonction

committor”⁸ du processus. Cette fonction renvoie la probabilité qu’une trajectoire de PDMP réalise l’événement d’intérêt avant une date fixée, sachant son état actuel.

Cette fonction committor n’étant évidemment pas connue, il faut l’approximer en pratique. Comme on l’a dit, Thomas Galtier et ses co-auteurs ont proposé une méthode d’approximation ayant donné de bons résultats sur de petits systèmes industriels dans [Chr+19], mais beaucoup moins performantes pour des systèmes complexes et/ou de grande taille. L’objectif de ma thèse est de proposer une méthodologie plus complète et systématique de l’approximation de la fonction committor de PDMPs modélisant des phénomènes complexes. Ceci implique de connaître a priori ou d’apprendre au cours des simulations, les mécanismes et les chemins conduisant une trajectoire de PDMP à réaliser l’événement d’intérêt.

Structure du manuscrit

Le corps du manuscrit, en dehors de cette introduction et de ce qui la précède, est divisé en trois parties, chacune constituée de plusieurs chapitres. Leur contenu est brièvement décrit ci-dessous.

État de l’art

L’objectif de cette partie est d’introduire les notions permettant d’énoncer le problème que traite cette thèse, et de décrire les méthodes déjà existantes sur lesquelles nous pourrions nous reposer. Pour des raisons de cohérence dans les chapitres et de concision, cet état de l’art ne cherche pas à présenter l’ensemble des notions utilisées pour la résolution du problème. Celles-ci seront décrites en temps utile dans les chapitres concernés de la [section 3.3.2](#) dédiée aux contributions.

Prérequis

Cet état de l’art se veut accessible aux mathématiciens débutants. On suppose néanmoins le lecteur familier des notions suivantes :

- ◊ Bases de théorie de la mesure : tribus, mesures, mesures sigma-finies, boréliens, intégration par rapport à une mesure, Radon-Nikodym et densité par rapport à une mesure. Ces notions interviennent principalement dans le [chapitre 3](#).
- ◊ Probabilités élémentaires en lien avec la théorie de la mesure : espérance, variance, conditionnement, distributions de probabilités classiques et leurs propriétés (gaussiennes, uniformes, exponentielles).
- ◊ Statistique mathématique rudimentaire : moyenne empirique et variance empirique, comportement asymptotique (“loi forte des grands nombres” et “théorème central limite”), intervalles de confiance.

⁸Une traduction possible serait “fonction d’engagement”.

- ◇ Processus stochastiques à temps discret : martingales, chaînes de Markov, temps d'arrêt.

Chapitre 1 : Monte-Carlo et événements rares

En simulation numérique probabiliste, une tâche courante consiste à évaluer une quantité d'intérêt définie comme l'espérance d'une fonction sous une distribution de référence. Selon la nature et la dimension de l'espace, il n'est pas toujours possible d'approcher cette intégrale à l'aide d'une méthode déterministe. On a alors recours à une méthode de Monte Carlo : l'espérance est approchée par une moyenne empirique de tirages i.i.d. (indépendants et identiquement distribués) par ordinateur. Lorsque la distribution de référence attribue peu de masse là où la fonction à intégrer est grande (en valeur absolue), la variance relative de l'estimateur de Monte Carlo devient grande. Il est alors préférable d'utiliser des méthodes dites de réduction de variance. Nous nous concentrerons en particulier sur le cas de la simulation d'événements rares. L'objectif de ce chapitre est tout d'abord de rappeler les bases de la simulation de variables aléatoires et des méthodes de Monte Carlo. Nous décrivons ensuite deux familles de méthodes de réduction de la variance adaptées aux événements rares : les méthodes d'échantillonnage d'importance et les méthodes dites de "splitting".

Chapitre 2 : Fondamentaux des PDMPs

La classe des processus de Markov déterministes par morceaux (PDMP) est l'objet mathématique central de cette thèse. L'objectif de ce chapitre est de fournir au lecteur toutes les notions sur les PDMPs nécessaires pour comprendre pleinement les contributions de cette thèse. Un PDMP est un processus stochastique à temps continu avec sauts, non diffusif. Sa particularité (qui en fait en réalité un processus très général) est son comportement hybride. D'une part, il prend des valeurs dans un espace composé d'une partie continue et d'une partie discrète. D'autre part, il évolue selon une dynamique déterministe entre les sauts, et saute à la fois à des instants aléatoires et déterministes. Ce comportement est en fait assez intuitif pour les lecteurs familiers avec les processus de Poisson inhomogènes. Nous rappelons d'abord les fondamentaux des processus de Poisson, avant d'aborder les caractéristiques essentielles des PDMPs. Enfin, nous proposons des approches classiques pour simuler des trajectoires de PDMP.

Chapitre 3 : Échantillonnage préférentiel de PDMPs

Ce chapitre vise à présenter les fondements de l'échantillonnage d'importance de PDMPs. Cette méthode nécessite l'évaluation de la vraisemblance d'une trajectoire de PDMP. Nous donnons tout d'abord la définition d'une mesure dominante sur l'espace des trajectoires de PDMP par rapport à laquelle elles admettent une densité de probabilité. Nous décrivons ensuite le processus de distribution d'importance optimale. Ce processus optimal d'importance est un processus déterministe par morceaux (pas nécessairement markovien). Il préserve l'espace d'états et le

flot du PDMP original. La dynamique de saut du processus d'importance optimale peut être décrite explicitement. Dans le cas de l'estimation de probabilités d'événements rares, il existe une reformulation markovienne de ce processus optimal. L'expression de son intensité de saut et de son noyau de saut est alors directement liée à la fonction dite *committor* du processus. Cette fonction renvoie la probabilité qu'une trajectoire réalise l'événement d'intérêt étant donné son état actuel. Cette fonction committor n'est pas connue mais peut être approchée. À la fin du chapitre, nous présentons une implémentation de l'échantillonnage d'importance basée sur une approximation simple de la fonction committor.

Contributions

Les trois chapitres de contributions sont une réorganisation de deux articles rédigés pendant la thèse.

- ◊ Le premier article [Che+24a] correspond à une version antérieure de la méthodologie décrite dans le chapitre 4, et introduit les scores de proximité adaptés à la modélisation de systèmes industriels décrits dans le chapitre 5. Il a été publié en 2024 par le *SIAM Journal of Uncertainty Quantification*.

Guillaume CHENNETIER, Hassane CHRAIBI, Anne DUTFOY et Josselin GARNIER.
 “Adaptive importance sampling based on fault tree analysis for piecewise deterministic Markov process”. In : *SIAM/ASA Journal on Uncertainty Quantification* 12.1 (2024), p. 128-156

- ◊ Le second article correspond aux ajouts proposés dans la méthodologie du chapitre 4 par rapport à celle de l'article [Che+24a], et introduit de nouveaux scores de proximité, décrits dans le chapitre 6, basés sur les temps d'atteinte d'une marche aléatoire sur un graphe. Il sera prochainement soumis auprès d'un éditeur.

Guillaume CHENNETIER, Hassane CHRAIBI, Anne DUTFOY et Josselin GARNIER.
 “Graph-informed importance sampling for rare event estimation with piecewise deterministic Markov process”. In : (Submission in 2024)

Chapitre 4 : Méthodologie pour l'échantillonnage préférentiel adaptatif

On souhaite estimer la probabilité qu'une trajectoire de PDMP atteigne une région cible \mathcal{F} de son espace d'états avant une date donnée. Les liens entre la fonction committor du processus et la distribution optimale pour l'échantillonnage d'importance de PDMP ont été rappelées dans le chapitre 3. Ce chapitre vise à présenter une méthodologie complète d'échantillonnage d'importance adaptative basée sur l'approximation de la fonction committor d'un PDMP. Nous introduisons d'abord la forme générale d'une famille d'approximations de la fonction committor. Nous détaillons ensuite une procédure d'entropie croisée avec recyclage des échantillons passés

permettant de déterminer conjointement un bon candidat au sein de la famille d'approximations et d'estimer la probabilité d'intérêt. Nous prouvons l'optimalité et la normalité asymptotique de notre méthode sous des conditions à la fois simples à vérifier et à interpréter. Enfin, nous discutons des choix d'implémentation de cette méthode.

Chapitre 5 : Fonction d'importance pour systèmes multicomposants

La famille d'approximations de la fonction committor proposée au chapitre 4 repose sur une transformation paramétrique d'un score de proximité β entre tout régime du PDMP et l'ensemble $\mathcal{V}_{\mathcal{F}}$ des régimes permettant d'accéder à la région critique \mathcal{F} . Dans ce chapitre, nous considérons des PDMPs modélisant des systèmes industriels multi-composants. L'événement rare d'intérêt correspond à la défaillance critique du système. Cette défaillance critique ne peut se produire que lorsque certaines combinaisons clés de composants restent en panne pendant une durée suffisamment longue. Il existe d'autres combinaisons de composants appelés chemins minimaux qui sont les plus petits groupes de composants dont le bon fonctionnement empêche la défaillance du système. Nous proposons un score de proximité basé sur la proportion de chemins minimaux endommagés dans un régime donné. Notre méthode d'échantillonnage d'importance adaptative est testée avec ce score de proximité sur des systèmes série/parallèle simples, puis sur un cas test industriel plus complexe : la piscine de combustible nucléaire usagé.

Chapitre 6 : Fonction d'importance informée par graphe

Nous avons proposé dans le chapitre 5 un score de proximité β^{MPS} basé sur la proportion de chemins minimaux endommagés. Celui-ci s'est avéré très efficace lorsque le PDMP modélise un système industriel multicomposants cohérent. Dans ce chapitre, nous souhaitons aller plus loin et proposer un score de proximité à la fois plus général, au sens où l'on ne souhaite pas faire d'hypothèse sur ce que modélise le PDMP, et plus expressif encore. Le régime d'un PDMP évolue selon une marche aléatoire $(V_t)_t$ sur un graphe dont les régimes sont les sommets. Le temps moyen d'atteinte de la région $\mathcal{V}_{\mathcal{F}}$ pour cette marche aléatoire partant du sommet v serait un score de proximité naturel pour le régime v . Ce temps moyen d'atteinte est explicite lorsque la marche aléatoire est markovienne et homogène en temps. La marche aléatoire $(V_t)_t$ donnée par le PDMP étant typiquement non markovienne, nous calculons ces temps moyens d'atteinte pour une marche aléatoire simplifiée $(Y_t)_t$ approchant $(V_t)_t$. Notre méthode d'échantillonnage d'importance est implémentée avec ce score de proximité sur le cas de test final du chapitre précédent : la piscine de combustible nucléaire usagé.

Conclusion

Chapitre 7 : Conclusion et perspectives

Enfin, les résultats principaux de la thèse sont résumés dans le [chapitre 7](#) et discutés. Différentes perspectives s'offrent pour enrichir la méthode proposée ou pour étendre son application à de nouveaux cas tests. Nous proposons plusieurs pistes.

Notations and abbreviations

TABLE 1 **Table of notations.** PDMPs.

Notation	Meaning
Z_t	Position (continuous component) of the PDMP at time t
V_t	Regime (discrete component) of the PDMP at time t
$X_t = (Z_t, V_t)$	State of the PDMP at time t
\mathcal{Z}	Space of the position of the PDMP
\mathcal{V}	Set of the regimes of the PDMP
$\mathcal{X} = \mathcal{Z} \times \mathcal{V}$	The state space of the process $(X_t)_t$
Ψ	Deterministic flow for the state of the PDMP
ψ	Deterministic flow for the position only
λ	Jump intensity of the PDMP
$\lambda_x(t)$	$\lambda(\Psi(x, t))$
$\Lambda(x, s)$	$\int_0^s \lambda_x(t) dt$
Q	Jump kernel of the PDMP
$q(\cdot x)$	Jump kernel density of the PDMP with respect to a measure $\nu_q(\cdot x)$
$t_\partial(x)$	Waiting time from state x before reaching the boundaries of \mathcal{X} by following Ψ
$T_\lambda(x)$	Waiting time from state x before a random jump with intensity λ
\mathbf{T}	Space of PDMP trajectories
ζ	Dominant measure on \mathbf{T}
\mathbf{p}	p.d.f. characterizing the reference distribution of a PDMP trajectory (of finite duration) with respect to the dominant measure ζ
s_{\max}	Maximal duration of a PDMP trajectory
\mathbf{X}	PDMP trajectory $(X_t)_{t \in [0, s_{\max}]}$

TABLE 2 **Table of notations.** Committor function and rare events.

Notation	Meaning
\mathcal{F}	Critical region of the state space. We are looking for the hitting probability of this set before date s_{\max}
\mathbf{F}	Subset of PDMP trajectories of \mathbf{T} that reach the critical region \mathcal{F} before time s_{\max}
$\mathcal{V}_{\mathcal{F}}$	Subset of PDMP regimes (in \mathcal{V}) from which the critical region \mathcal{F} can be accessed by following the flow Ψ (without jumping)
\mathbf{I}	Function of interest whose expectation is sought under the reference distribution \mathbf{p} . Most often corresponds to the indicator function of the event $\{\mathbf{X} \in \mathbf{F}\}$
\mathcal{I}	$\mathbb{E}_{\mathbf{p}}[\mathbf{I}(\mathbf{X})]$, i.e. the quantity of interest. Most often corresponds to the probability of the rare event of interest
ξ^*	Committor function of the PDMP associated with the event $\{\mathbf{X} \in \mathbf{F}\}$
ξ^{+-}	Edge committor function of the PDMP associated with the event $\{\mathbf{X} \in \mathbf{F}\}$
ξ_{θ}	Importance function, i.e. an approximation of the committor function parameterized by $\theta \in \Theta$
ξ_{θ}^-	Edge importance function, i.e. an approximation of the edge committor function parameterized by $\theta \in \Theta$
ϕ_j	Basis function involved in the expression of the importance function
β	Proximity score between regimes and $\mathcal{V}_{\mathcal{F}}$, involved in the expression of the importance function
\mathbf{g}^*	Optimal importance sampling distribution $\propto \mathbf{I} \times \mathbf{p}$
\mathbf{g}_{θ}	Importance distribution parameterized by $\theta \in \Theta$

TABLE 3 **Table of notations.** Sets and basic functions.

Notation	Meaning
$]a, b[$	$\{x : a < x < b\}$, open interval (French notation)
$\llbracket a, b \rrbracket$	$[a, b] \cap \mathbb{Z} = \{a, a + 1, \dots, b - 1, b\}$ with $a, b \in \mathbb{Z}$ and $a \leq b$
$a \wedge b$	$\min(a, b)$
$a \vee b$	$\max(a, b)$
$\bigwedge_{i \in I} x_i$	$\min \{x_i \mid i \in I\}$
$\bigvee_{i \in I} x_i$	$\max \{x_i \mid i \in I\}$
$\text{Leb}(B)$	Lebesgue measure of a set B
$\text{supp}(f)$	$\{x : f(x) \neq 0\}$, support of the function f
$f[A]$	$\{f(x) : x \in A\}$, image of the set A under the function f

TABLE 4 **Table of notations.** Random variables and standard distributions.

Notation	Meaning
\mathcal{X}	State space of the random variable X
ζ	Dominant measure of the probability space \mathcal{X}
\mathbf{p}	P.d.f. characterizing the reference distribution with respect to the dominant measure ζ
$\mathbb{E}_{\mathbf{p}}[X]$	Expectation of the random variable X distributed according to \mathbf{p}
$\stackrel{\mathcal{L}}{=} \text{ or } \stackrel{d}{=}$	Equality in distribution
Φ	C.d.f. of the standard Gaussian distribution (mean 0 and variance 1)
Φ^{-1}	Quantile function of the standard Gaussian distribution
$\mathcal{N}(\mu, \sigma^2)$	Gaussian distribution with mean μ and variance σ^2
$\text{Unif}(A)$	Uniform distribution on the set A
$\text{Exp}(\lambda)$	Exponential distribution with parameter λ (inverse of the mean)
$\text{Bin}(n, p)$	Binomial distribution with number of trials n and probability of success p
$\text{Ber}(p)$	Bernoulli distribution with probability of success p
$\text{Geom}(p)$	Geometric distribution with probability of success p
$\text{Pois}(\lambda)$	Poisson distribution with intensity λ

TABLE 5 **Table of abbreviations.**

Abbreviation	Meaning
w.r.t.	with respect to
i.i.d.	independent and identically distributed
p.d.f.	probability density function
c.d.f.	cumulative distribution function
a.s.	almost surely
a.e.	almost everywhere
càdlag	right continuous with left limits (from french)
(S)LLN	(Strong) Law of large numbers
CLT	Central limit theorem
PDP(s)	Piecewise deterministic process(es)
PDMP(s)	Piecewise deterministic Markov process(es)
CMC	Classical Monte Carlo
IS	Importance sampling
MIS	Multiple importance sampling
AIS	Adaptive importance sampling
CE	Cross entropy
mCE	multilevel Cross entropy
iCE	improved Cross entropy
IF	Importance function
FTA	Fault tree analysis
MPS(s)	Minimal path set(s)
MCS(s)	Minimal cut set(s)
BFGS	The Broyden-Fletcher-Goldfarb-Shanno method in optimization
MHT	Mean hitting times

STATE OF THE ART

CHAPTER 1

Monte Carlo simulation and rare events

In probabilistic numerical simulation, a common task is the evaluation of a quantity of interest defined as the expectation of a function under a reference distribution. Depending on the nature and dimension of the space, it is not always possible to approximate this integral using a deterministic method. One then resorts to a Monte Carlo method: the expectation is approximated by an empirical average of i.i.d. draws by computer. When the reference distribution puts little mass where the function to be integrated is large (in absolute value), the relative variance of the Monte Carlo estimator becomes large. It is then preferable to use so-called variance reduction methods. We will focus in particular on the case of rare event simulation. The objective of this chapter is firstly to recall the basics of random variable simulation and Monte Carlo methods. We then describe two families of variance reduction methods adapted to rare events: importance splitting and importance sampling methods.

Content of this chapter

1.1	Introduction to Monte Carlo methods	4
1.1.1	Generating random variables	4
1.1.2	Monte Carlo methods	9
1.2	Particle-based methods	12
1.2.1	Subset simulation	13
1.2.2	Adaptive multilevel splitting	15
1.3	Importance sampling	15
1.3.1	The essentials	17
1.3.2	Adaptive procedures	23
1.3.3	Curse of dimensionality	29

1.1 Introduction to Monte Carlo methods

Computer generation of random variables generation is a fundamental technique in applied mathematics. Far from being limited to the simulation of directly probabilistic models, it is an essential component of many methods for solving deterministic problems. This is the very idea behind Monte Carlo methods: approximating an integral through an empirical mean of random draws. The introduction of randomness, particularly when controlled by convergence theorems or concentration inequalities, allows for efficient exploration of the space. This approach is also fruitful in optimization, where traditional deterministic methods may struggle with the dimensionality of the space, the multimodality of the objective function, its lack of smoothness, or its black-box nature (typically with an absence of gradients). Famous examples include genetic algorithms [Mit98] or simulated annealing [NJ10; DCM19]. Randomized space exploration is also at the heart of many multi-arm bandit and reinforcement learning algorithms.

In practice, there is not always a direct method to generate a random variable according to an arbitrary distribution. This remains an active area of research, especially for the efficient simulation of high-dimensional multi-modal distributions. Here, we present the fundamental principles.

1.1.1 Generating random variables

The first challenge is to generate random variables with a uniform distribution on the interval $[0, 1]$. As we will see, it is theoretically possible to simulate any other distribution using these uniform variables. Producing “true” randomness requires equipment exploiting the quantum properties of physical processes. This involves significant effort, considering that similar benefits can be obtained from a deterministic sequence of numbers whose statistical properties make them almost indistinguishable from a sequence of truly random, uniform, and independent variables. An algorithm capable of generating such a deterministic sequence is called a pseudo-random number generator.

Pseudo-random generator

The idea behind a pseudo-random generator is to iterate the application of a transformation f starting from an initial value called the seed x_0 . Early attempts at algorithms appeared in the 19th century for encrypting messages, but the “middle-square digits” algorithm proposed in 1946 by John Von Neumann is generally considered the first pseudo-random generator. This generator takes an input seed $x_0 \in \mathbb{N}$ with d digits and returns the d -digit number in the middle of the number x_0^2 (which has $2d$ digits). Later came the linear congruential generators (for example, the Lehmer generator introduced in 1948: $X_{n+1} = aX_n \bmod m$), which were widely used throughout the 20th century. Today, the most frequently implemented algorithm is the Mersenne-Twister

generator (notably the default generator for the R language and the `numpy.random` library in Python). While better generators exist, few applications (typically cryptography) require a more advanced generator.

In the following, we will assume the availability of an exact and inexpensive pseudo-random generator for independent uniform variables on the interval $[0, 1]$. Let us now discuss methods for generating a variable from an arbitrary one-dimensional distribution using the uniform random variable generator.

Inversion method

To simulate a real random variable X , it is sufficient to apply the generalized inverse of the cumulative distribution function (c.d.f.) of X to a uniform variable $U \sim \text{Unif}([0, 1])$.

DEFINITION 1 · Generalized inverse

Let F be a non-decreasing function from $A \subset \mathbb{R}$ to $B \subset \mathbb{R}$. We call the generalized inverse of F the following function :

$$F^{-1} : u \in B \mapsto \inf\{t \in A : F(t) \geq u\} = \sup\{t \in A : F(t) < u\}. \quad (\text{I.1})$$

If F is continuous and strictly increasing, it is bijective and F^{-1} is then its reciprocal bijection [La 20]. When F is a c.d.f. i.e. $B = [0, 1]$ and F is càdlàg (right continuous with left limits), F^{-1} is the corresponding quantile function.

PROPOSITION 1 · Inversion method

Let X be a random variable with c.d.f. F and $U \sim \text{Unif}([0, 1])$ be a random uniform variable.

- (i) The random variable $F^{-1}(U)$ has the same distribution as X ,
- (ii) if X is continuous then $F(X)$ has the same distribution as U .

Proof. (i) Note that $F^{-1}(U) \leq x \iff \sup\{t \in \mathbb{R} : F(t) < U\} \leq x \iff U \leq F(x)$. It results $\mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x)$.

(ii) Next, by the same argument, we have $\mathbb{P}(F(X) \leq x) = \mathbb{P}(X \leq F^{-1}(x)) = F(F^{-1}(x))$. Since F is continuous and increasing, $F(\inf\{t \in \mathbb{R} : F(t) \geq x\}) = x$. ■

The variable $-\log(U)/\lambda$, for instance, follows an exponential distribution with parameter λ . For more complex distributions, the explicit form of F^{-1} is seldom known. In some cases,

numerically inverting F is not too costly, but it leads to approximate simulations. Exact simulation methods employ two types of mechanisms: first, the transformation of random variables that we can simulate (as seen in the inversion method just discussed), and most importantly, acceptance-rejection procedures.

Acceptance-rejection method

An acceptance-rejection procedure involves simulating random variables and keeping only those that satisfy a given criterion. The criterion is chosen in such a way that the retained random variables follow a target distribution. Therefore, we deliberately draw our variables in a “too large” space in which simulation is easy. Obtaining a sample from the target distribution by simply removing points from the initial sample is possible again thanks to the properties of the uniform distribution. It is stable under rejection: drawing points uniformly in a set A and retaining only those falling into a subset $B \subset A$ is equivalent to drawing uniformly in B . This property generalizes under the term “fundamental theorem of simulation” [RC99]. We aim to simulate a real random variable X with density \mathbf{p} .

PROPOSITION 2 · Fundamental theorem of simulation

Let \mathbf{p} be a p.d.f. on a space $\mathcal{X} \subseteq \mathbb{R}^d$ and B be the following set :

$$B := \{(x, y) \in \mathcal{X} \times \mathbb{R}_+ : 0 \leq y \leq \mathbf{p}(x)\}. \quad (1.2)$$

If $(X, Y) \sim \text{Unif}(B)$ then the marginal p.d.f. of X is \mathbf{p} .

Proof. If $(X, Y) \sim \text{Unif}(B)$, the density of the couple is $(x, y) \mapsto \frac{1}{\text{Leb}(B)} \mathbf{1}_{(x,y) \in B}$. We have :

$$\text{Leb}(B) = \int_{\mathcal{X}} \int_{\mathbb{R}_+} \mathbf{1}_{(x,y) \in B} dy dx = \int_{\mathcal{X}} \int_0^{\mathbf{p}(x)} dy dx = \int_{\mathcal{X}} \mathbf{p}(x) dx = 1. \quad (1.3)$$

The marginal p.d.f. of X is given by :

$$x \mapsto \int_{\mathbb{R}_+} \frac{1}{\text{Leb}(B)} \mathbf{1}_{(x,y) \in B} dy = \int_0^{\mathbf{p}(x)} dy = \mathbf{p}(x). \quad (1.4)$$

■

We would like to draw a couple (X, U) uniformly on B through rejection, i.e. by uniformly drawing the couple in a larger set until $U \leq \mathbf{p}(X)$. This method is formalized as follows: we assume that we know a probability density function (p.d.f.) \mathbf{g} on \mathcal{X} that we can simulate and a constant $m > 0$ such that $\mathbf{p}(x) \leq m \cdot \mathbf{g}(x)$ for all $x \in \mathcal{X}$. One just has to draw couples $(Y_k, U_k)_{k \in \mathbb{N}}$ with $Y_k \sim \mathbf{g}$ and $U_k \sim \text{Unif}([0, 1])$ until $U_k \leq \frac{\mathbf{p}(Y_k)}{m \cdot \mathbf{g}(Y_k)}$ to get a variable with p.d.f. \mathbf{p} :

PROPOSITION 3 · Accept-reject method

Let \mathbf{p} and \mathbf{g} be two p.d.f., $(U_k)_{k \in \mathbb{N}^*}$ be a sequence of i.i.d. uniform random variables on $[0, 1]$ and $(Y_k)_{k \in \mathbb{N}^*}$ be a sequence of i.i.d. random variables (also independent from the U_k) with p.d.f. \mathbf{g} . Finally, let $N = \inf\left\{k \in \mathbb{N}^* : U_k \leq \frac{\mathbf{p}(Y_k)}{m \cdot \mathbf{g}(Y_k)}\right\}$. Then \mathbf{p} is the p.d.f. of the variable $X = Y_N$.

Proof. Let $U \sim \text{Unif}([0, 1])$ and $Y \sim g$. First, note that all terms of the two sequences $(U_k)_{k \in \mathbb{N}^*}$ and $(Y_k)_{k \in \mathbb{N}^*}$ are distributed as U and Y . The probability of success of each trial is then given by :

$$\begin{aligned} \mathbb{P}\left(U \leq \frac{\mathbf{p}(Y)}{m \cdot \mathbf{g}(Y)}\right) &= \mathbb{E}_{Y \sim \mathbf{g}}\left[\mathbb{P}_{U \sim \text{Unif}([0,1])}\left(U \leq \frac{\mathbf{p}(Y)}{m \cdot \mathbf{g}(Y)} \mid Y\right)\right] = \mathbb{E}_{\mathbf{g}}\left[\frac{\mathbf{p}(Y)}{m \cdot \mathbf{g}(Y)}\right] \\ &= \int_{\mathbb{R}} \frac{\mathbf{p}(y)}{m \cdot \mathbf{g}(y)} \mathbf{g}(y) \, dy = \frac{1}{m} \int_{\mathbb{R}} \mathbf{p}(y) \, dy = \frac{1}{m}. \end{aligned} \quad (1.5)$$

Each trial is independent of the previous ones, so N follows a geometric distribution of parameter $\frac{1}{m}$. Using the same arguments¹, for any $x \in \mathbb{R}$, we have :

$$\begin{aligned} \mathbb{P}\left(Y \leq x, U \leq \frac{\mathbf{p}(Y)}{m \cdot \mathbf{g}(Y)}\right) &= \mathbb{E}_{Y \sim \mathbf{g}}\left[\mathbb{P}\left(Y \leq x, U \leq \frac{\mathbf{p}(Y)}{m \cdot \mathbf{g}(Y)} \mid Y\right)\right] \\ &= \mathbb{E}_{\mathbf{p}}\left[\mathbb{1}_{Y \leq x} \frac{\mathbf{p}(Y)}{m \cdot \mathbf{g}(Y)}\right] = \mathbb{E}_{\mathbf{p}}\left[\mathbb{1}_{X \leq x} \frac{1}{m}\right] \\ &= \frac{1}{m} \mathbb{P}_{\mathbf{p}}(X \leq x). \end{aligned} \quad (1.6)$$

Finally :

$$\begin{aligned} \mathbb{P}(X \leq x) &= \sum_{n \geq 1} \mathbb{P}(Y_n \leq x, N = n) \\ &= \sum_{n \geq 1} \mathbb{P}\left(Y \leq x, U \leq \frac{\mathbf{p}(Y)}{m \cdot \mathbf{g}(Y)}\right) \times \mathbb{P}\left(U > \frac{\mathbf{p}(Y)}{m \cdot \mathbf{g}(Y)}\right)^{n-1} \\ &= \frac{\mathbb{P}_{\mathbf{p}}(X \leq x)}{m} \times \sum_{n \geq 1} \left(1 - \frac{1}{m}\right)^{n-1} = \frac{\mathbb{P}_{\mathbf{p}}(X \leq x)}{m} \times \frac{1}{1 - \left(1 - \frac{1}{m}\right)} \\ &= \mathbb{P}_{\mathbf{p}}(X \leq x). \end{aligned} \quad (1.7)$$

■

Therefore, the rejection criterion is random. The uniform variables only come into play as auxiliary variables for the acceptance test. The simulation cost of a random variable with distribution \mathbf{p} is the number N of pairs (Y_k, U_k) that must be drawn before $U_k < \frac{\mathbf{p}(Y_k)}{m \cdot \mathbf{g}(Y_k)}$. Since the successive pairs are independent, the variable N follows a geometric distribution

¹The equality $\mathbb{E}_{\mathbf{p}}[X] = \mathbb{E}_{\mathbf{g}}\left[X \frac{\mathbf{p}(X)}{\mathbf{g}(X)}\right]$ is known as the ‘‘importance sampling’’ trick and is detailed in [section 1.3.1](#).

with parameter $\frac{1}{m}$. Therefore, we generate on average m pairs before obtaining a realization distributed according to \mathbf{p} . For a given \mathbf{g} , we cannot do smaller than $m^* := \sup_x \frac{\mathbf{p}(x)}{\mathbf{g}(x)}$. Let \mathcal{G} be the set of densities for which we assume we know how to simulate. The auxiliary density that is most efficient on average for the accept-reject procedure belongs to :

$$\mathcal{G}^* = \arg \min_{\mathbf{g} \in \mathcal{G}} \left\{ \sup_x \frac{\mathbf{p}(x)}{\mathbf{g}(x)} \right\}. \quad (1.8)$$

When trying to sample from a high-dimensional distribution \mathbf{p} , it becomes very difficult to find an auxiliary distribution \mathbf{g} that can be simulated and is not much smaller than \mathbf{p} at any point in space. A second drawback of the rejection method is that the density \mathbf{p} must be explicitly evaluated at each point drawn according to the auxiliary density \mathbf{g} . In many applications, \mathbf{p} is only known up to a multiplicative constant (very common in statistical physics and Bayesian statistics). Calculating the associated normalization constant via a numerical integration method is much too expensive in high dimensions. The vanilla version of the accept-reject method is therefore generally limited to low dimensions.

Monte Carlo Markov Chain (MCMC)

The accept-reject method is slow because a candidate is proposed according to the same distribution at each attempt, always with the same small chance of success. This is the price to pay for the produced variable to be exactly distributed according to \mathbf{p} . The simulation can be accelerated by giving up exactness. We present a method that, when iterated a sufficient number of times, produces a variable approximately distributed according to \mathbf{p} . The idea is to construct, using successive accept-reject proposals, a Markov chain whose invariant distribution is \mathbf{p} . Under suitable conditions, the state of the Markov chain is asymptotically distributed according to \mathbf{p} . These methods are called Markov Chain Monte Carlo (MCMC) methods. I refer the reader interested in an in-depth study of the properties of convergence of Markov chains to [BH22]. We present the most famous MCMC method: the Metropolis-Hastings method [Has70] (generalization of Metropolis method [Met+53]). We aim to simulate according to a target distribution \mathbf{p} on a space \mathcal{X} . Let Q be a transition kernel on \mathcal{X} , i.e., a family of distributions $(Q(\cdot | x))_{x \in \mathcal{X}}$ on \mathcal{X} indexed by each element of \mathcal{X} . For any $x \in \mathcal{X}$, $Q(\cdot | x)$ is a probability mass function if \mathcal{X} is finite or countable, and it is a probability density function if \mathcal{X} is continuous (with respect to some dominating measure μ on \mathcal{X}). For any pair $x, y \in \mathcal{X}$, we also define the acceptance ratio:

$$r : (x, y) \in \mathcal{X}^2 \mapsto \min \left(1, \frac{\mathbf{p}(y)Q(x | y)}{\mathbf{p}(x)Q(y | x)} \right). \quad (1.9)$$

The Metropolis-Hastings algorithm corresponds to the following procedure: at each iteration, a new candidate Y is proposed from the state X according to the kernel $Q(\cdot | X)$, and is accepted

with probability $r(X, Y)$. Assume Q is irreducible, i.e. for any pair of states $(x_0, y) \in \mathcal{X}^2$ there

ALGORITHM 1 · Metropolis-Hastings

```

1 Initialization state  $X_0 \in \mathcal{X}$ .
2 for  $k = 1, \dots, n$  do
3     Sampling step
                                      $Y_k \sim Q(\cdot \mid X_{k-1}).$            (1.10)
4     Accept-reject step
                                      $U_k \sim \text{Unif}([0, 1]),$            (1.11)
                                      $X_k = X_{k-1} \mathbb{1}_{U > r(X_{k-1}, Y_k)} + Y_k \mathbb{1}_{U \leq r(X_{k-1}, Y_k)}.$    (1.12)
5 end
6 Return  $X_n$ .
```

exists a path of non-zero probability between them. Said otherwise, either $Q(y \mid x_0) > 0$, or there are $n \geq 1$ and $(x_1, \dots, x_n) \in \mathcal{X}^n$ such that:

$$\left(\prod_{k=1}^n Q(x_k \mid x_{k-1}) \right) \times Q(y \mid x_n) > 0. \tag{1.13}$$

Moreover, assume that $Q(y \mid x_0) > 0 \implies Q(x_0 \mid y) > 0$. Then, it can be proved that the asymptotic distribution of the produced sequence $(X_n)_n$ (when $n \rightarrow \infty$) is \mathbf{p} .

Metropolis-Hastings performs much better than the standard accept-reject method in high dimensions, but convergence remains slow. Proposing transitions according to a non-reversible kernel allows for faster independence from the initial state of the chain. Modern MCMC methods use different strategies to reduce the correlation between iterations to speed up convergence. These include, among others, the Metropolis-adjusted Langevin algorithm (MALA) [GC11], Hamiltonian Monte Carlo (HMC) [Bet17], the No-U-Turn Sampler (NUTS) [HG+14], or PD-MCMC methods based on PDMPs (see chapter 2). The article [RR98] is an original review of a large number of theoretical and practical questions concerning MCMC methods.

1.1.2 Monte Carlo methods

The Monte Carlo method emerged in the United States during the 1940s with the advent of the first electronic computers. Driven primarily by Stanislaw Ulam and John von Neumann, its early applications enabled much faster resolution of the neutronic calculations involved in the development of nuclear weapons. It was first referred to as Monte Carlo in [MU49]. The name

of the method was proposed by Nicholas Metropolis, in a nod to the Monte Carlo casino where Stanislaw Ulam's uncle used to gamble. For readers interested in the origins of the method, I recommend the following short articles, which also cover the first attempts at pseudo-random simulation and the birth of the accept-reject method: [Met+87; Eck87].

The principle of the Monte Carlo method is to approximate an expectation by an empirical mean, using independent and identically distributed random samples. Suppose we aim to estimate:

$$\mathcal{I} := \mathbb{E}_{X \sim \mathbf{p}}[\mathbf{I}(X)], \quad (1.14)$$

where $\mathbf{I} : \mathcal{X} \rightarrow \mathbb{R}$ is a target function, and X follows a reference distribution (also called nominal distribution) \mathbf{p} . The classical Monte Carlo method (CMC) then simply consists in estimating the expectation \mathcal{I} by the empirical mean of variables drawn i.i.d. according to the reference distribution \mathbf{p} .

$$\hat{\mathcal{I}}_n^{\text{CMC}} = \frac{1}{n} \sum_{k=1}^n \mathbf{I}(X_k) \quad \text{with } X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathbf{p}. \quad (1.15)$$

No specific assumptions about the space \mathcal{X} are necessary as long as a probability distribution can be defined. Important remark: this method more generally allows estimating the integral of a function $\varphi : \mathcal{X} \rightarrow \mathbb{R}$ with respect to a measure μ on \mathcal{X} . It suffices to choose a reference distribution characterized by a p.d.f. \mathbf{p} absolutely continuous with respect to μ and whose support contains that of φ , i.e. $\varphi(x) \neq 0 \implies \mathbf{p}(x) > 0$ for μ -a.e. $x \in \mathcal{X}$. Then by noting $\mathbf{I} = \frac{\varphi}{\mathbf{p}}$, we have²:

$$\int_{\mathcal{X}} \varphi(x) \mu(dx) = \int_{\mathcal{X}} \frac{\varphi(x)}{\mathbf{p}(x)} \mathbf{p}(x) \mu(dx) = \mathbb{E}_{X \sim \mathbf{p}} \left[\frac{\varphi(X)}{\mathbf{p}(X)} \right] = \mathcal{I}. \quad (1.16)$$

Properties of the classical Monte Carlo estimator

It is immediately clear from the linearity of expectation that the estimator $\hat{\mathcal{I}}_n^{\text{CMC}}$ is unbiased. By a direct application of the strong law of large numbers, we have :

PROPOSITION 4 · Consistency

$$\hat{\mathcal{I}}_n^{\text{CMC}} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \mathcal{I}. \quad (1.17)$$

The variance of $\hat{\mathcal{I}}_n^{\text{CMC}}$ is given by $\frac{\hat{\sigma}_{\mathbf{p}}^2}{n}$ with $\sigma_{\mathbf{p}}^2 := \text{Var}_{\mathbf{p}}(\mathbf{I}(X))$. This variance can be estimated, from the same sample X_1, \dots, X_n , by :

$$\hat{\sigma}_{\mathbf{p}}^2 = \frac{1}{n-1} \sum_{k=1}^n (\mathbf{I}(X_k) - \hat{\mathcal{I}}_n^{\text{CMC}})^2. \quad (1.18)$$

²This trick will be generalized to give rise to the importance sampling method (discussed in [section 1.3](#)).

Assuming $\sigma_{\mathbf{p}}^2 < \infty$, we have from the central limit theorem :

PROPOSITION 5 · Asymptotic normality

$$\sqrt{n}(\hat{\mathcal{I}}_n^{\text{CMC}} - \mathcal{I}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma_{\mathbf{p}}^2). \quad (1.19)$$

By application of Slutsky's lemma, we also have :

$$\sqrt{n} \frac{\hat{\mathcal{I}}_n^{\text{CMC}} - \mathcal{I}}{\hat{\sigma}_{\mathbf{p}}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1). \quad (1.20)$$

We deduce an asymptotic confidence interval for \mathcal{I} . Let $\alpha \in [0, 1]$, we note $\Phi^{-1}(1 - \alpha/2)$ the $1 - \alpha/2$ -quantile of the $\mathcal{N}(0, 1)$ distribution. We have :

$$\mathbb{P}\left(\mathcal{I} \in \left[\hat{\mathcal{I}}_n^{\text{CMC}} - \Phi^{-1}(1 - \alpha/2) \frac{\hat{\sigma}_{\mathbf{p}}}{\sqrt{n}}, \hat{\mathcal{I}}_n^{\text{CMC}} + \Phi^{-1}(1 - \alpha/2) \frac{\hat{\sigma}_{\mathbf{p}}}{\sqrt{n}}\right]\right) \xrightarrow[n \rightarrow \infty]{} 1 - \alpha. \quad (1.21)$$

The estimator therefore converges to \mathcal{I} at the classical rate of $\mathcal{O}(n^{-1/2})$. This rate is independent of the dimension of the space \mathcal{X} . This is the great strength of the Monte Carlo method. Deterministic numerical methods, in comparison, converge more and more slowly as the dimension increases.

Rare event simulation and variance reduction

The performance of the Monte Carlo estimator $\hat{\mathcal{I}}_n^{\text{CMC}}$ can be measured by its relative variance (square of the coefficient of variation): $\sigma_{\mathbf{p}}^2/\mathcal{I}^2$. This relative variance is large when the nominal distribution \mathbf{p} puts weight where the function \mathbf{I} is close to zero. Indeed, most of the realizations X_1, \dots, X_n then fall where $|\mathbf{I}|$ is small. The estimator $\hat{\mathcal{I}}_n^{\text{CMC}}$ is therefore the result of a large number of almost zero contributions and a small number of important contributions. Unless we sample enough, we find ourselves in one of the two following cases :

- (i) More frequently, \mathcal{I} is underestimated because we do not observe the regions where $|\mathbf{I}|$ is large.
- (ii) Less frequently, we quickly draw a point where $|\mathbf{I}|$ is large, but we are unable to reweight it by the small probability of this event, resulting in a significant overestimation of \mathcal{I} .

The situation that is most representative of this problem, and the one that is of interest in this thesis, is the case where $\mathbf{I} = \mathbb{1}_{\mathcal{F}}$ is the indicator function of a region \mathcal{F} that is rarely visited in the space \mathcal{X} . This is known as rare event probability estimation. In this case, we have :

$$\mathcal{I} = \mathbb{E}_{X \sim \mathbf{p}}[\mathbf{I}(X)] = \mathbb{E}_{X \sim \mathbf{p}}[\mathbb{1}_{X \in \mathcal{F}}] = \mathbb{P}_{\mathbf{p}}(X \in \mathcal{F}). \quad (1.22)$$

The rarer the event $\{X \in \mathcal{F}\}$ is under \mathbf{p} , the more difficult it is to estimate \mathcal{I} using classical Monte Carlo. For $n \ll \mathcal{I}^{-1}$, the estimator $\hat{\mathcal{I}}_n^{\text{CMC}}$ is most often 0 but exceeds \mathcal{I} by much when at least one realization belongs to \mathcal{F} . This is indeed a variance problem since the estimator is often far from \mathcal{I} on average. Its relative variance is given by :

$$\frac{\text{Var}(\hat{\mathcal{I}}_n^{\text{CMC}})}{\mathbb{E}[\hat{\mathcal{I}}_n^{\text{CMC}}]^2} = \frac{1 - \mathcal{I}}{n\mathcal{I}}. \quad (1.23)$$

The relative variance is approximately inversely proportional to \mathcal{I} when \mathcal{I} is small. A rule of thumb is to keep $n\mathcal{I} > 100$ to ensure that the coefficient of variation is under 0.1.

There are many alternative sampling methods that produce estimators with smaller variance than classical Monte Carlo. These are known as variance reduction methods.

In this chapter, we discuss two major families of variance reduction methods adapted to the simulation of rare events: splitting methods in [section 1.2](#), and importance sampling in [section 1.3](#). However, other variance reduction methods should also be mentioned, on a non-exhaustive basis. Some of which are general-purpose, such as quasi Monte Carlo [[Sob90](#)], control variates [[GS02](#)], antithetic variables [[HM56](#)]. Others are specifically designed for rare event simulation, such as line sampling [[Lu+08](#)] and directional sampling [[Bje88](#)]. The FORM-SORM method [[ZO99](#)], which is not a sampling method per se, is also very popular in industrial reliability. A recent literature review [[SK23](#)] examines the application of all these methods in structural reliability. We do not cover them in this thesis, as most are difficult to adapt to the case of dynamic rare events. A rare event is said to be static when X is a random vector, and dynamic when X is the trajectory of a stochastic process. In the latter case, \mathcal{F} generally corresponds to a subregion of the process state space, and the rare event of interest occurs if \mathcal{F} is reached before a given stopping time. The dynamic case is not necessarily more difficult than the static case (as we will see in the next section), but it has several specific features that often require adaptation of classical variance reduction methods.

1.2 Particle-based methods

The methods presented in this section belong to the broad family of sequential Monte Carlo (SMC) methods [[CP20](#)]. They involve simulating a large number of interacting particles that are iteratively selected, cloned, and mutated so that the distribution of the particle sample approaches a target distribution. As a result, we also speak of particle-based methods. There is a very wide variety of methods based on this principle, developed by different communities, under different names, and for different applications. The term SMC was traditionally associated with filtering methods such as the celebrated Kalman filter [[KB+18](#)]. The formalism of Feynman-Kac models [[Moro4](#)] has allowed a more general formulation of particle methods, which now

encompasses most cases of interest.

In particular, particle methods include algorithms specifically designed for the simulation of rare events called “splitting methods”. They were originally introduced in [KH51] to study the transmission of particles through an obstacle. They are sometimes referred to as importance splitting, as opposed to importance sampling methods (which are covered in section 1.3). The history of this method is detailed in [CGR19]. The principle is simple: a small probability can be expressed as the product of larger probabilities, which are therefore simpler to estimate. More specifically, if we seek to estimate:

$$\mathcal{I} = \mathbb{P}_{X \sim \mathbf{p}}(X \in \mathcal{F}), \quad (1.24)$$

and we know a sequence of nested sets

$$\mathcal{X} =: \mathcal{F}_0 \supset \mathcal{F}_1 \supset \dots \supset \mathcal{F}_L = \mathcal{F}, \quad (1.25)$$

then we can re-express \mathcal{I} using Bayes theorem :

$$\mathcal{I} = \mathbb{P}(X \in \mathcal{F}_L) = \prod_{\ell=1}^L \mathbb{P}(X \in \mathcal{F}_\ell \mid X \in \mathcal{F}_{\ell-1}). \quad (1.26)$$

We seek to estimate the probabilities $\mathbb{P}(X \in \mathcal{F}_\ell \mid X \in \mathcal{F}_{\ell-1})$ one after the other. To do this, we need to be able to simulate at each step $\ell = 1, \dots, L$ according to $\mathcal{L}(X \mid X \in \mathcal{F}_{\ell-1})$, i.e., according to a density proportional to $\mathbf{1}_{\mathcal{F}_{\ell-1}} \mathbf{p}$. If we could do this directly, we could estimate \mathcal{I} with a single draw (see proposition 6). The trick is that it is much easier (but not trivial) to transform a variable X already in \mathcal{F}_ℓ into another variable Y distributed according to \mathbf{p} on \mathcal{F}_ℓ , than to simulate Y from scratch. Let us first assume that we can simulate $Y \sim \mathbf{p}|_{\mathcal{F}_\ell}$ from a particle $x \in \mathcal{F}_\ell$ and a transition kernel $\tilde{Q}_\ell(\cdot \mid x)$. We will discuss the construction of such a kernel below.

1.2.1 Subset simulation

We first present a standard version of the algorithm. This is known as subset simulation or subset sampling (SS) [AB01] in reliability engineering, and as multilevel splitting in molecular dynamics (derived from [KH51]). However, one can find slight differences in formulation and implementation between these two versions depending on the sources and the time period. The operation of the method is described in Algorithm 2. The method is initialized with n independent particles drawn according to \mathbf{p} . The algorithm stops after L iterations (one for each subset \mathcal{F}_ℓ). At each iteration ℓ , a selection step is performed during which the particles that have not reached \mathcal{F}_ℓ are removed and replaced by copies of particles that have reached \mathcal{F}_ℓ . The copied particles are drawn uniformly. This selection step is followed by a mutation step where each particle makes a transition using the Markov kernel \tilde{Q}_ℓ . The intermediate probability

$\mathbb{P}(X \in \mathcal{F}_{\ell+1} \mid X \in \mathcal{F}_\ell)$ can therefore be estimated by the proportion of particles reaching $\mathcal{F}_{\ell+1}$ after mutation. The final estimator of the rare event probability \mathcal{I} is simply the product of the estimates of the intermediate probabilities.

There is a risk that the method will get stuck before iteration ℓ . This happens when no particle reaches the target region \mathcal{F}_ℓ . To avoid this, it is therefore necessary to choose carefully the proposal kernel \tilde{Q} , the number of subsets \mathcal{F}_ℓ , the subsets themselves, and the number of particles n .

Kernel choice

If a kernel Q is known for which the distribution \mathbf{p} is reversible, it is sufficient to draw a proposal according to Q and accept it only if it belongs to \mathcal{F}_ℓ to sample according to an admissible kernel \tilde{Q}_ℓ . One can apply Q several times without losing invariance to increase the chances of accepting at least one transition. Even in case of acceptance, it is recommended to apply Q several times to reduce the correlation between successive states of each particle, which also reduces the variance of the final estimator [Cér+12]. Once again, this is less difficult than drawing directly from \mathcal{F}_ℓ proportionally to \mathbf{p} , because the kernel Q usually makes local proposals around a point $x \in \mathcal{F}_\ell$. Most of the time, no such kernel Q is known, but it can be constructed using an MCMC method (see the Metropolis-Hastings method described in [algorithm 1](#)). A detailed analysis of the use of MCMC methods for subset sampling can be found in [Pap+15].

ALGORITHM 2 · Subset simulation

1 **Initialization** with n particles: $X_1^{(1)}, \dots, X_n^{(1)} \sim \mathbf{p}$.

2 **for** $\ell = 1, \dots, L$ **do**

3 **Selection step.** Let

$$J_\ell = \{k \in \llbracket 1, n \rrbracket : X_k^{(\ell)} \in \mathcal{F}_\ell\}. \quad (1.27)$$

4 **for** $k \in J_\ell$ **do** $Y_k^{(\ell)} = X_k^{(\ell)}$.

5 **for** $k \notin J_\ell$ **do** $K \sim \text{Unif}(J_\ell)$ and $Y_k^{(\ell)} = X_K^{(\ell)}$.

6 **Mutation step.**

7 **for** $k \in \llbracket 1, n \rrbracket$ **do** $X_k^{(\ell+1)} \sim \tilde{Q}_\ell(\cdot \mid Y_k^{(\ell)})$

8 **end**

9 **Estimation step.**

$$\hat{\mathcal{I}}_{n,L}^{ss} = n^{-L} \prod_{\ell=1}^L \text{Card}(J_\ell). \quad (1.28)$$

1.2.2 Adaptive multilevel splitting

Another way to consider the subsets \mathcal{F}_ℓ for $\ell = 1, \dots, L$ is to define a so-called importance function $\xi : \mathcal{X} \rightarrow \mathbb{R}$ and levels $-\infty = S_0 < S_1 < \dots < S_L = S^*$ such that $\xi(x) > S_\ell \iff x \in \mathcal{F}_\ell$. We are therefore trying to estimate $\mathcal{I} = \mathbb{P}(\xi(X) > S^*)$. This is a natural formulation for many rare events that can be written as threshold exceedances³.

It is known (see [C  r+12]) that the optimal distribution of the levels S_1, \dots, S_L (the one that minimizes the variance of the estimator $\hat{\mathcal{I}}^{ss}$) is evenly spaced in terms of probability of success. This means that the optimal choice is such that $\mathbb{P}(\xi(X) > S_{\ell+1} \mid \xi(X) > S_\ell)$ is constant regardless of ℓ . A natural way to free oneself from the choice of these thresholds is to determine them sequentially in the form of quantiles of the performance achieved by the particles (according to the importance function ξ). This idea was first mentioned in [Garoo] and then formalized in [CG07]. More precisely, at each iteration ℓ , the threshold $\bar{S}^{(\ell)}$ is set so that a proportion ρ (or a number $\lfloor \rho n \rfloor$) of particles have a score higher according to ξ . Therefore, this time we have a fixed and no longer random number of particles selected at each iteration. The intermediate probability $\mathbb{P}(\xi(X) > \bar{S}^{(\ell+1)} \mid \xi(X) > \bar{S}^{(\ell)})$ can be estimated by $\frac{\lfloor \rho n \rfloor}{n}$.

Dynamic rare event case

The historical applications of the method correspond to the case where X is the trajectory of a stochastic process, and one is interested in the probability that this trajectory reaches a region \mathcal{F} before a given stopping time. One usually considers an importance function of the form $\xi([X_t]_t) = \sup_t \beta(X_t)$ where β is a scoring function that quantifies the proximity of X_t to the set \mathcal{F} . The choice of β is then crucial [Dio+17; Vil07].

When Markovian, the dynamic case has a major advantage over the static case: one always knows a transition kernel that allows to simulate in \mathcal{F}_ℓ proportionally to \mathbf{p} . It is sufficient to clone the best trajectories up to the instant when they reached their peak (according to β), and to simulate the rest of the trajectories until the planned stopping time.

There are many theoretical results (convergence and variance analysis) for this version of the method [Br  +16; DG21; C  r+19].

1.3 Importance sampling

Importance Sampling, abbreviated as IS, is arguably the most well-known method for variance reduction and perhaps the simplest to describe in its raw form. The method is introduced under

³Care should be taken not to confuse multilevel splitting methods with multilevel Monte Carlo methods [Gil15]. The latter consists in estimating an expectation $\mathbb{E}_{\mathbf{p}}[\mathbf{I}(X)]$ using a sequence of auxiliary variables Y_1, \dots, Y_ℓ which approximates \mathbf{p} with increasing accuracy but with increasing cost. The term ‘‘level’’ refers to the level of accuracy of each auxiliary variable.

ALGORITHM 3 · Adaptive multilevel splitting

1 Initialization with n particles: $X_1^{(1)}, \dots, X_n^{(1)} \sim \mathbf{p}$. Set $\ell = 1$ and $\bar{S}^{(\ell)} = -\infty$.
2 while $\bar{S}^{(\ell)} \leq S^*$ **do**
3 Selection step.
4 for $k \in \llbracket 1, n \rrbracket$ **do** $S_k^{(\ell)} = \xi(X_k^{(\ell)})$.
5 Let σ_ℓ be a permutation on $\llbracket 1, n \rrbracket$ such that :

$$S_{\sigma_\ell(1)}^{(\ell)} \leq \dots \leq S_{\sigma_\ell(n)}^{(\ell)}. \quad (1.29)$$

 Let $\bar{S}^{(\ell)} := S_{\sigma_\ell(\lceil(1-\rho)n\rceil)}^{(\ell)}$ be the $1 - \rho$ quantile of $(S_k^{(\ell)})_{k=1}^n$.
6

$$J_\ell = \left\{ k \in \llbracket 1, n \rrbracket : S_k^{(\ell)} \leq \bar{S}^{(\ell)} \right\}. \quad (1.30)$$

7 for $k \in J_\ell$ **do** $Y_k^{(\ell)} = X_k^{(\ell)}$.
8 for $k \notin J_\ell$ **do** $K \sim \text{Unif}(J_\ell)$ and $Y_k^{(\ell)} = X_K^{(\ell)}$.
9 Mutation step.
10 for $k \in \llbracket 1, n \rrbracket$ **do** $X_k^{(\ell+1)} \sim \tilde{Q}_\ell(\cdot \mid Y_k^{(\ell)})$
11 end
12 Estimation step.

$$\hat{\mathcal{I}}_n^{\text{AMS}} = \left(\frac{\lceil \rho n \rceil}{n} \right)^\ell. \quad (1.31)$$

this name in 1951 by Kahn and Harris [KH51], but the idea was proposed in 1949, also by Kahn and other researchers working on the problem of neutron and gamma ray attenuation in thick shields⁴.

In this section, we present the fundamental principles of importance sampling, its application to rare event simulation, and some refinements. For more detailed information, we refer the reader to [Owe13, Chapter 9] for a general overview, [RT+09; Buco4] for an approach more focused on rare events, and [Bug+17; EM21] for reviews of recent advances.

We consider the framework of Monte Carlo methods (see section 1.1.2). The goal is to estimate a quantity of the form $\mathcal{I} := \mathbb{E}_{\mathbf{p}}[\mathbf{I}(X)]$, where X is a random variable taking values in a space \mathcal{X} and whose distribution is characterized by a probability density function (p.d.f.) \mathbf{p} with respect to a dominant measure μ . Finally, $\mathbf{I} : \mathcal{X} \rightarrow \mathbb{R}$ is a function of interest that we will sometimes refer to

⁴Readers interested in the historical development of the method can find more details in [And22].

as the target function.

1.3.1 The essentials

The principle of importance sampling is to estimate \mathcal{I} with a sample drawn not according to the reference distribution \mathbf{p} but under an alternative distribution characterized by a p.d.f. \mathbf{g} (still with respect to the measure μ), referred to as the “importance distribution” (also called “proposal”). The bias introduced by the change in distribution is corrected using weights given by the likelihood ratio \mathbf{p}/\mathbf{g} . More formally, under the assumption of support $\mathbf{I}(x)\mathbf{p}(x) \neq 0 \implies \mathbf{g}(x) \neq 0$ for μ -a.e. $x \in \mathcal{X}$, we have:

$$\begin{aligned} \mathcal{I} &:= \mathbb{E}_{\mathbf{p}}[\mathbf{I}(X)] = \int_{\mathcal{X}} \mathbf{I}(x) \mathbf{p}(x) \mu(dx) \\ &= \int_{\mathcal{X}} \mathbf{I}(x) \frac{\mathbf{p}(x)}{\mathbf{g}(x)} \mathbf{g}(x) \mu(dx) = \mathbb{E}_{\mathbf{g}}\left[\mathbf{I}(X) \frac{\mathbf{p}(X)}{\mathbf{g}(X)}\right]. \end{aligned} \quad (1.32)$$

This leads to the definition of an estimator for \mathcal{I} using importance sampling. Let X_1, \dots, X_n be an i.i.d. sample from the distribution \mathbf{g} , we have:

$$\hat{\mathcal{I}}_n^{\text{IS}} := \frac{1}{n} \sum_{k=1}^n \mathbf{I}(X_k) \frac{\mathbf{p}(X_k)}{\mathbf{g}(X_k)}. \quad (1.33)$$

Similar to the standard Monte Carlo estimator⁵, $\hat{\mathcal{I}}_n^{\text{IS}}$ is an unbiased, consistent, and asymptotically normal estimator of \mathcal{I} . Its variance is given by $\text{Var}(\hat{\mathcal{I}}_n^{\text{IS}}) = \sigma_{\mathbf{g}}^2/n$ where:

$$\sigma_{\mathbf{g}}^2 := \mathbb{E}_{\mathbf{g}}\left[\mathbf{I}(X)^2 \frac{\mathbf{p}(X)^2}{\mathbf{g}(X)^2}\right] - \mathcal{I}^2 = \mathbb{E}_{\mathbf{p}}\left[\mathbf{I}(X)^2 \frac{\mathbf{p}(X)}{\mathbf{g}(X)}\right] - \mathcal{I}^2. \quad (1.34)$$

It can be conventionally estimated, using the same i.i.d. sample X_1, \dots, X_n from the distribution \mathbf{g} , by:

$$\hat{\sigma}_{\mathbf{g}}^2 := \frac{1}{n-1} \sum_{k=1}^n \left(\mathbf{I}(X_k) \frac{\mathbf{p}(X_k)}{\mathbf{g}(X_k)} - \hat{\mathcal{I}}_n^{\text{IS}} \right)^2$$

Of course $\hat{\sigma}_{\mathbf{g}} = \sqrt{\hat{\sigma}_{\mathbf{g}}^2}$ is the estimator of the corresponding standard deviation. For any $\alpha \in [0, 1]$, we denote $\Phi^{-1}(\alpha)$ as the α -quantile of the $\mathcal{N}(0, 1)$ distribution. By applying the central limit theorem and Slutsky’s lemma, we have asymptotic validity of the following confidence interval:

$$\mathbb{P}\left(\mathcal{I} \in \left[\hat{\mathcal{I}}_n^{\text{IS}} - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{\hat{\sigma}_{\mathbf{g}}}{\sqrt{n}}; \hat{\mathcal{I}}_n^{\text{IS}} + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{\hat{\sigma}_{\mathbf{g}}}{\sqrt{n}} \right]\right) \xrightarrow{n \rightarrow \infty} 1 - \alpha. \quad (1.35)$$

⁵The importance sampling estimator in eq. (1.33) is ultimately just a classical Monte Carlo estimator for the last expectation in eq. (1.32).

Benefits of importance sampling

Importance sampling is a generalization of classical Monte Carlo estimation, which is recovered when choosing $\mathbf{g} = \mathbf{p}$. It is an interesting alternative to MCMC methods when it is not possible to directly simulate from the distribution \mathbf{p} . Another scenario arises when there is no sampling step, but the practitioner has access to a sample under a known distribution \mathbf{g} different from \mathbf{p} . In this case, using the trick in eq. (1.32), we can estimate $\mathbb{E}_{\mathbf{p}}[\mathbf{I}(X)]$ with the sample from the distribution \mathbf{g} . This is more commonly referred to as importance weighting (IW), and this “what-if” approach [Owe13, Section 9.14] has numerous applications, for example in deep learning [BL19], off-policy reinforcement learning [Mun+16], or in sensitivity analysis [Cha+18].

In this thesis, we focus on cases where sampling according to \mathbf{p} is a poor strategy to estimate \mathcal{I} due to the large variance $\sigma_{\mathbf{p}}^2$. The variance of the estimator $\hat{\mathcal{I}}_n^{\text{IS}}$ essentially relies on the choice of the importance distribution \mathbf{g} . It is observed that:

$$\sigma_{\mathbf{g}}^2 - \sigma_{\mathbf{p}}^2 = \mathbb{E}_{\mathbf{p}} \left[\mathbf{I}(X)^2 \frac{\mathbf{p}(X)}{\mathbf{g}(X)} \right] - \mathbb{E}_{\mathbf{p}}[\mathbf{I}(X)^2] = \mathbb{E}_{\mathbf{p}} \left[\mathbf{I}(X)^2 \left(\frac{\mathbf{p}(X)}{\mathbf{g}(X)} - 1 \right) \right]. \quad (1.36)$$

This implies that the variance of the importance sampling estimator for the importance density \mathbf{g} is smaller than the variance of the classical Monte Carlo estimator if and only if $\mathbb{E}_{\mathbf{p}} \left[\mathbf{I}(X)^2 \left(\frac{\mathbf{p}(X)}{\mathbf{g}(X)} - 1 \right) \right] < 0$.

Optimal importance sampling

There exists an optimal importance distribution that minimizes the variance $\sigma_{\mathbf{g}}^2$, and its expression is known.

PROPOSITION 6 · Optimal importance sampling

Assume that $\sigma_{\mathbf{p}}^2 < \infty$. Then the following function is the probability density function of the optimal importance distribution:

$$\mathbf{g}^* : x \mapsto \frac{|\mathbf{I}(x)| \mathbf{p}(x)}{\int_{\mathcal{X}} |\mathbf{I}(y)| \mathbf{p}(y) d\mu(y)}. \quad (1.37)$$

For any distribution \mathbf{g} we have: $\sigma_{\mathbf{g}}^2 \geq \sigma_{\mathbf{g}^*}^2$. Moreover, if the sign of \mathbf{I} is constant, then $\sigma_{\mathbf{g}^*}^2 = 0$. In other words, sampling only one realization according to \mathbf{g}^* would be sufficient to estimate \mathcal{I} exactly.

Proof. The application $x \mapsto x^2$ being convex, we deduce from Jensen's inequality that

$$\begin{aligned}\sigma_{\mathbf{g}}^2 &= \text{Var}_{\mathbf{g}}\left(\mathbf{I}(X) \frac{\mathbf{p}(X)}{\mathbf{g}(X)}\right) = \mathbb{E}_{\mathbf{g}}\left[\mathbf{I}(X)^2 \frac{\mathbf{p}(X)^2}{\mathbf{g}(X)^2}\right] - \mathcal{I}^2 \\ &\geq \mathbb{E}_{\mathbf{g}}\left[\left|\mathbf{I}(X)\right| \frac{\mathbf{p}(X)}{\mathbf{g}(X)}\right]^2 - \mathcal{I}^2 = \mathbb{E}_{\mathbf{p}}[|\mathbf{I}(X)|]^2 - \mathcal{I}^2.\end{aligned}\tag{1.38}$$

The equality case only occurs when the random variable $|\mathbf{I}(X)|\mathbf{p}(X)/\mathbf{g}(X)$ is a.s. constant. In that case, there exists a constant $c > 0$ such that for μ -a.e. $x \in \text{supp}(\mathbf{g})$ we have

$$\mathbf{g}(x) = \frac{|\mathbf{I}(x)|\mathbf{p}(x)}{c}.\tag{1.39}$$

This relationship is verified only by a single probability density function \mathbf{g}^* with the normalization constant $c = \int_{\mathcal{X}} |\mathbf{I}(x)|\mathbf{p}(x)\mu(dx)$. This density thus minimizes the variance $\sigma_{\mathbf{g}}^2$, the minimal value of which is therefore

$$\sigma_{\mathbf{g}^*}^2 = \mathbb{E}_{\mathbf{p}}[|\mathbf{I}(X)|]^2 - \mathcal{I}^2.\tag{1.40}$$

One immediately observes that this variance is zero if the sign of \mathbf{I} is constant. ■

The apparent power of this result is offset, on the one hand, by the fact that we do not know⁶ the quantity $\int_{\mathcal{X}} |\mathbf{I}(x)|\mathbf{g}(x)\mu(dx)$, and, on the other hand, by the fact that we do not know a priori how to sample proportionally to $|\mathbf{I}| \times \mathbf{p}$. Nevertheless, it provides an idea of what a good importance distribution should accomplish. The high variance of the Monte Carlo estimator arises because $|\mathbf{I}|$ and \mathbf{p} do not take high values in the same regions of \mathcal{X} . Therefore, it takes time to explore the “important” regions for $|\mathbf{I}|$ by sampling under \mathbf{p} . The correct approach is to sample according to a probability density function proportional to both $|\mathbf{I}|$ and \mathbf{p} to correct the mismatch and not favor one over the other.

Warning! Explosive variance

Determining a suitable importance distribution is highly challenging in practice, and it is far easier to end up with an increase in variance than a reduction in variance compared to a classical Monte Carlo method. While achieving optimal variance reduction is generally out of reach, the unwary practitioner may easily find themselves with an estimator of infinite variance⁷.

EXAMPLE 1 · Infinite-variance IS estimator

We seek to estimate $\mathcal{I} = \mathbb{E}_{\mathbf{p}}[X]$ where $\mathbf{p}(x) = \lambda e^{-\lambda x} \mathbf{1}_{x \geq 0}$ is the p.d.f. of an exponential random variable with parameter λ (thus $\mathcal{I} = \frac{1}{\lambda}$). Consider $\mathbf{g}(x) = \mu e^{-\mu x} \mathbf{1}_{x \geq 0}$ the p.d.f. of

⁶This is precisely the quantity we seek to estimate in the case where \mathbf{I} has a constant sign.

⁷The following example comes from the first version of [CD18] on arXiv: <https://arxiv.org/abs/1511.01437v1> but it does not appear on the final version.

an exponential random variable with parameter μ as the importance distribution. We have from eq. (1.34):

$$\begin{aligned}\sigma_{\mathbf{g}}^2 &= \text{Var}_{\mathbf{g}}\left(X \frac{\mathbf{p}(X)}{\mathbf{g}(X)}\right) - \mathbb{E}_{\mathbf{g}}\left[X \frac{\mathbf{p}(X)}{\mathbf{g}(X)}\right]^2 = \mathbb{E}_{\mathbf{p}}\left[X^2 \frac{\mathbf{p}(X)}{\mathbf{g}(X)}\right] - \mathbb{E}_{\mathbf{p}}[X]^2, \\ &= \frac{\lambda}{\mu} \mathbb{E}_{\mathbf{p}}\left[X^2 e^{-(\lambda-\mu)X}\right] - \frac{1}{\lambda^2} = \frac{\lambda}{\mu} \int_0^{\infty} x^2 e^{-(2\lambda-\mu)x} dx - \frac{1}{\lambda^2}.\end{aligned}\quad (1.41)$$

This last integral diverges when $\mu \geq 2\lambda$, thus $\sigma_{\mathbf{g}}^2 = +\infty$ in that case.

The likelihood ratio \mathbf{p}/\mathbf{g} is more likely to take high values when the proposal \mathbf{g} is more concentrated or has lighter tails than \mathbf{p} .

Application to rare event simulation

The most classical application of importance sampling as a variance reduction method is precisely the one that concerns us in this thesis: the simulation of rare events. We return to the framework of section 1.1.2. Let $\mathcal{F} \subset \mathcal{X}$, and define $\mathbf{I} : x \mapsto \mathbf{1}_{x \in \mathcal{F}}$. We aim to estimate the following quantity:

$$\mathcal{I} = \mathbb{E}_{\mathbf{p}}[\mathbf{I}(X)] = \mathbb{E}_{\mathbf{p}}[\mathbf{1}_{X \in \mathcal{F}}] = \mathbb{P}_{\mathbf{p}}(X \in \mathcal{F}). \quad (1.42)$$

As mentioned earlier, the classical Monte Carlo estimator proves to be inefficient when the probability \mathcal{I} is small. The use of importance sampling is clear: we generate a sample under an importance distribution \mathbf{g} that favors the occurrence of the event. The optimal importance distribution is proportional to $x \mapsto \mathbf{p}(x|x \in \mathcal{F})$. The event $X \in \mathcal{F}$ can often be reformulated as a threshold crossing $f(X) > s_{\mathcal{F}}$ ⁸. It is then necessary to find an importance distribution that strikes the right balance between pushing observations into the region \mathcal{F} and preserving a tail behavior that decreases like that of \mathbf{p} (and certainly not faster!). The problem is extensively addressed in the literature on industrial reliability, and a recent review can be found in [TJG22]. Let us take the application of the method to the event $X > s_{\mathcal{F}}$ where X follows a centered Gaussian distribution. For example if $s_{\mathcal{F}} = 4$ and $X \sim \mathcal{N}(0, 1)$ then $\mathcal{I} \approx 3.16 \times 10^{-5}$. It is demonstrated that the best importance distribution among all Gaussians with the same variance as X is the one with mean $s_{\mathcal{F}}$. Table 1.1 presents the values of four independent estimates of \mathcal{I} for a sample size of 10 variables and then for 100 variables. We can see that the right order of magnitude can be obtained with just a few dozen simulations, whereas at least 1 million would be needed to achieve the same degree of accuracy with a standard Monte Carlo method. We also illustrate the method on a less rare case for visualization in fig. 1.1.

⁸This connects the problem to extreme value theory and large deviations theory

TABLE 1.1 **Results of the estimation of $\mathbb{P}(\mathcal{N}(0, 1) > 4)$ using importance sampling with an importance distribution $\mathcal{N}(4, 1)$.** The experiment is independently repeated 4 times for each of the two sample sizes: $N = 10$ and $N = 100$.

Experiment	no. 1	no. 2	no. 3	no. 4
$\hat{\mathcal{I}}$ for $N = 10$	1.17×10^{-5}	1.79×10^{-5}	6.36×10^{-5}	2.08×10^{-5}
$\hat{\mathcal{I}}$ for $N = 100$	3.79×10^{-5}	3.44×10^{-5}	3.47×10^{-5}	2.52×10^{-5}

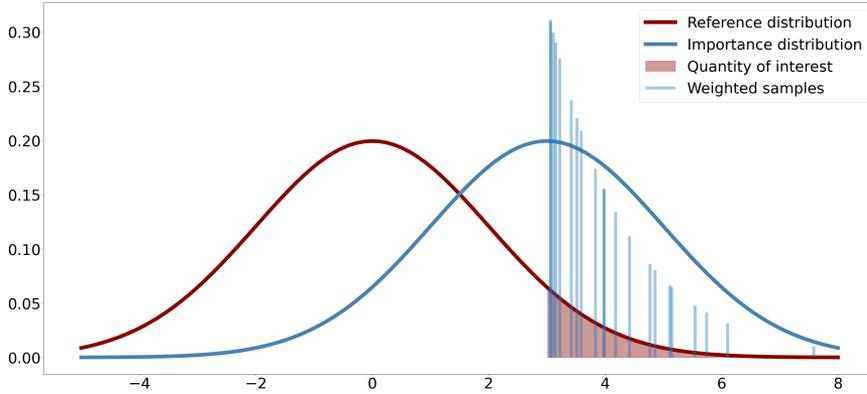


FIGURE 1.1 **Importance sampling estimation of $\mathbb{P}(X > s_{\mathcal{F}})$ with reference distribution $\mathcal{N}(0, 4)$, importance distribution $\mathcal{N}(s_{\mathcal{F}}, 4)$ and $s_{\mathcal{F}} = 3$.** The estimate is given by the mean value of the weighted samples, which corresponds to the mean height of the blue bars. It is not a very rare event but the weighted samples would have been too large and the red area too small for a clear visualization of the method with larger $s_{\mathcal{F}}$.

Self-normalized importance sampling

Importance sampling requires being able to explicitly evaluate the densities \mathbf{p} and \mathbf{g} at each sampled point. Sometimes, only unnormalized versions $\mathbf{p}_{\text{un}} = c_{\mathbf{p}}\mathbf{p}$ and $\mathbf{g}_{\text{un}} = c_{\mathbf{g}}\mathbf{g}$ of these densities are evaluable. The self-normalized importance sampling (SNIS) estimator circumvents this issue [Owe13, Chapter 9]. Let X_1, \dots, X_n be an i.i.d. sample from the distribution \mathbf{g} , we have:

$$\hat{\mathcal{I}}_n^{\text{SNIS}} = \frac{\sum_{k=1}^n \mathbf{I}(X_k) \frac{\mathbf{p}_{\text{un}}(X_k)}{\mathbf{g}_{\text{un}}(X_k)}}{\sum_{\ell=1}^n \frac{\mathbf{p}_{\text{un}}(X_{\ell})}{\mathbf{g}_{\text{un}}(X_{\ell})}} = \frac{\hat{\mathcal{I}}_n^{\text{IS}}}{\frac{1}{n} \sum_{k=1}^n \frac{\mathbf{p}(X_k)}{\mathbf{g}(X_k)}}. \quad (1.43)$$

Unlike classical importance sampling, this estimator is biased⁹, but the bias converges asymptoti-

⁹A recent variant aimed at reducing this bias can be found in [Car+22].

cally to zero. Indeed, by the strong law of large numbers, we have:

$$\frac{1}{n} \sum_{k=1}^n \frac{\mathbf{p}(X_k)}{\mathbf{g}(X_k)} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \mathbb{E}_{\mathbf{g}} \left[\frac{\mathbf{p}(X)}{\mathbf{g}(X)} \right] = 1, \quad (1.44)$$

and thus $\hat{\mathcal{I}}_n^{\text{SIS}} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \mathcal{I}$.

Multiple importance sampling

Importance sampling can be generalized to the case where multiple importance distributions are used and combined to construct an estimator for \mathcal{I} . This is referred to as multiple importance sampling (MIS) [VG95]. There are numerous sampling and weighting strategies, most of which are documented in [Elv+19].

Suppose we have n samples. For $i \in \llbracket 1, n \rrbracket$ each element of the sample $(X_{i,1}, \dots, X_{i,n_i})$ is distributed according to \mathbf{g}_i . An MIS estimator is determined by a weighting strategy, i.e., a sequence of functions $(w_i)_{i=1}^n$ from \mathcal{X} to \mathbb{R}_+ such that $\sum_{i=1}^n w_i(x) = 1$ for all $x \in \mathcal{X}$. A general formulation of an MIS estimator is as follows:

$$\hat{\mathcal{I}}^{\text{MIS}} = \sum_{i=1}^n \frac{1}{n_i} \sum_{k=1}^{n_i} w_i(X_{i,k}) \mathbf{I}(X_{i,k}) \frac{\mathbf{p}(X_{i,k})}{\mathbf{g}_i(X_{i,k})}. \quad (1.45)$$

This estimator is unbiased:

$$\begin{aligned} \mathbb{E}[\hat{\mathcal{I}}^{\text{MIS}}] &= \sum_{i=1}^n \frac{1}{n_i} \sum_{k=1}^{n_i} \mathbb{E}_{\mathbf{g}_i} \left[w_i(X) \mathbf{I}(X) \frac{\mathbf{p}(X)}{\mathbf{g}_i(X)} \right] = \sum_{i=1}^n \int_{\mathcal{X}} \mathbf{I}(x) \frac{\mathbf{p}(x)}{\mathbf{g}_i(x)} \mathbf{g}_i(x) w_i(x) \mu(dx) \\ &= \int_{\mathcal{X}} \mathbf{I}(x) \mathbf{p}(x) \sum_{i=1}^n w_i(x) \mu(dx) = \mathcal{I}. \end{aligned} \quad (1.46)$$

Let us first mention the two most common weighting schemes in MIS.

- ◊ We can adopt a classical approach and consider the weight $w_i(x) = \frac{1}{n}$ for all $x \in \mathcal{X}$ and all $i \in \llbracket 1, n \rrbracket$.
- ◊ Alternatively, we may consider that the observation $X_{i,k}$ was not drawn from \mathbf{g}_i but from a mixture of $(\mathbf{g}_j)_{j=1}^n$. We then use the weights $w_i(x) = \frac{\mathbf{g}_i(x)}{\sum_{j=1}^n \mathbf{g}_j(x)}$. This weighting scheme is called the balance heuristic and yields the following estimator:

$$\hat{\mathcal{I}}^{\text{MIS-BH}} = \sum_{i=1}^n \frac{1}{n_i} \sum_{k=1}^{n_i} \mathbf{I}(X_{i,k}) \frac{\mathbf{p}(X_{i,k})}{\sum_{j=1}^n \mathbf{g}_j(X_{i,k})}. \quad (1.47)$$

MIS methods are generally more robust than classical importance sampling. They make it easier to tackle multimodal target distributions using simple importance distributions. They have

gained popularity, particularly through the work of Eric Veach since the late 1990s, in the field of computer graphics (image synthesis, ray tracing, etc.)¹⁰. In his article [VG95, Theorem 1], he also proves that the variance of the MIS estimator with balance heuristic cannot be much larger than the best possible weighting scheme. For any MIS estimator $\hat{\mathcal{I}}^{\text{MIS}}$ with weighting scheme w_1, \dots, w_n , the following bound holds:

$$\text{Var}(\hat{\mathcal{I}}^{\text{MIS-BH}}) \leq \text{Var}(\hat{\mathcal{I}}^{\text{MIS}}) + \left(\frac{1}{\min(n_i)_{i=1}^n} - \frac{1}{\sum_{i=1}^n n_i} \right) \mathcal{I}^2. \quad (1.48)$$

The balance heuristic still serves as a reference [Pap+19], but can be too conservative. Numerous recent works attempt to generalize and surpass it [Elv+19; SE22], allowing for negative weights [HO14], optimizing weights [SS23; Kar+19], drawing more from certain \mathbf{g}_k [SHS16; SHS18], or incorporating information via estimated variances [Gri+19]. Also noteworthy is a generalization to a continuum of importance densities [Wes+20].

1.3.2 Adaptive procedures

Determining a good importance distribution (in terms of the variance of the generated IS estimator) is a complex task. Combining several distributions, as seen with the MIS method, reduces the risk of exploding variance but does not necessarily lead to a drastic reduction. The most commonly used approach in importance learning nowadays is sequential: the importance distribution is updated throughout the simulations to better approximate the target distribution. This is known as adaptive importance sampling (AIS).

For a given family of importance distributions \mathcal{G} , the process is as follows. At iteration $\ell = 1, \dots, L$:

- (i) **Sampling step:** draw a sample according to $\mathbf{g}^{(\ell)} \in \mathcal{G}$.
- (ii) **Adaptation step:** determine a new distribution $\mathbf{g}^{(\ell+1)} \in \mathcal{G}$ most often based on the drawn sample.

After a certain number of iterations L , we should have an effective importance distribution $\mathbf{g}^{(L)}$ that can be used for importance sampling estimation.

There are a large number of variants, including the choice of the proposal distribution family, sampling strategy (sample size per iteration, MIS, number of iterations), and adaptation strategy (updating the used importance distribution). Here, we provide the basics of the main methods and refer interested readers to [Bug+17] for a comprehensive review. More specific aspects, such as the convergence of recycling schemes of past samples [MPS19; DP18], will be addressed in the [chapter 4](#).

¹⁰Eric Veach's doctoral research on the MIS method was even honored with a technical Oscar in 2014 for "foundational research on efficient Monte Carlo path tracing for image synthesis."

Empirical variance minimization

The most natural idea is undoubtedly to directly search for the distribution $\mathbf{g} \in \mathcal{G}$ that minimizes the expected variance of the IS estimator [Buc04, Chapter 14], [RK16, Chapter 5.7].

Suppose that at iteration ℓ , we observe a sample X_1, \dots, X_n under an importance distribution $\mathbf{g}^{(\ell)} \in \mathcal{G}$. For any other importance distribution $\mathbf{g} \in \mathcal{G}$, we have:

$$\begin{aligned} \sigma_{\mathbf{g}}^2 + \mathcal{I}^2 &= \mathbb{E}_{\mathbf{g}} \left[\mathbf{I}(X)^2 \frac{\mathbf{p}(X)^2}{\mathbf{g}(X)^2} \right] = \mathbb{E}_{\mathbf{g}^{(\ell)}} \left[\mathbf{I}(X)^2 \frac{\mathbf{p}(X)^2}{\mathbf{g}(X)^2} \frac{\mathbf{g}(X)}{\mathbf{g}^{(\ell)}(X)} \right] \\ &= \mathbb{E}_{\mathbf{g}^{(\ell)}} \left[\mathbf{I}(X)^2 \frac{\mathbf{p}(X)}{\mathbf{g}(X)} \frac{\mathbf{p}(X)}{\mathbf{g}^{(\ell)}(X)} \right] \\ &\approx \frac{1}{n} \sum_{k=1}^n \mathbf{I}(X_k)^2 \frac{\mathbf{p}(X_k)}{\mathbf{g}(X_k)} \frac{\mathbf{p}(X_k)}{\mathbf{g}^{(\ell)}(X_k)}. \end{aligned} \quad (1.49)$$

Thus we can “improve” the quality of the next sample by choosing a new proposal $\mathbf{g}^{(\ell+1)}$:

$$\mathbf{g}^{(\ell+1)} \in \arg \min_{\mathbf{g} \in \mathcal{G}} \sum_{k=1}^n \mathbf{I}(X_k)^2 \frac{\mathbf{p}(X_k)}{\mathbf{g}(X_k)} \frac{\mathbf{p}(X_k)}{\mathbf{g}^{(\ell)}(X_k)}. \quad (1.50)$$

To solve the optimization problem eq. (1.50), one typically considers a parametric family of importance distributions $\mathcal{G} = (\mathbf{g}_{\theta})_{\theta \in \Theta}$ differentiable with respect to $\theta \in \Theta \subset \mathbb{R}^{d_{\Theta}}$. The optimization problem eq. (1.50) is then solved using gradient descent to find a zero of the function:

$$\theta \mapsto \sum_{k=1}^n \mathbf{I}(X_k)^2 \frac{\mathbf{p}(X_k)^2}{\mathbf{g}^{(\ell)}(X_k)} \frac{\nabla_{\theta} \mathbf{g}_{\theta}(X_k)}{\mathbf{g}_{\theta}(X_k)^2}. \quad (1.51)$$

The main drawback of this method is its instability. It tries to improve the quality of the next sample by relying on the current sample. If the current sample is of poor quality, the solution to eq. (1.50) can be a very poor candidate for minimizing $\sigma_{\mathbf{g}}^2$. Initialization is crucial. Moreover, we minimize a variance estimated by importance sampling, which involves a product of likelihood ratio, increasing the instability of the estimation¹¹.

Cross entropy minimization

The most popular alternative to empirical variance minimization is the so-called “Cross entropy” method [RK04]. Instead of anticipating the variance of the estimator, it seeks the importance distribution closest to the target distribution \mathbf{g}^* in terms of Kullback-Leibler (KL) divergence¹².

¹¹As we will see in section 1.3.3, likelihood ratios are subject to a degeneracy phenomenon that makes IS estimators unstable.

¹²The KL divergence is also called “Cross entropy” divergence, hence the method’s name.

DEFINITION 2 · Kullback-Leibler divergence

The KL divergence of two distributions characterized by p.d.f. \mathbf{p} and \mathbf{g} on \mathcal{X} is given by:

$$\mathcal{D}_{\text{KL}}(\mathbf{p}\|\mathbf{g}) = \mathbb{E}_{\mathbf{p}} \left[\log \left(\frac{\mathbf{p}(X)}{\mathbf{g}(X)} \right) \right] \quad (1.52)$$

The KL divergence quantifies the proximity between two distributions but it is not a distance (hence the term “divergence”) since it is not symmetric. Nevertheless it verifies the following property.

PROPOSITION 7 · Kullback-Leibler nonnegativity

For any pair of p.d.f. \mathbf{p} and \mathbf{g} with respect to a dominant measure μ we have:

- ◇ $\mathcal{D}_{\text{KL}}(\mathbf{p}\|\mathbf{g}) \geq 0$,
- ◇ $\mathcal{D}_{\text{KL}}(\mathbf{p}\|\mathbf{g}) = 0 \iff \mathbf{p}(x) = \mathbf{g}(x)$ for μ -a.e. $x \in \mathcal{X}$.

Proof. ◇ Recall that $\log x \leq x - 1$ for all $x > 0$. Indeed, $x \mapsto \log(x) - x + 1$ is a concave function whose derivative vanishes at 1, and therefore, it is bounded above by $\log(1) - 1 + 1 = 0$. Let us apply this inequality to $-\mathcal{D}_{\text{KL}}(\mathbf{p}\|\mathbf{g})$:

$$-\mathcal{D}_{\text{KL}}(\mathbf{p}\|\mathbf{g}) = \int_{\mathcal{X}} \mathbf{p}(x) \log \left(\frac{\mathbf{g}(x)}{\mathbf{p}(x)} \right) \mu(dx) \leq \int_{\mathcal{X}} \mathbf{p}(x) \left(\frac{\mathbf{g}(x)}{\mathbf{p}(x)} - 1 \right) \mu(dx) \quad (1.53)$$

And thus:

$$\mathcal{D}_{\text{KL}}(\mathbf{p}\|\mathbf{g}) \geq - \left(\int_{\mathcal{X}} \mathbf{g}(x) \mu(dx) - \int_{\mathcal{X}} \mathbf{p}(x) \mu(dx) \right) = 1 - 1 = 0. \quad (1.54)$$

- ◇ Since $\log x = x - 1$ if and only if $x = 1$, we have $\mathcal{D}_{\text{KL}}(\mathbf{p}\|\mathbf{g}) = 0$ if and only if $\frac{\mathbf{g}(x)}{\mathbf{p}(x)} = 1$ for μ -a.e. $x \in \mathcal{X}$. This gives us the desired result. ■

We are once again in the parametric framework $\mathcal{G} = (\mathbf{g}_{\theta})_{\theta \in \Theta}$ and we aim to determine the parameter θ^* minimizing $\theta \mapsto \mathcal{D}_{\text{KL}}(\mathbf{g}^* \|\mathbf{g}_{\theta})$. The following calculations explain the choice of this specific divergence¹³. It is possible to estimate an unbiased objective function equivalent to the KL divergence from a sample X_1, \dots, X_n distributed according to any proposal \mathbf{g} .

¹³It is worth adding that the KL divergence plays a central role in information theory and has many other interesting properties. An example of interpretation: the value of θ that minimizes an unbiased estimator of $\mathcal{D}_{\text{KL}}(\mathbf{g}^* \|\mathbf{g}_{\theta})$ with observations distributed under \mathbf{g}^* is the maximum likelihood estimator.

$$\begin{aligned}
\arg \min_{\theta \in \Theta} \mathcal{D}_{\text{KL}}(\mathbf{g}^* \parallel \mathbf{g}_\theta) &= \arg \min_{\theta \in \Theta} \mathbb{E}_{\mathbf{g}^*}[\log \mathbf{g}^*(X)] - \mathbb{E}_{\mathbf{g}^*}[\log \mathbf{g}_\theta(X)] \\
&= \arg \max_{\theta \in \Theta} \mathbb{E}_{\mathbf{g}^*}[\log \mathbf{g}_\theta(X)] \\
&= \arg \max_{\theta \in \Theta} \mathbb{E}_{\mathbf{p}}[|\mathbf{I}(X)| \log \mathbf{g}_\theta(X)] \\
&= \arg \max_{\theta \in \Theta} \mathbb{E}_{\mathbf{g}}\left[|\mathbf{I}(X)| \frac{\mathbf{p}(X)}{\mathbf{g}(X)} \log \mathbf{g}_\theta(X)\right].
\end{aligned} \tag{1.55}$$

This provides us with the following adaptation strategy. At each iteration ℓ , we have an importance distribution $\mathbf{g}^{(\ell)} := \mathbf{g}_{\theta^{(\ell)}}$, and we apply the two-step procedure:

(i) **Sampling step:** generate $X_1^{(\ell)}, \dots, X_n^{(\ell)} \sim \mathbf{g}^{(\ell)}$.

(ii) **Adaptation step:** solve

$$\theta^{(\ell+1)} \in \arg \max_{\theta \in \Theta} \sum_{k=1}^n |\mathbf{I}(X_k)| \frac{\mathbf{p}(X_k)}{\mathbf{g}^{(\ell)}(X_k)} \log \mathbf{g}_\theta(X_k). \tag{1.56}$$

A clear advantage over empirical variance minimization is the presence of only one likelihood ratio in the objective function estimation: the second likelihood ratio in eq. (1.50) is replaced by $\log \mathbf{g}_\theta$ in eq. (1.56).

Another reason for the popularity of this method is given by the following result [KR04]:

PROPOSITION 8 · Explicit minimization of CE for exponential families

Suppose the proposal belongs to a natural exponential family (NEF) [ML09], i.e., there exist known functions $x \mapsto h(x)$, $x \mapsto T(x)$, $\theta \mapsto \eta(\theta)$, and $\theta \mapsto A(\theta)$ such that:

$$\mathbf{g}_\theta(x) = h(x) \exp[\langle \eta(\theta), T(x) \rangle - A(\theta)]. \tag{1.57}$$

Then the optimization problem eq. (1.50) has an explicit solution.

Proof. First, we note that the gradient of the proposal is:

$$\nabla_\theta \log g_\theta(x) = T(x) - \nabla_\theta A(\theta). \tag{1.58}$$

We deduce that:

$$\nabla_{\theta} \sum_{k=1}^n |\mathbf{I}(X_k)| \frac{\mathbf{p}(X_k)}{\mathbf{g}^{(\ell)}(X_k)} \log \mathbf{g}_{\theta}(X_k) = 0 \quad (1.59)$$

$$\iff \sum_{k=1}^n |\mathbf{I}(X_k)| \frac{\mathbf{p}(X_k)}{\mathbf{g}^{(\ell)}(X_k)} [T(X_k) - \nabla_{\theta} A(\theta)] = 0 \quad (1.60)$$

$$\iff \nabla_{\theta} A(\theta) = \frac{\sum_{k=1}^n |\mathbf{I}(X_k)| \frac{\mathbf{p}(X_k)}{\mathbf{g}^{(\ell)}(X_k)} T(X_k)}{\sum_{k=1}^n |\mathbf{I}(X_k)| \frac{\mathbf{p}(X_k)}{\mathbf{g}^{(\ell)}(X_k)}}. \quad (1.61)$$

■

Exponential families encompass a wide range of common distributions but exclude certain heavy-tailed distributions, such as Student's t-distributions, which are valuable for importance sampling. A recent article [GCE23] considers an extension that includes Student's t-distributions: the λ -exponential family. It demonstrates that for these families, one can explicitly solve the minimization problem of an α -Rényi divergence [Rén61], of which the KL divergence is a special (limit) case.

Multilevel cross entropy

In rare event simulation, with $\mathbf{I}(X) = \mathbb{1}_{f(X) > s_{\mathcal{F}}}$, one either needs to generate a very large number of variables in the first iteration to ensure that at least one realizes the event of interest, or initialize the Cross Entropy method with a good importance distribution from the start. A refinement of the method called “multilevel Cross Entropy” (mCE) has been proposed to address this situation (it is detailed in [KR04; RK16; De +05]).

The idea is as follows: instead of seeking an importance distribution adapted to the estimation of $\mathbb{P}_{\mathbf{p}}(f(X) > s_{\mathcal{F}})$ at each iteration, we set intermediate thresholds $\gamma_1, \dots, \gamma_T \leq s_{\mathcal{F}}$ that are easier to reach. At iteration ℓ , we seek:

$$\theta^{(\ell+1)} \in \arg \max_{\theta \in \Theta} \sum_{k=1}^n \mathbb{1}_{f(X_k) > \gamma_{\ell}} \frac{\mathbf{p}(X_k)}{\mathbf{g}^{(\ell)}(X_k)} \log \mathbf{g}_{\theta}(X_k). \quad (1.62)$$

This allows distributing the effort across iterations. There exists a sequential procedure for setting these thresholds. Given a sample X_1, \dots, X_n , let $Y_k = f(X_k)$ for all $k \in \llbracket 1, n \rrbracket$, and $Y_{(k)}$ denote the k -th order statistic. Let $\rho \in [0, 1]$, and the following rule is applied:

$$\gamma_{\ell} = \min(s_{\mathcal{F}}, Y_{(\lceil (1-\rho)n \rceil)}). \quad (1.63)$$

Thus, at each iteration, we determine the most suitable importance density within our family \mathcal{G} to estimate the probability of achieving better than a proportion $(1 - \rho)$ of the observed draws.

This principle also allows the construction of original optimization methods (especially combinatorial optimization) [Bot+13]. When seeking to maximize a function \mathbf{I} , random points can be drawn, and the multilevel Cross entropy method can be used to sequentially determine an importance distribution that draws points for which \mathbf{I} takes increasingly higher values.

Finally, a drawback of the method is the lack of smoothness of the indicator function $\mathbb{1}_{f(X) > \gamma}$. It introduces threshold effects (a distribution adapted to the event $f(X) > \gamma$ may be unable to achieve the event $f(X) > \gamma + \varepsilon$), and implies using only a fragment of the simulated variables. An alternative version of the Multilevel Cross entropy method, called the improved Cross entropy method (iCE), has been proposed in [PGS19]. It replaces the indicator with a smoother function: $I_\gamma(X) = \Phi\left(-\frac{s_F - f(X)}{\gamma}\right)$ with Φ being the cumulative distribution function of a standard normal distribution. We have $I_\gamma(X) \xrightarrow{\gamma \rightarrow 0} \mathbb{1}_{f(X) > s_F}$. The hyperparameter γ is not updated using quantiles this time but is chosen as small as possible such that the anticipated coefficient of variation is below a fixed threshold δ_{target} :

$$\gamma_\ell = \arg \min_{\gamma \in (0, \gamma_{\ell-1})} \left(\frac{\hat{\sigma}_g(\gamma)}{\hat{\mathcal{I}}_n^{\text{IS}}(\gamma)} - \delta_{\text{target}} \right)^2. \quad (1.64)$$

Non parametric adaptive importance sampling

To make the parametric approaches described so far work, one needs either a good intuition about the form of the parametric family used or a family of very high dimension. However, the latter confronts us with other problems, as described in section 1.3.3. Nonparametric approaches, inspired by traditional density estimation methods in statistics, have been proposed to overcome the problem.

When referring to nonparametric adaptive importance sampling (NAIS), we generally mean an approach introduced in the 1990s based on a kernel density estimation (KDE), also called kernel smoothing [WJ94]. In its simplest version [AAT92], the importance density used at iteration $\ell + 1$ is constructed from a sample X_1, \dots, X_n from distribution $\mathbf{g}^{(\ell)}$:

$$\mathbf{g}^{(\ell+1)}(x) := \frac{1}{n} \sum_{k=1}^n \mathbf{I}(X_k) \frac{\mathbf{p}(X_k)}{\mathbf{g}^{(\ell)}(X_k)} K_{\mathbf{H}}(x - X_k), \quad (1.65)$$

where $K_{\mathbf{H}}$ is most often a Gaussian kernel (i.e., the density of a multivariate Gaussian distribution with mean zero and covariance matrix \mathbf{H}). There are many refinements by varying the kernel [Ned09], sequentially adjusting its hyperparameters [GR96], and estimating the importance density using all past samples [Zha96]. Note that the use of kernels is proposed in a generalized version of the Cross entropy method [BK11].

Finally, recent advances in machine learning and deep learning methods provide excellent tools for approximating nonlinear functions in high dimensions. The family of non-parametric importance sampling methods thus includes techniques based on kriging [BMM13; ZWS20],

support vector machines [LL21], variational autoencoders [Dem+23], and normalizing flows [Mül+19; DJ24].

1.3.3 Curse of dimensionality

Like many methods in applied mathematics, importance sampling is not immune to the curse of dimensionality. Determining an effective importance distribution in high-dimensional spaces is extremely challenging. This is not surprising in the context of AIS methods because either the family \mathcal{G} is not flexible enough to approximate the target distribution, or it is, but the adaptation step becomes impractical with the allowed sample sizes. However, the major issue remains the phenomenon of weight degeneracy, which can occur even in low dimensions but appears inevitable as the dimension increases.

Degeneracy of the weights

When the importance distribution differs too much from the target distribution, especially in the tails of the distribution, we observe a phenomenon called weight degeneracy. This is a situation where the overwhelming majority of likelihood ratios of simulated variables are close to 0, and only a few carry most of the weight. This inevitably increases the estimator's variance since we find ourselves in a situation similar to the rare event case: we waste the majority of simulations and assign disproportionate weight to a few observations that are not representative of the target distribution if we have not generated enough variables. Here is a simple example from [RK16, Section 5.7] where this situation occurs.

EXAMPLE 2 · Degeneracy of the weights for vectors with i.i.d. coordinates

We are interested in the behavior of the variable $\frac{\mathbf{p}(X)}{\mathbf{g}(X)}$ when $X \sim \mathbf{g}$. We assume that $X \in \mathbb{R}^d$ and that $\mathbf{p}(X) = \prod_{i=1}^d p(X_i)$ and $\mathbf{g}(X) = \prod_{i=1}^d g(X_i)$. On the one hand, independently of the dimension d :

$$\mathbb{E}_{\mathbf{g}} \left[\frac{\mathbf{p}(X)}{\mathbf{g}(X)} \right] = \int_{\mathbb{R}^d} \mathbf{p}(x) \mu(dx) = 1. \quad (1.66)$$

On the other hand:

$$\frac{\mathbf{p}(X)}{\mathbf{g}(X)} = \prod_{i=1}^d \frac{p(X_i)}{g(X_i)} = \exp \left[\sum_{i=1}^d \log \left(\frac{p(X_i)}{g(X_i)} \right) \right]. \quad (1.67)$$

By the law of large numbers, for large d , we have the approximation:

$$\sum_{i=1}^d \log\left(\frac{p(X_i)}{g(X_i)}\right) \approx d \mathbb{E}_{\mathbf{g}}\left[\log\left(\frac{p(X_1)}{g(X_1)}\right)\right] = -d \mathbb{E}_{\mathbf{g}}\left[\log\left(\frac{g(X_1)}{p(X_1)}\right)\right] < 0. \quad (1.68)$$

The last inequality follows from the positivity of the Kullback-Leibler divergence. We finally deduce that:

$$\frac{\mathbf{p}(X)}{\mathbf{g}(X)} \approx \exp\left(-d \mathbb{E}_{\mathbf{g}}\left[\log\left(\frac{g(X_1)}{p(X_1)}\right)\right]\right) \xrightarrow{d \rightarrow \infty} 0. \quad (1.69)$$

The paradox in [example 2](#) is that the weights tend to zero as the dimension increases, while their mean value remains 1. This implies that in high dimensions, a few weights will take very large values to compensate for the others. The phenomenon can be observed in [fig. 1.2](#), a specific numerical application of [example 2](#) inspired by [\[BBL08\]](#). The reference distribution \mathbf{p} and the importance distribution are both multivariate Gaussian of dimension d with an identity covariance matrix. The mean is zero everywhere for the \mathbf{p} distribution and equal to 0.1 everywhere for the \mathbf{g} distribution. We depict the histogram of the ratio $\frac{\mathbf{p}}{\mathbf{g}}$ for several values of d , and we observe weights that collapse toward 0 while their empirical mean remains close to 1.

Stabilization of importance sampling

Several approaches aim to modify the weights, the target distribution, or the arrangement of proposals to stabilize the IS estimator.

The most well-known method is probably the truncated importance sampling method [\[Iono8\]](#), also called clipping [\[Mar+18\]](#). It belongs to a broader class of methods known as nonlinear importance sampling (NIS) [\[Mig17; MMV18\]](#), which involves applying a nonlinear transformation to the importance weights to prevent their degeneracy. By normalizing the transformed weights, an estimator with asymptotically vanishing bias is obtained. Clipping applies the transformation $w \mapsto \min(w, s)$, where s is a threshold set in advance or based on the data, to each weight. This prevents a small number of weights from being much larger than the others. A smoother version of clipping is presented in [\[Veh+15\]](#).

Another family of approaches can be grouped under the name tempering [\[APB22\]](#). In its simplest version, it involves replacing the target distribution $g^*(x)$ with $g^*(x)^\beta$ with $\beta < 1$. The new target distribution is then “flattened” making it easier to approximate (and the likelihood ratios are less sensitive to small variations in x). The power β can be fixed or evolve sequentially. This philosophy might bring to mind a weighted likelihood method [\[HZ02\]](#) to give less weight to the data, simulated annealing [\[DCM19\]](#) and distribution interpolation [\[ABV23\]](#) to connect a simple pseudo-target distribution to a complex target distribution.

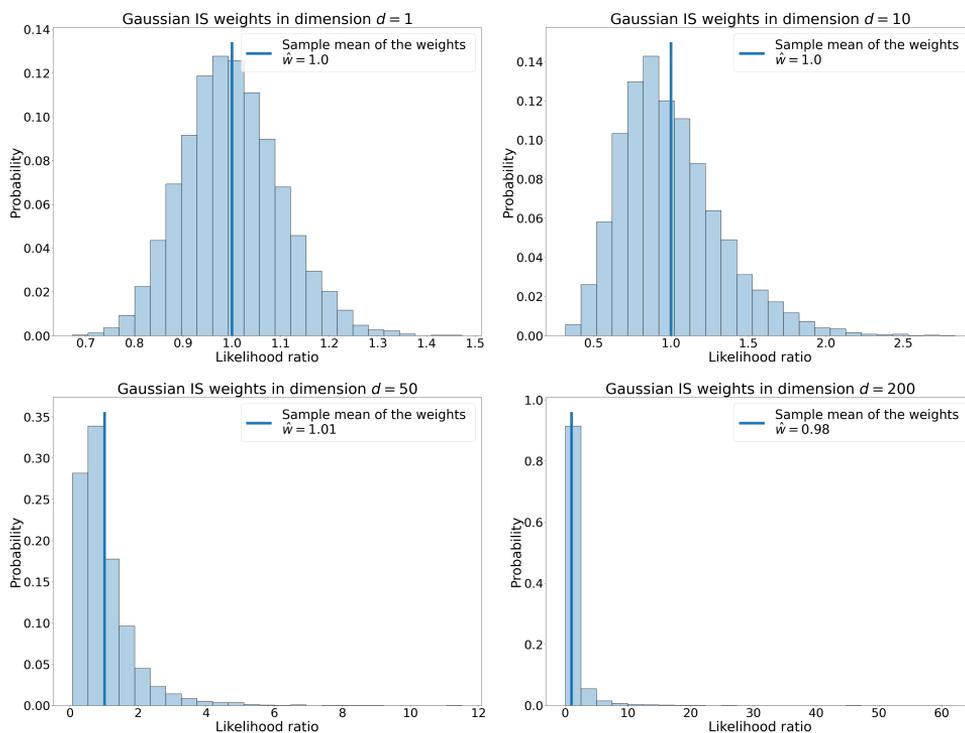


FIGURE 1.2 **Histogram of likelihood ratios between two multivariate Gaussian variables for different values of the dimension d .** Reference distribution: $\otimes_{i=1}^d \mathcal{N}(0, 1)$. Importance distribution: $\otimes_{i=1}^d \mathcal{N}(0.1, 1)$. Sample size of 10^4 in the four cases.

It is often considered good practice to include the reference distribution \mathbf{p} as a component of a mixture for static importance sampling (in this case, MIS) or as a member of the family \mathcal{G} in AIS to achieve at least the performance of classical Monte Carlo. This is known as defensive importance sampling [Hes95]. It is worth noting that a mixture $\mathbf{g}_\lambda = \lambda\mathbf{p} + (1-\lambda)\mathbf{g}$ upper-bounds the likelihood ratio:

$$\frac{\mathbf{p}(x)}{\mathbf{g}_\lambda(x)} \leq \frac{\mathbf{p}(x)}{\lambda\mathbf{p}(x)} = \frac{1}{\lambda}. \quad (1.70)$$

This stabilization strategy for importance sampling using a mixture has been generalized into “safe importance sampling” [OZ00; DP21].

Many recent works propose new, more advanced methods to overcome the dimensionality challenge. Notably, there are dimensionality reduction methods employing projection into lower-dimensional subspaces [EMS21a; EMS21b; Uri+21; Ehr+23]. Other approaches suggest more efficient non-informed proposals in high dimensions compared to Gaussian mixtures (which are strongly discouraged beyond $d > 50$), such as von Mises-Fisher mixtures [WS16; PGS19].

CHAPTER 2

Piecewise deterministic Markov processes

The central mathematical object of this thesis is the PDMP (piecewise deterministic Markov process). The aim of this chapter is to provide the reader with all the notions about PDMPs needed to fully understand the contributions of this thesis. A PDMP is a non-diffusive continuous-time jump stochastic process. Its specificity (which actually makes it a very general process) is its hybrid behavior. On the one hand, it takes values on a space composed of a continuous part and a discrete part. On the other hand, it evolves according to a deterministic dynamics between jumps, and jumps both at random and deterministic times. This behavior is actually quite intuitive for readers familiar with inhomogeneous Poisson processes. We first recall the basics of these processes (which are special cases of PDMPs), before addressing the essential characteristics of PDMPs. Finally, we propose classical approaches for simulating PDMP trajectories.

Content of this chapter

2.1	Poisson process	34
2.1.1	Introduction to Markovian jump processes	34
2.1.2	The simple Poisson process	39
2.1.3	Inhomogeneous Poisson process	44
2.2	PDMP essentials	46
2.2.1	Local characteristics of the PDMP	47
2.2.2	Strong Markov PDP	52
2.3	Modeling and simulation with PDMPs	56
2.3.1	PDMP superposition	57
2.3.2	PDMP thinning	61
2.3.3	The inversion method with a twist	62

2.1 Poisson process

In this section, we introduce the fundamentals of Poisson processes on \mathbb{R}_+ . The intuition behind these processes is simple. Events occur successively at random times. The counting process associated with these events provides, at each instant $t > 0$, the number of events already observed up to time t . A Poisson process is a counting process associated with events whose occurrence times are mutually independent. We refer to a simple Poisson process (also called homogeneous or stationary) if the probability of observing an event in a given time interval is identical for any other interval of the same length. Otherwise, the Poisson process is said to be inhomogeneous or non-stationary.

Beyond modeling numerous physical, biological, or social phenomena, Poisson processes also describe the mechanics of jump times occurring in piecewise deterministic Markov processes at the heart of this thesis.

2.1.1 Introduction to Markovian jump processes

A stochastic process is a collection of random variables with values in a given state space, usually (but not necessarily) indexed by time. A stochastic process indexed by time is said to be Markovian when it satisfies a memoryless property: the distribution of its future states given current and past states only depends on the current state. The randomness of continuous-time Markovian stochastic processes can manifest itself in two ways: continuously at every instant of the trajectory (these are referred to as diffusion processes), and punctually at given instants of the trajectory (these are known as jump processes). A Markov process can exhibit both diffusive and jump characteristics. In the context of this thesis, we focused on piecewise deterministic Markov processes, a class that encompasses all non-diffusive Markov processes. We will, therefore, leave diffusion processes aside and concentrate on Markovian pure jump processes.

DEFINITION 3 · Stochastic process

Let \mathfrak{T} be an abstract index set. A stochastic process indexed on \mathfrak{T} is a collection of random variables $(X_t)_{t \in \mathfrak{T}}$ from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to a measurable space $(\mathcal{X}, \mathcal{X})$.

Only stochastic processes indexed by time will be considered in the following. Thus \mathfrak{T} is a totally ordered set representing time and $\inf(\mathfrak{T}) = 0$ (so X_0 is the initial state of the stochastic process $(X_t)_{t \in \mathfrak{T}}$). If \mathfrak{T} is a countable set (typically \mathbb{N}), then $(X_t)_{t \in \mathfrak{T}}$ is a discrete-time stochastic process. If $\mathfrak{T} = [0, a)$ with $a \in \mathbb{R}_+ \cup \{+\infty\}$, then $(X_t)_{t \in \mathfrak{T}}$ is a continuous-time stochastic process. We will also adopt the following notation conventions:

- ◇ for any $B \subseteq \mathfrak{T}$, we note $X_B = (X_t)_{t \in B}$,

- ◇ uppercase variables are used for random variables (or processes) and lowercase variables for fixed realizations.

In order to more easily express the conditioning of a random variable with respect to the past of the process, we also introduce the definition of a filtration.

DEFINITION 4 · Filtration

Let $(\mathcal{F}_t)_{t \in \mathfrak{T}}$ be a collection of sub- σ -algebras of \mathcal{F} . The collection $(\mathcal{F}_t)_{t \in \mathfrak{T}}$ is called a filtration if it is non-decreasing in terms of set inclusion: $\mathcal{F}_s \subseteq \mathcal{F}_{s+t}$ for any $s, t \in \mathfrak{T}$.

Equipped with this filtration, the probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathfrak{T}}, \mathbb{P})$ is called a filtered probability space.

EXAMPLE 3 · Natural filtration of the process

For any $s \in \mathfrak{T}$, let \mathcal{F}_s be the σ -algebra generated by the process until time s (included), i.e.:

$$\mathcal{F}_s := \sigma(X_{[0,s]}) = \sigma(\{X_t^{-1}(A_t) : A_t \in \mathcal{X}, t \leq s\}). \quad (2.1)$$

Then $(\mathcal{F}_t)_{t \in \mathfrak{T}}$ is a filtration and is called the natural filtration of \mathcal{F} with respect to $X_{\mathfrak{T}}$. We also note $\mathcal{F}_s^- := \sigma(X_{[0,s)})$ the σ -algebra generated by the process before time s .

Markov property

A stochastic process is said to be Markovian if it verifies the following memoryless property.

DEFINITION 5 · Markov property

The process $X_{\mathfrak{T}}$ is Markovian if, for any $B \in \mathcal{X}$ and any s, t such that $s, s+t \in \mathfrak{T}$, it verifies:

$$\mathbb{P}(X_{s+t} \in B \mid \mathcal{F}_s) = \mathbb{P}(X_{s+t} \in B \mid X_s). \quad (2.2)$$

An alternative formulation is the following, for any bounded measurable function $f : \mathcal{X} \rightarrow \mathbb{R}$, we have:

$$\mathbb{E}[f(X_{s+t}) \mid \mathcal{F}_s] = \mathbb{E}[f(X_{s+t}) \mid X_s]. \quad (2.3)$$

In other words, knowing the value of X_s gives us the same insight into the future state distribution as having complete knowledge of the entire process history up to time s . Although this property might appear restrictive at first glance, it is often feasible to consider an analogous process that satisfies it on an extended version of the state space \mathcal{X} . This extension is chosen such that each state encompasses the necessary information from previous states that makes the

distribution of the next state explicit. Here is a very simple example with an autoregressive model of order 2.

EXAMPLE 4 · Autoregressive model AR(2)

Let $a_0, a_1, a_2 \in \mathbb{R}$ and $(\varepsilon_n)_{n \in \mathbb{N}}$ be a sequence of i.i.d. standard Gaussian variables $\mathcal{N}(0, 1)$. We define $(X_n)_{n \in \mathbb{N}}$ the autoregressive model of order 2 (denoted by AR(2)) satisfying for any $n \geq 2$:

$$\begin{aligned} X_0 &= a_0 + \varepsilon_0, \\ X_1 &= a_0 + a_1 X_1 + \varepsilon_1, \\ X_n &= a_0 + a_1 X_{n-1} + a_2 X_{n-2} + \varepsilon_n. \end{aligned} \tag{2.4}$$

The process $X_{\mathbb{N}}$ is obviously not Markovian since the distribution of the state X_{n+1} given $(X_k)_{k \leq n}$, also relies on X_{n-1} (and not only X_n) for any $n \geq 2$. However, the process $(Y_n)_{n \in \mathbb{N}^*}$ with values in \mathbb{R}^2 and defined by $Y_n = (X_n, X_{n-1})$ for $n \geq 1$, is Markovian.

Another example of such extension of the state space is the time-augmented process where the elapsed time completes the state variable X . This trick will be used several times in this manuscript.

DEFINITION 6 · Time-augmented process

Let $(X_t)_{t \in \mathfrak{S}}$ be a stochastic process with values in a state space \mathcal{X} and indexed by time in a set \mathfrak{S} . The time-augmented process associated with $(X_t)_t$ is the stochastic process $(X_t, t)_{t \in \mathfrak{S}}$ with values in $\mathcal{X} \times \mathfrak{S}$.

A specific case of Markov processes that is simpler to study is the time-homogeneous case (sometimes simply referred to as "homogeneous").

DEFINITION 7 · Time-homogeneous process

The stochastic process $X_{\mathfrak{S}}$ is said time-homogeneous if it is Markovian and the distribution of its future states does not depends on elapsed time, i.e. for any $B \in \mathcal{X}$, any $x \in \mathcal{X}$ and any s, t such that $s, s + t \in \mathfrak{S}$ we have:

$$\mathbb{P}(X_{s+t} \in B \mid X_s = x) = \mathbb{P}(X_t \in B \mid X_0 = x) \tag{2.5}$$

Under the right conditions, the distribution of successive states of a homogeneous Markov process can be characterised by operators whose properties are very useful in practice (semigroup, infinitesimal generator, etc.). A process can generally be made homogeneous by considering the associated time-augmented process, but at the cost of making the state space more complex

(especially if \mathcal{X} is finite or countable whereas \mathfrak{T} is continuous).

Strong Markov property

The Markov property holds that the distribution of states of the process after a fixed date $s \in \mathfrak{T}$ given all the past of the process up to that date depends only on the state of the process at time s . Under the right conditions, this property can be extended to the case of a random date τ . This is known as the strong Markov property. This random date must be a stopping time for the filtered probability space.

DEFINITION 8 · Stopping time

A random variable τ , with values in \mathfrak{T} , is called a stopping time with respect to the filtration $(\mathcal{F}_t)_{t \in \mathfrak{T}}$ if for any $t \in \mathfrak{T}$:

$$\{\tau \leq t\} \in \mathcal{F}_t. \quad (2.6)$$

To condition a process on all its past before stopping time τ , we define the σ -algebra on τ -past:

$$\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty : \forall t \in \mathfrak{T}, A \cap \{\tau \leq t\} \in \mathcal{F}_t\}, \quad (2.7)$$

with $\mathcal{F}_\infty = \sigma(\cup_{t \in \mathfrak{T}} \mathcal{F}_t)$.

DEFINITION 9 · Strong Markov property

The process $X_{\mathfrak{T}}$ is a strong Markov process if, for any stopping time τ with respect to the filtration $(\mathcal{F}_t)_{t \in \mathfrak{T}}$, any time $t \in \mathfrak{T}$ and any $B \in \mathcal{X}$, it verifies:

$$\mathbb{P}(X_{\tau+t} \in B \mid \mathcal{F}_\tau) = \mathbb{P}(X_{\tau+t} \in B \mid X_\tau). \quad (2.8)$$

An alternative formulation is the following, for any bounded measurable function $f : \mathcal{X} \rightarrow \mathbb{R}$, we have:

$$\mathbb{E}[f(X_{\tau+t}) \mid \mathcal{F}_\tau] = \mathbb{E}[f(X_{\tau+t}) \mid X_\tau]. \quad (2.9)$$

Put simply, the distribution of the future states of a strong Markov process after a stopping time τ depends neither on the past of the process before τ (as in the classical Markov property), nor on the value of τ itself (conditioning on X_τ). All discrete-time Markov processes (\mathfrak{T} countable) are strong Markov processes. It is trickier for continuous-time processes. An example of a Markov process that is not strong Markov is given in [FG13, Problem 13 pp. 626–627, Chapter 31]¹.

¹The example is given in [FG13] as an exercise. The interested reader can find a solution in the *Stochastic*

Jump times of a Markovian jump process

A pure jump process is a process in which randomness only occurs at specific time points along its trajectory, causing events called jumps. These jumps trigger a change in the state of the process. The distribution of a pure jump process is therefore entirely characterised by the distribution of its jump times and the distribution of its state after each jump. If the jump process is Markovian, then the distribution of the next jump time relies only on the current state of the process and not on past states. In particular, conditioning on the current state, it does not depend on the time elapsed since the previous jump. Thus, not all distributions are permitted.

EXAMPLE 5 · A non-Markov jump process

Let $(X_t)_{t \in \mathbb{R}_+}$ be a time-continuous jump process with values in \mathbb{N} . Let T_0, T_1, \dots be i.i.d. random variables with uniform distribution on $[0, 1]$, and for any $n \in \mathbb{N}^*$, let $S_n = S_{n-1} + T_{n-1}$ be the time of the n -th jump (with $S_0 = 0$). The process $(X_t)_{t \in \mathbb{R}_+}$ starts at $X_0 = 0$, is constant between jumps and increments by 1 after each jump. In other words, for any $s, t \in \mathbb{R}_+$, $X_{s+t} - X_s$ counts the number of jumps occurred in interval $(s, s+t]$:

$$\begin{aligned} X_0 &= 0, \\ X_{s+t} &= X_s + \sum_{n=1}^{\infty} \mathbb{1}_{S_n \in (s, s+t]}. \end{aligned} \quad (2.10)$$

The process $X_{\mathbb{R}_+}$ is not Markovian. Indeed, the distribution of the next jump time knowing the past of the trajectory up to the current state depends on the time of the last jump and therefore not only on the current state. Let $s, t > 0$, $\varepsilon \in (0, \min(s, 1))$ and let $n \in \mathbb{N}^*$. On the one hand, we have:

$$\begin{aligned} \mathbb{P}(X_{s+t} = n+1 \mid X_s = n, S_n = s - \varepsilon) &= \mathbb{P}(T_n \leq t + \varepsilon \mid T_n > \varepsilon), \\ &= \frac{F_{T_n}(t + \varepsilon) - F_{T_n}(\varepsilon)}{1 - F_{T_n}(\varepsilon)}, \\ &= \min\left(\frac{t}{1 - \varepsilon}, 1\right), \end{aligned} \quad (2.11)$$

with $F_{T_n} : x \in \mathbb{R} \mapsto \min(1, \max(0, x))$ the c.d.f. of $T_n \sim \text{Unif}([0, 1])$. On the other hand, $\mathbb{P}(X_{s+t} = n+1 \mid X_s = n)$ cannot be a function of ε .

Let us now assume that the waiting times between jumps from [example 5](#) are drawn i.i.d. according to an exponential distribution with parameter $\lambda > 0$ instead of a uniform distribution on $[0, 1]$. We note $\tilde{F} : x \in \mathbb{R} \mapsto 1 - e^{-\lambda x} \mathbb{1}_{x \geq 0}$ their c.d.f. The memoryless property of the

Processes (Advanced Probability II) course from Cosma Shalizi's website: [Example 167 pp. 96, Chapter 13. Strong Markov, Martingales.](#)

exponential distribution leads to the following consequence in eq. (2.11):

$$\mathbb{P}(X_{s+t} = n + 1 \mid X_s = n, S_n = s - \varepsilon) = \frac{\tilde{F}(t + \varepsilon) - \tilde{F}(\varepsilon)}{1 - \tilde{F}(\varepsilon)} = 1 - e^{-\lambda t}. \quad (2.12)$$

The distribution of the waiting time before the next jump no longer depends on ε and therefore on the time already waited since the previous jump. The jump process is then Markovian, and more precisely in that case, a *simple Poisson process*.

2.1.2 The simple Poisson process

A Poisson process is a counting process, i.e. an increasing, continuous-time stochastic process $(N_t)_{t \in \mathbb{R}_+}$ with values in \mathbb{N} . Events are assumed to occur punctually. The process counts the number of events occurring over time. In particular, for any $s, t \in \mathbb{R}_+$, $N_{s+t} - N_s$ represents the number of events that occur in the time interval $(s, s + t]$. In the case of Poisson processes, this number of events follows a Poisson distribution (hence the name). One key property of a Poisson process is its Markovian nature: the distribution of the number of events occurred in the time interval $(s, s + t]$ does not depend on what happened up to time s . It can also be seen and more easily understood as a continuous-time version of a Bernoulli process.

Bernoulli process

A Bernoulli process of parameter $\lambda \in [0, 1]$ is simply a sequence $(X_n)_{n \in \mathbb{N}^*}$ of independent and identically distributed Bernoulli random variables with parameter λ . From a modeling perspective, this implies that at each discrete time step, an event may occur with a probability of λ . Consequently, we can deduce the following:

- (i) the waiting time until the next event is discrete and follows a geometric distribution with parameter λ ,
- (ii) let $m, n \in \mathbb{N}^*$ be discrete times, the number of events that occur between time m and time $m + n$ follows a binomial distribution $\text{Bin}(n, \lambda)$.

The Bernoulli process is obviously a time-homogeneous Markov process since it is a sequence of i.i.d. variables. Let $(N_t)_{t \geq 0}$ be the time-continuous counting process associated with the Bernoulli process, i.e. for any time $s \geq 0$:

$$N_s = \sum_{k=1}^{\lfloor s \rfloor} X_k \sim \text{Bin}(\lfloor s \rfloor, \lambda). \quad (2.13)$$

Let us now consider a Bernoulli process whose Bernoulli trials are spaced by a time increment of $h > 0$ (and not necessarily 1) but succeed with probability λh (instead of λ). When h is small,

the trials are $1/h$ times more frequent but fail h times more often, such that the frequency of successes remains the same as in the case $h = 1$. A Poisson process with intensity λ corresponds to the limit of this Bernoulli process as h tends to 0.

PROPOSITION 9 · Convergence of accelerated Bernoulli process

Let $(\tilde{X}_n^{(h)})_{n \in \mathbb{N}^*}$ be a sequence of i.i.d. random variables of Bernoulli distribution with parameter λh . Let also $(\tilde{N}_t^{(h)})_{t \geq 0}$ be the associated counting process knowing a Bernoulli trial is attempted at each time step h , i.e., for any time $t \geq 0$:

$$\tilde{N}_t^{(h)} = \sum_{k=1}^{\lfloor \frac{t}{h} \rfloor} \tilde{X}_k^{(h)}. \quad (2.14)$$

Then:

- (i) The number of events in the time interval $(s, s + t]$ is asymptotically distributed according to a Poisson distribution whose parameter is proportional to the elapsed time. For any $k \in \mathbb{N}$:

$$\lim_{h \rightarrow 0} \mathbb{P}(\tilde{N}_t^{(h)} = k) = \exp(-\lambda t) \frac{(\lambda t)^k}{k!}. \quad (2.15)$$

- (ii) The waiting time before the next event asymptotically follows an exponential distribution of parameter λ . For any time $s, t \geq 0$:

$$\lim_{h \rightarrow 0} \mathbb{P}(\tilde{N}_{s+t}^{(h)} = \tilde{N}_s^{(h)}) = \exp(-\lambda t). \quad (2.16)$$

Proof. First, we clearly have $\tilde{N}_t^{(h)} \sim \text{Bin}(\lfloor \frac{t}{h} \rfloor, \lambda h)$.

- (i) It should be noted that $\lim_{h \rightarrow 0} \lambda h \lfloor \frac{t}{h} \rfloor = \lambda t$ since $h \lfloor \frac{t}{h} \rfloor \in (t - h, t]$. Using $\lambda h \sim \frac{\lambda t}{\lfloor \frac{t}{h} \rfloor}$, we obtain:

$$\begin{aligned} \mathbb{P}(\tilde{N}_t^{(h)} = k) &= \binom{\lfloor \frac{t}{h} \rfloor}{k} (\lambda h)^k (1 - \lambda h)^{\lfloor \frac{t}{h} \rfloor - k} \underset{h \rightarrow 0}{\sim} \frac{\lfloor \frac{t}{h} \rfloor^k}{k!} \frac{(\lambda t)^k}{\lfloor \frac{t}{h} \rfloor^k} \left(1 - \frac{\lambda t}{\lfloor \frac{t}{h} \rfloor}\right)^{\lfloor \frac{t}{h} \rfloor - k}, \\ &\underset{h \rightarrow 0}{\sim} \frac{(\lambda t)^k}{k!} \exp\left[\left\lfloor \frac{t}{h} \right\rfloor \log\left(1 - \frac{\lambda t}{\lfloor \frac{t}{h} \rfloor}\right)\right] \xrightarrow{h \rightarrow 0} \frac{(\lambda t)^k}{k!} e^{-\lambda t}. \end{aligned} \quad (2.17)$$

(ii) The event $\{\widetilde{N}_{s+t}^{(h)} = \widetilde{N}_t^{(h)}\}$ corresponds to an absence of success between time s and time $s + t$.

$$\begin{aligned} \mathbb{P}(\widetilde{N}_{s+t}^{(h)} = \widetilde{N}_t^{(h)}) &= \mathbb{P}\left(\sum_{k=\lfloor \frac{s}{h} \rfloor + 1}^{\lfloor \frac{s+t}{h} \rfloor} \widetilde{X}_k^{(h)} = 0\right) = (1 - \lambda h)^{\lfloor \frac{s+t}{h} \rfloor - \lfloor \frac{s}{h} \rfloor - 1} \\ &= \exp\left[\left(\left\lfloor \frac{s+t}{h} \right\rfloor - \left\lfloor \frac{s}{h} \right\rfloor - 1\right) \log(1 - \lambda h)\right] \\ &\underset{h \rightarrow 0}{\sim} \exp[-\lambda \times (s+t) + \lambda s + \lambda h] \underset{h \rightarrow 0}{\rightarrow} e^{-\lambda t}. \end{aligned} \quad (2.18)$$

■

The emergence of the Poisson distribution as the probability of occurrence decreases is what earns it the nickname “the law of rare events”.

Construction of the Poisson process

Poisson processes are, in a sense, the limit version of Bernoulli processes: at any given instant, an event can occur independently of what came before, with a “rate” λ . There are several equivalent definitions of a Poisson process. The classical characterisation gives the expected properties of such a counting process.

DEFINITION 10 · Poisson process, characterization as a counting process

A Poisson process with intensity $\lambda > 0$ is defined as the counting process $(N_t)_{t \geq 0}$ (starting at $N_0 = 0$ a.s.) satisfying:

- (i) The number of events in non-overlapping intervals are independent, i.e. for any $0 \leq t_1 \leq t_2 \leq t_3 \leq t_4$, the random variable $N_{t_4} - N_{t_3}$ is independent of $N_{t_2} - N_{t_1}$.
- (ii) The probability of observing one event in a time interval is proportional to the length of this interval when it tends to 0 (with proportionality coefficient λ):

$$\mathbb{P}(N_{t+h} - N_t = 1) = \lambda h + o(h) \quad \text{for any } t \geq 0. \quad (2.19)$$

- (iii) The probability of observing strictly more than one event in a time interval is zero when its length tends to 0:

$$\mathbb{P}(N_{t+h} - N_t > 1) = o(h) \quad \text{for any } t \geq 0. \quad (2.20)$$

Here $o(h)$ is the Landau notation of any function such that $\frac{o(h)}{h} \xrightarrow{h \rightarrow 0^+} 0$.

The disadvantage of this definition is that it does not directly indicate how to construct a Poisson

process. The results of [proposition 9](#) suggest a way by focusing on the waiting times between events.

DEFINITION 11 · Poisson process, characterization as a jump process

Let $(T_n)_{n \in \mathbb{N}}$ be a sequence of independent and identically distributed random variables following an exponential distribution with parameter λ . We note

- ◊ $S_n = \sum_{k=0}^{n-1} T_k$ the sum of the n first times for any $n \in \mathbb{N}^*$.
- ◊ $N_t = \sum_{n=1}^{\infty} \mathbf{1}_{S_n \leq t} = \sup\{n \in \mathbb{N} : S_n \leq t\}$ the number of events occurring before time t for any $t \geq 0$.

The stochastic process $(N_t)_{t \geq 0}$ is referred to as a Poisson process of intensity λ .

THEOREM 1 · Equivalence between Poisson process characterizations

[Definition 11](#) and [definition 10](#) are equivalent. More precisely:

- (i) Let $(N_t)_{t \in \mathbb{R}_+}$ be a counting process with values in \mathbb{N} and satisfying the three properties of [definition 10](#). For any $n \in \mathbb{N}$, we define the random variable:

$$S_n = \inf\{t \geq 0 : N_t \geq n\}. \quad (2.21)$$

Then the variables $T_n = S_{n+1} - S_n$ are mutually independent for any $n \in \mathbb{N}$ and follow an exponential distribution with parameter λ .

- (ii) Let $(T_n)_{n \in \mathbb{N}}$ be a sequence of i.i.d. random variables following an exponential distribution with parameter λ . The process $(N_t)_t$ defined for any $t \geq 0$ by:

$$N_t = \sup\left\{n : \sum_{k=0}^{n-1} T_k \leq t\right\}, \quad (2.22)$$

satisfies the three properties of [definition 10](#).

Moreover, in both cases, the number of events occurred in any time interval of length t follows a Poisson distribution with parameter λt . For any times $s, t > 0$ and $k \in \mathbb{N}$, we have:

$$\mathbb{P}(N_{s+t} - N_s = k) = \exp(-\lambda t) \frac{(\lambda t)^k}{k!}. \quad (2.23)$$

Proof. (i) This result is known as the *interval theorem*. A proof is given in [[Kin92](#), Section 4.1 pp. 39].

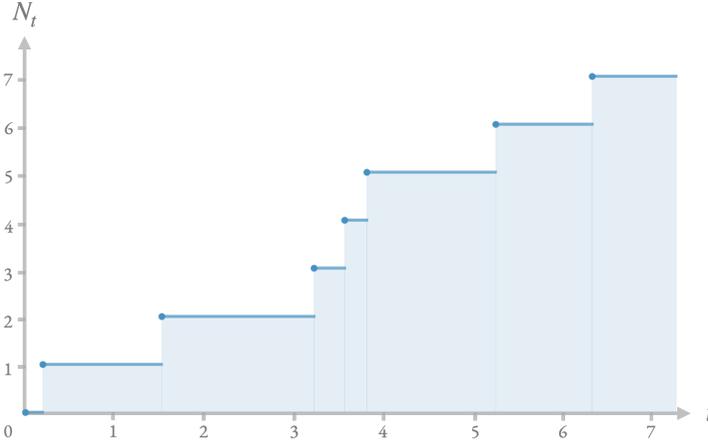


FIGURE 2.1 **Representation of a simple Poisson trajectory** with intensity $\lambda = 1$.

(ii) The process $(N_t)_t$ has clearly independent increments. Let T_1 and T_2 be two independent random variables with exponential distribution with parameter λ .

$$\begin{aligned} \mathbb{P}(N_{t+h} - N_t = 1) &= \mathbb{P}(T_1 \leq h, T_1 + T_2 > h) = \mathbb{E}\left[\mathbf{1}_{T_1 \leq h} \mathbb{P}(T_2 > h - T_1 \mid T_1)\right] \\ &= \int_0^h e^{-\lambda(h-t)} \lambda e^{-\lambda t} dt = \lambda h e^{-\lambda h}. \end{aligned} \quad (2.24)$$

Thus:

$$\frac{1}{\lambda h} \mathbb{P}(N_{t+h} - N_t = 1) \xrightarrow{h \rightarrow 0^+} 1. \quad (2.25)$$

Similarly:

$$\begin{aligned} \mathbb{P}(N_{t+h} - N_t > 1) &= \mathbb{P}(T_1 + T_2 < h) = \mathbb{E}\left[\mathbf{1}_{T_1 \leq h} \mathbb{P}(T_2 < h - T_1 \mid T_1)\right] \\ &= \mathbb{E}\left[\mathbf{1}_{T_1 \leq h} (1 - e^{-\lambda(h-T_1)})\right] = 1 - e^{-\lambda h} - \lambda h e^{-\lambda h}. \end{aligned} \quad (2.26)$$

Thus:

$$\frac{1}{h} \mathbb{P}(N_{t+h} - N_t \leq 1) = (1 - \lambda h) \frac{e^{-\lambda h}}{h} \xrightarrow{h \rightarrow 0^+} 1. \quad (2.27)$$

■

A trajectory of a Poisson process with intensity λ can therefore be generated by drawing the inter-jump waiting times independently according to an exponential distribution of parameter λ . A realization of such a trajectory is shown in [fig. 2.1](#).

2.1.3 Inhomogeneous Poisson process

A typical application of Poisson processes is the modeling of queues: individuals arrive one by one at random times and join a queue. In [example 5](#), we have a non-Markovian queue: the waiting time before the next arrival in the queue depends on the time elapsed since the last arrival. On the other side, if a queue is described by a simple Poisson process, then:

- (i) *Markov process*: at any time, the waiting time before the next arrival does not depend on the number or the dates of the previous arrivals,
- (ii) *Time-homogeneous process*: the probability of observing $k \in \mathbb{N}$ arrivals in a given time slot is the same for any other slot of the same duration.

If we are interested in the queue of a store, it is quite realistic to assume that consumers do not coordinate and arrive independently of each other (property (i)). On the other hand, property (ii) is probably too restrictive. Some time slots are busier than others. The Poisson process can be generalized to the case of a non-constant intensity λ but a function of the total elapsed time.

Construction of the inhomogeneous Poisson process

The intensity $\lambda : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is from now on a non-negative function of time. The greater its value at a time t , the more likely an event is to occur at that time. It is perhaps more natural to generalize the [definition 10](#) in that case. The corresponding process is called an inhomogeneous Poisson process or a non-stationary Poisson process.

DEFINITION 12 · Inhomogeneous Poisson process (counting process characterization)

A counting process $(N_t)_{t \in \mathbb{R}_+}$ (starting at $N_0 = 0$ a.s.) is an inhomogeneous Poisson process of intensity function λ if and only if:

- (i) The increments are independent (same condition as the first of [definition 10](#)).
- (ii) The probability of observing one event in a time interval is proportional to the length of this interval when it tends to 0 (but the proportionality coefficient λ changes with time):

$$\mathbb{P}(N_{t+h} - N_t = 1) = \lambda(t)h + o(h) \quad \text{for any } t \geq 0. \quad (2.28)$$

- (iii) The probability of observing strictly more than one event in a time interval is zero when its length tends to 0:

$$\mathbb{P}(N_{t+h} - N_t > 1) = o(h) \quad \text{for any } t \geq 0. \quad (2.29)$$

As before, this definition does not explicit a construction method for the process. There also exists a generalization of the second characterization ([definition 11](#)). Let us note $\Lambda : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ the integrated intensity defined as:

$$\Lambda : s \mapsto \int_0^s \lambda(u) \, du. \quad (2.30)$$

DEFINITION 13 · Poisson process, characterization as a jump process

Let $(T_n)_{n \in \mathbb{N}}$ be a sequence of random variables in \mathbb{R}_+ . For any $n \in \mathbb{N}$ and any $t \in \mathbb{R}_+$ we note $S_n = \sum_{k=0}^{n-1} T_k$ and $N_t = \sum_{n=1}^{\infty} \mathbb{1}_{S_n \leq t}$. Assume that for any $n \in \mathbb{N}^*$, the survival function of T_n given S_n is:

$$s \mapsto \mathbb{P}(T_n > s \mid S_n) = \exp[-(\Lambda(s + S_n) - \Lambda(S_n))]. \quad (2.31)$$

Equivalently we have:

$$\mathbb{P}(S_{n+1} > s \mid S_n) = \exp[-(\Lambda(s) - \Lambda(S_n))] \mathbb{1}_{s \geq S_n}. \quad (2.32)$$

The stochastic process $(N_t)_{t \geq 0}$ is referred to as an inhomogeneous Poisson process of intensity λ .

The successive times of an inhomogeneous Poisson process can be interpreted as exponentially distributed draws whose parameter fluctuates continuously over time. Once again, the two definitions are equivalent and the number of points in any interval follows a Poisson distribution.

THEOREM 2 · Equivalent definitions of inhomogeneous Poisson processes

[Definition 13](#) and [definition 12](#) are equivalent. More precisely:

(i) Let $(N_t)_{t \in \mathbb{R}_+}$ be a counting process with values in \mathbb{N} satisfying the four properties of [definition 12](#). Then, for any $n \in \mathbb{N}$, [eq. \(2.32\)](#) is the survival function of the random variable

$$S_n = \inf\{t \geq 0 : N_t \geq n\}. \quad (2.33)$$

(ii) Let $(T_n)_{n \in \mathbb{N}}$ be sequentially defined as random variables with survival function [eq. \(2.31\)](#). The process $(N_t)_t$ defined for any $t \geq 0$ by:

$$N_t = \sup\left\{n : \sum_{k=0}^{n-1} T_k \leq t\right\}, \quad (2.34)$$

satisfies the four properties of [definition 12](#).

Moreover, in both cases, the number of events occurred in the time interval $(s, s + t]$ for any time $s, t \geq 0$ follows a Poisson distribution with parameter $\Lambda(s + t) - \Lambda(s)$. Thus, for any times $s, t > 0$ and $k \in \mathbb{N}$, we have:

$$\mathbb{P}(N_{s+t} - N_s = k) = \exp[-(\Lambda(s + t) - \Lambda(s))] \frac{(\Lambda(s + t) - \Lambda(s))^k}{k!}. \quad (2.35)$$

The proof of (i) is given in [[Tij03](#), Theorem 1.3.1]². The proof of (ii) is similar to the proof of [theorem 1](#).

Connection with the Poisson binomial distribution

Remember that the simple Poisson process could be seen as a passage to the limit of the counting process associated with a Bernoulli process. The inhomogeneous Poisson process corresponds to the limit of a sum of independent Bernoulli (when the step tends to 0) but with a different parameter λ_k at each discrete time step. The distribution of the sum is actually known as the Poisson binomial distribution.

DEFINITION 14 · Poisson binomial distribution

A Poisson binomial variable K with parameters $n \in \mathbb{N}^*$ and $(\lambda_k)_{k=1}^n \in [0, 1]^n$ is distributed as the sum of n independent Bernoulli variables, the k -th having parameter λ_k for $k = 1, \dots, n$.

$$p_n(k) = \mathbb{P}(K = k) = \sum_{A \in F_k} \prod_{i \in A} \lambda_i \prod_{j \in A^c} (1 - \lambda_j), \quad (2.36)$$

where $F_k = \{A \subseteq \llbracket 1, n \rrbracket : \text{Card}(A) = k\}$.

2.2 PDMP essentials

Piecewise deterministic Markov processes (PDMPs) were introduced in 1984 by Mark H. A. Davis, originally to provide a framework for the rigorous study of problems arising in queuing systems, inventory theory, resource allocation, in his words. Since their formalization (initially in the article [[Dav84](#)] then detailed in Davis's reference book on Markovian processes [[Dav18](#)]) PDMPs have proved to be a very rich probabilistic modeling tool and have found applications in a great variety of scientific fields. These include industrial reliability [[DDZ15](#); [Des+21](#); [Ben+16](#)], insurances [[Sch98](#); [Kri+19](#); [ES94](#)], and in particular many applications in biology [[RT17](#); [Kre23](#);

²More precisely, the author proves that the process $(N_t)_t$ of [definition 12](#) is Poisson-distributed. The survival function [eq. \(2.31\)](#) can then be deduced from [eq. \(2.35\)](#) with $k = 0$.

[Clo+17], for example to model the evolution of a population [ASS18; GHL22], the electrical activity of a neuron [LTT18; Tho19] or model gene regulatory networks [LB18; Her+17]. The use of PDMPs is not limited to probabilistic modeling. They have gained popularity in the statistical community over the past decade, forming the foundation of the so-called PD-MCMC methods [Van+17]. As variants of MCMC (Markov Chain Monte Carlo) methods, PDMPs are instrumental processes employed in place of traditional Markov chains to simulate an intractable distribution. Notable examples include the Bouncy Particle Sampler [BVD18] and the Zig-Zag sampler [BFR19]. These methods provide motivation for further research on the ergodic properties of PDMPs [DGM21; CDo8; Ben+15]. These two recent PhD theses on PD-MCMC methods [Ber23; Gra23] provide a good overview of the state of the art in the field³.

Hybrid process

This variety of application is not surprising, as PDMPs form a very broad class of stochastic processes, in fact those of all non-diffusive Markovian processes. However their expressive power comes with the price of an intricate formalism. These are continuous-time stochastic jump processes with values in a space known as hybrid, as it is the product of a Euclidean subspace and a finite or countable set. Thus, the state variable of the process consists of a continuous part, called position, and a discrete part, called regime.

A PDMP is a generalization of a continuous-time Markov chain (CTMC) on this hybrid space. It differs in two aspects. Firstly, the state variable is not fixed between jumps. The regime remains constant between two jumps, and the position of the process evolves continuously along a deterministic trajectory dependent on the regime. Secondly, the process has two different jump mechanisms: a CTMC-style random one, and a deterministic one. It is subject to spontaneous jumps at random times drawn from an inhomogeneous Poisson process (with intensity depending on the state variable and not just time). It also jumps at “deterministic” times when the position hits the boundary of the state space (if it exists). The destination of the jumps is drawn from a transition kernel on \mathcal{X} similarly to the CTMC case. After the jump, the position follows a new deterministic trajectory depending on the new regime.

2.2.1 Local characteristics of the PDMP

A PDMP is a stochastic process with values in a hybrid space $\mathcal{X} = \mathcal{Z} \times \mathcal{V}$ with $\mathcal{Z} \subset \mathbb{R}^{d_z}$ and \mathcal{V} a finite or countable set. Its state over time $X_t = (Z_t, V_t) \in \mathcal{X}$ is described by both the continuous variable $Z_t \in \mathcal{Z}$ called position, and by the discrete variable $V_t \in \mathcal{V}$ called regime. The complete behavior of the PDMP is characterized by three functions.

³I also recommend this webpage maintained by Joris Bierkens: <https://diamhomes.ewi.tudelft.nl/~jorisbierkens/pdmps.html>.

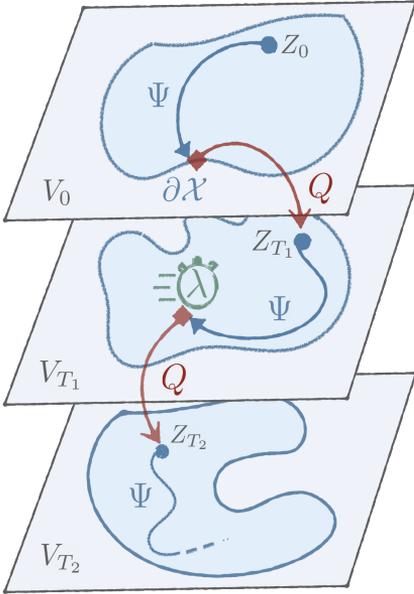


FIGURE 2.2 **Representation of a PDMP trajectory** with local characteristics (Ψ, λ, Q) . The position follows the flow Ψ between the jumps starting from an initial state $X_0 = (Z_0, V_0)$. The process first jumps when it hits the boundary of the state space at time T_1 to state (Z_{T_1}, V_{T_1}) drawn with jump kernel Q . It then jumps at a random time T_2 drawn with jump intensity λ to state (Z_{T_2}, V_{T_2}) drawn with jump kernel Q .

- (i) The *flow* Ψ gives the deterministic trajectory of the position between the jumps. This is generally the solution of a system of differential equations.
- (ii) The *jump intensity* λ characterizes the distribution of the time of the spontaneous jumps. This is a function that associates a non-negative real value with each state of the space. The higher the value, the more likely the process is to jump when it is in that state.
- (iii) The *transition kernel* Q determines the distribution of the state of the process after a jump. The distribution of the destination of a jump only depends on the state of the process right before the jump.

The triplet (Ψ, λ, Q) is called the *local characteristics* of the PDMP on \mathcal{X} .

Piecewise deterministic motion

Between two jumps, the regime of the process remains constant while the position follows a deterministic motion that depends on the regime. This deterministic motion is given by a flow function $\Psi : \mathcal{X} \times \mathbb{R}_+ \mapsto \mathcal{X}$ in the following way: for any $s, t \geq 0$ and state $X_s \in \mathcal{X}$, if there is no jump between the time s and the time $s + t$, then $X_{s+t} = \Psi(X_s, t)$. A flow function verifies the following properties:

- (i) for any $x \in \mathcal{X}$, $\Psi(x, 0) = x$,
- (ii) and the flow has the semigroup property $\Psi(\Psi(x, t), s) = \Psi(x, s + t)$.

Notice that, since only the position moves between the jumps, we can also define, in a given regime $v \in \mathcal{V}$, a flow function ψ_v on $\mathcal{Z} \times \mathbb{R}_+$ associated with the position only. For any $x = (z, v) \in \mathcal{Z} \times \mathcal{V}$ and any $t \geq 0$, we have :

$$\Psi((z, v), t) = (\psi_v(z, t), v). \quad (2.37)$$

In the most common case, the position of the PDMP evolves according to one or more ordinary differential equations (ODEs), from which the flow is deduced. Let $x = (z, v) \in \mathcal{Z} \times \mathcal{V}$ and let $\Xi_v : \mathcal{Z} \rightarrow \mathcal{Z}$ be a Lipschitz continuous function of the position. Then by Picard-Lindelöf theorem, there exists a unique solution function ψ_v to the ODE problem:

$$\begin{cases} \frac{d}{dt} \psi_v(z, t) = \Xi_v(\psi_v(z, t)) \\ \psi_v(z, 0) = z \end{cases} \quad (2.38)$$

The case of time-dependent ODEs is handled by considering the time-augmented process (see [definition 6](#)) that includes the elapsed time in the position variable. We also define \mathfrak{X} , the differential operator of the first order.

DEFINITION 15 · Differential operator of the first order

Let $f : \mathcal{X} \rightarrow \mathbb{R}$ be such that for any $v \in \mathcal{V}$, the application $z \in \mathcal{Z} \mapsto f((z, v))$ is C^1 (continuously differentiable), and let $x = (z, v) \in \mathcal{X}$. We have:

$$\mathfrak{X}f(x) := \langle \nabla_z f(x), \Xi_v(z) \rangle = \sum_{i=1}^{d_z} \frac{\partial f}{\partial z_i}((z, v)) \times \left[\frac{d}{dt} \psi_v(z, t) \right]_i. \quad (2.39)$$

Note that \mathfrak{X} satisfies (for the same f and x):

$$\frac{d}{dt} f(\Psi(x, t)) = \mathfrak{X}f(\Psi(x, t)). \quad (2.40)$$

This operator is mainly involved in the expression of the infinitesimal generator of the PDMP (see [proposition 12](#)).

Jumps at the boundary

A PDMP trajectory jumps when the position reaches the state space boundary (if it exists) following the flow. Let $\mathcal{X}_v := \{(z', v') \in \mathcal{X} \mid z' \in \mathcal{Z}, v' = v\}$ be the subspace of states that can only be reached in the regime $v \in \mathcal{V}$. We have $\mathcal{X} = \bigcup_{v \in \mathcal{V}} \mathcal{X}_v$. If \mathcal{X}_v has boundaries noted $\partial \mathcal{X}_v$, the process jumps when the position reaches them. For any $x = (z, v) \in \mathcal{X}_v$, the waiting time before

this jump is thus defined by:

$$t_\theta(x) = \inf \{s > 0 : \Psi(x, s) \in \partial\mathcal{X}_v\} \quad \text{with} \quad \inf\{\emptyset\} = +\infty.$$

Boundary jumps are often referred to as deterministic jumps, because they occur on a fixed date. This can be misleading, as the destination of the jumps remains random, and the date of the jumps is “deterministic” provided that a spontaneous jump does not occur beforehand. Similarly, we prefer to refer to jumps occurring at random times as spontaneous jumps rather than random jumps.

Jump intensity function

As we said, the spontaneous jumps occur at random times. From any initial state $x \in \mathcal{X}$, the waiting time before the next spontaneous jump is drawn according to an inhomogeneous Poisson process of intensity $\lambda \circ \Psi(x, \cdot)$ where $\lambda : \mathcal{X} \rightarrow \mathbb{R}_+$ is the jump intensity of the PDMP. It associates to each state $x \in \mathcal{X}$ a weight $\lambda(x) \in [0, +\infty)$. The larger $\lambda(x)$ is, the more likely it is that the PDMP jumps when it passes through state x . To simplify notation, we define the two following functions from λ for any $x \in \mathcal{X}$:

$$\lambda_x : u \in \mathbb{R}_+ \mapsto \lambda(\Psi(x, u)), \quad (2.41)$$

$$\Lambda : (x, s) \in \mathcal{X} \times \mathbb{R}_+ \mapsto \int_0^s \lambda_x(u) du. \quad (2.42)$$

It is assumed that λ_x is locally integrable. From state $x \in \mathcal{X}$, let $T_\lambda^x(x)$ be the waiting time drawn according to the inhomogeneous Poisson process before the next spontaneous jump⁴. Its survival function is given by:

$$\mathbb{P}(T_\lambda^x(x) > s) = \exp(-\Lambda(x, s)). \quad (2.43)$$

The waiting time $T_\lambda^x(x)$ admits a probability density function with respect to the Lebesgue measure that can be deduced from its cumulative distribution function:

$$s \mapsto \frac{d}{ds} \mathbb{P}(T_\lambda^x(x) \leq s) = \lambda_x(s) \exp(-\Lambda(x, s)). \quad (2.44)$$

⁴Throughout this document, we adopt a conventional notation. Uppercase variables generally correspond to random variables, and lowercase variables to observed realizations. The boundary hitting time t_θ is therefore always in lowercase. The waiting time before a spontaneous jump will be written in uppercase T_λ^x or lowercase t_λ^x depending on whether we consider the random variable or its realization.

True jump time

Ultimately, starting from a state $x \in \mathcal{X}$, the waiting time $T(x)$ before the next jump is the shortest time between the (deterministic) waiting time to reach the boundary of the state space, and the (random) waiting time drawn according to the inhomogeneous Poisson process. Thus:

$$T(x) = \min(t_\partial(x), T_\lambda(x)). \quad (2.45)$$

Its survival function is given by:

$$\mathbb{P}(T(x) > s) = \mathbb{1}_{s \leq t_\partial(x)} \exp(-\Lambda(x, s)). \quad (2.46)$$

It also admits the probability density function:

$$s \mapsto p_\lambda(s | x) = p_{\lambda_x}(s) = \lambda_x(s) \mathbb{1}_{s < t_\partial(x)} \exp(-\Lambda(x, s)) \mathbb{1}_{s \leq t_\partial(x)}, \quad (2.47)$$

with respect to the measure:

$$\nu_\lambda(\cdot | x) : B \in \mathcal{B}([0, t_\partial(x)]) \mapsto \text{Leb}(B \cap [0, t_\partial(x))) + \delta_{t_\partial(x)}(B), \quad (2.48)$$

where Leb is the Lebesgue measure⁵.

The state space \mathcal{X} as a metric space

The jumps of a PDMP can simultaneously modify its position and regime. Thus, the state of the process after a jump is randomly drawn in \mathcal{X} . To define a probability measure on \mathcal{X} , we need to define a Borelian σ -algebra on \mathcal{X} and therefore a metric on this space. To our knowledge, there is no natural distance on a hybrid space. However, such a distance can be constructed from a distance $\text{dist}_{\mathcal{Z}}$ on \mathcal{Z} . Let $C_{\mathcal{Z}}$ be a constant upperbound of $\text{dist}_{\mathcal{Z}}$, i.e. $C_{\mathcal{Z}} \geq \sup_{(z, z') \in \mathcal{Z}^2} \text{dist}_{\mathcal{Z}}(z, z')$ and $C_{\mathcal{Z}} = \infty$ if \mathcal{Z} is unbounded. For any $x_1 = (z_1, v_1), x_2 = (z_2, v_2) \in \mathcal{X}$, we define the distance $\text{dist}_{\mathcal{X}}$ on \mathcal{X} by:

$$\text{dist}_{\mathcal{X}}((z_1, v_1), (z_2, v_2)) = \begin{cases} \text{dist}_{\mathcal{Z}}(z_1, z_2) & \text{if } v_1 = v_2, \\ \frac{1}{2} C_{\mathcal{Z}} & \text{else.} \end{cases} \quad (2.49)$$

Choosing a constant $C_{\mathcal{Z}} < \sup_{(z, z') \in \mathcal{Z}^2} \text{dist}_{\mathcal{Z}}(z, z')$ would not ensure that $\text{dist}_{\mathcal{X}}$ verifies the triangle inequality. Indeed, let $x_1 = (z_1, v_1), x_2 = (z_2, v_2), x_3 = (z_3, v_3) \in \mathcal{X}$ be three states such that $v_1 = v_3 \neq v_2$ and $\text{dist}_{\mathcal{Z}}(z_1, z_3) > C_{\mathcal{Z}}$. We have:

$$\text{dist}_{\mathcal{X}}((z_1, v_1), (z_2, v_2)) + \text{dist}_{\mathcal{X}}((z_2, v_2), (z_3, v_3)) = C_{\mathcal{Z}} < \text{dist}_{\mathcal{X}}((z_1, v_1), (z_3, v_3)). \quad (2.50)$$

⁵Restricting the Lebesgue measure to the interval $[0, t_\partial(x))$ makes p_λ a p.d.f. without having to renormalize it in eq. (2.47).

Equipping the state space with a metric allows us to define the desired Borelian σ -algebra on \mathcal{X} noted $\mathcal{B}(\mathcal{X})$ (where $\mathcal{B}(\cdot)$ more generally indicates the Borel sets of a set).

Jump transition kernel

After a jump, the process resumes a new deterministic trajectory from a new state randomly chosen according to a Markovian transition kernel Q , i.e. a family $[Q(\cdot | x)]_{x \in \mathcal{X}}$ of probability distributions on \mathcal{X} indexed by each element of \mathcal{X} . If the process is in state $x \in \mathcal{X}$ at time $s \geq 0$ and the waiting time before the next jump is $t \geq 0$, then its probability to land on a subset $B \subset \mathcal{X}$ after its next jump is given by:

$$\mathbb{P}(X_{s+t} \in B | X_s = x, T(X_s) = t) = Q(B | \Psi(x, t)). \quad (2.51)$$

Let also $q = [q(\cdot | x)]_{x \in \mathcal{X}}$ be a family of probability density functions for Q with respect to a family of reference measures $\nu_q = [\nu_q(\cdot | x)]_{x \in \mathcal{X}}$ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Jumping from the state $x^- \in \mathcal{X}$, the probability to land on a subset $B \subset \mathcal{X}$ is thus given by:

$$Q(B | x^-) = \int_B q(x | x^-) \nu_q(dx | x^-). \quad (2.52)$$

A jump necessarily changes the state of the PDMP (no jump in place). The jump kernel therefore verifies $Q(\{x\} | x) = q(x | x) = 0$ for any $x \in \mathcal{X}$. We also introduce the following notation. Let f be a function of \mathcal{X} , for any $x^- \in \mathcal{X}$ we note $Qf(x^-)$ the mean value of f knowing the process jumps from state x^- :

$$Qf(x^-) = \mathbb{E}_{X \sim Q(\cdot | x^-)}[f(X)] = \int_{\mathcal{X}} f(x) q(x | x^-) \nu_q(dx | x^-). \quad (2.53)$$

In practice, PDMPs often model phenomena for which the number of possible transitions from any state $x^- \in \mathcal{X}$ is finite. The measure $\nu_q(\cdot | x^-)$ is then simply a counting measure, and $q(\cdot | x^-)$ can be represented by a vector.

2.2.2 Strong Markov PDP

It is clear that the process we have described so far is at least a piecewise deterministic process (PDP), but is it Markovian and even homogeneous? It is easy to have a local intuition. The randomness only concerns the time and the destination of the jumps. Firstly, the distribution of the next jump time given the entire past of the PDP only depends on its current state (see eq. (2.46)). Secondly, the distribution of the destination of a jump, given the entire past of the process before the jump and given the time of the jump, only depends on the state preceding the jump (see eq. (2.52)). To prove this globally, an additional assumption is needed: the non-

explosion of the process.

Explosion

A situation one wants to avoid is the explosion of the process. This is an acceleration of the jump dynamics such that an infinite number of jumps can occur over a finite time span. Let $(X_t)_t$ be a PDP on \mathcal{X} with local characteristics (Ψ, λ, Q) . Let us denote by S_k the time of the k -th jump and T_k the waiting time between the k -th and the $(k + 1)$ -th jump. More formally, starting at time $S_0 = 0$, we have:

$$\begin{aligned} T_k &= T(X_{S_k}) = \inf\{t \geq 0 : \Psi(X_{S_k}, t) \neq X_{S_k+t}\}, \\ S_{k+1} &= S_k + T_k. \end{aligned} \quad (2.54)$$

The first line can be interpreted as follows: a jump has occurred when the state of the PDMP is no longer the one that would have been obtained by following the flow since the previous jump.

DEFINITION 16 · Non-explosion

A jump process with jump times $(S_k)_{k \in \mathbb{N}^*}$ and interjump times $(T_k)_{k \in \mathbb{N}}$ is said non-explosive if:

$$\sup_{n \in \mathbb{N}^*} S_n = \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} T_k = +\infty \quad \text{a.s.} \quad (2.55)$$

Conversely, if $\lim_{n \rightarrow \infty} S_n < s$ for some $s > 0$, we have no proper definition of X_{s+t} , the state of the process at time $s + t$ for any $t \geq 0$. It is therefore very inconvenient to study the process in long (or even short) time without assuming that it is non-explosive.

PROPOSITION 10 · Sufficient conditions for non-explosion

Let $(X_t)_t$ be a PDP with values in \mathcal{X} and local characteristics (Ψ, λ, Q) . The non-explosivity of the process (definition 16) holds under the two following assumptions:

- (i) For any $x \in \mathcal{X}$ the jump intensity (as a function of time) λ_x is upper bounded on $[0, t_\theta(x)]$.
- (ii) Let $A_\varepsilon = \{x' \in \mathcal{X} \mid t_\theta(x') \geq \varepsilon\}$. There exists $\varepsilon > 0$ such that $Q(A_\varepsilon \mid x) = 1$ for any $x = (z, v) \in \partial\mathcal{X}_v$.

The proof is given in [Dav18, Proposition 2.4.6] under a slightly stronger second assumption: there exists $\varepsilon > 0$ such that $Q(A_\varepsilon \mid x) = 1$ for any $x \in \mathcal{X}$ and not only for any x in the boundary, but the proof remains the same. All we need is to ensure that a jump at boundary cannot occur an arbitrarily short time after another. This is easily illustrated by the following example:

EXAMPLE 6 · Explosive process

Let $(X_t)_t$ be the following PDMP:

- ◊ the state space is $\mathcal{X} = \mathcal{Z} \times \mathcal{V} = [0, 1] \times \mathbb{N}$,
- ◊ the flow function is $\Psi : (x, t) = ((z, v), t) \in \mathcal{X} \times \mathbb{R}_+ \mapsto z - t$,
- ◊ the jump intensity is constant and equal to zero: $\lambda(x) = 0$ for any $x \in \mathcal{X}$,
- ◊ the jump kernel is a Dirac distribution, for any $B \in \mathcal{X}$ and any departure state $x^- = (z^-, v^-)$:

$$Q(B \mid (z^-, v^-)) = \delta_{(v^--1)^-, v^--1}(B).$$

Jumps only occur at the boundary when the position returns linearly to 0. The post-jump position is closer and closer to 0 as jumps are made. Starting from $x_0 = (1, 0)$, the k -th waiting time before a jump is $T_{k-1} = \frac{1}{k^2}$. Thus $\sum_{k=1}^{\infty} T_{k-1} = \frac{\pi^2}{6} < \infty$ and we have explosion.

Generator of a PDMP

Let $(X_t)_t$ be a PDP with values in \mathcal{X} and local characteristics (Ψ, λ, Q) . Let $\{P_t\}_t$ be the family of operators defined, for any time $s > 0$, any state $x \in \mathcal{X}$, and any bounded measurable functions $f : \mathcal{X} \rightarrow \mathbb{R}$, by:

$$P_s f(x) := \mathbb{E}[f(X_s) \mid X_0 = x]. \quad (2.56)$$

The following proposition from [Dav18, Theorem 25.5] assures us that the process is Markovian with transition kernels $\{P_t\}_t$.

PROPOSITION 11 · Piecewise deterministic strong Markov process ([Dav18])

Let $(X_t)_t$ be a non-explosive PDP with values in \mathcal{X} and local characteristics (Ψ, λ, Q) .

Then, the process $(X_t)_{t \in \mathbb{R}_+}$ is a homogeneous strong Markov process, i.e. for any a.s. finite stopping time τ :

$$\mathbb{E}[f(X_{\tau+s}) \mid X_{[0, \tau]}] = \mathbb{E}[f(X_{\tau+s}) \mid X_{\tau}] = P_s f(X_{\tau}). \quad (2.57)$$

The family of transition kernels $\{P_t\}_t$ completely characterizes the distribution of the PDMP. They are the reference tool for studying Markov processes in simple cases, such as discrete time or countable state space. Their explicit expression and manipulation are much less straightforward in continuous time and on general state spaces. Indeed, explicitly knowing P_t for all $t \geq 0$ means knowing the distribution of all future states of the PDMP given its current state. It is simpler to

characterize the local behavior of the process (as allowed by the local characteristics (Ψ, λ, Q)) by studying the distribution of the “closest” future state. By differentiating the transition kernel with respect to time, we obtain the *infinitesimal generator* of the process.

DEFINITION 17 · Infinitesimal generator of a Markov process

The infinitesimal generator \mathcal{A} is the operator defined, for any state $x \in \mathcal{X}$, by:

$$\mathcal{A}f(x) = \lim_{t \rightarrow 0^+} \frac{P_t f(x) - f(x)}{t} = \lim_{t \rightarrow 0^+} \frac{\mathbb{E}[f(X_t) \mid X_0 = x] - f(x)}{t}, \quad (2.58)$$

for any function f in the so-called *domain of the generator* $\mathcal{D}(\mathcal{A})$. A function f belongs to the domain $\mathcal{D}(\mathcal{A})$ if:

- (i) f is continuous on \mathcal{X} and vanishes at infinity in the following sense: for any $\varepsilon > 0$, there is a compact subset $C \subset \mathcal{X}$ such that $|f(x)| < \varepsilon$ for any $x \in \mathcal{X} \setminus C$,
- (ii) the limit in eq. (2.58) exists with respect to the supremum distance:

$$\sup_{x \in \mathcal{X}} |t \times \mathcal{A}f(x) - (P_t f(x) - f(x))| \xrightarrow{t \rightarrow 0^+} 0. \quad (2.59)$$

In [Dav18, Theorem 26.14], Davis gave the explicit expression of the infinitesimal generator of a PDMP on an extended domain $\overline{\mathcal{D}}(\mathcal{A})$. This extended domain is the set of measurable functions satisfying the Dynkin formula (discussed in more detail in theorem 13), i.e. such that the process $(M_t)_t$, defined below for any $s > 0$, is a local martingale:

$$M_s = f(X_s) - f(x) - \int_0^s \mathcal{A}f(X_t) dt. \quad (2.60)$$

PROPOSITION 12 · Infinitesimal generator of a PDMP ([Dav18])

The infinitesimal generator of a PDMP with local characteristics (Ψ, λ, Q) on \mathcal{X} is given, for any state $x \in \mathcal{X}$ and any function $f \in \overline{\mathcal{D}}(\mathcal{A})$, by:

$$\mathcal{A}f(x) = \mathfrak{K}f(x) + \lambda(x) (Qf(x) - f(x)). \quad (2.61)$$

The characterization of the distribution of a PDMP via the generator allows to define an invariant measure for the PDMP, which paves the way for PD-MCMC methods [Van+17]. A measure μ is invariant for the PDMP with generator \mathcal{A} if and only if any function $f \in \overline{\mathcal{D}}(\mathcal{A})$ (see [Dav18, Proposition 34.7]):

$$\int \mathcal{A}f(x) \mu(dx) = 0. \quad (2.62)$$

2.3 Modeling and simulation with PDMPs

We need to generate PDMP trajectories in practice. This implies being able to:

- ◇ *Follow the flow Ψ between the jumps.* Most often, Ψ is unknown. We only have access to a system of ODEs of the form [eq. \(2.38\)](#). When the solution is not explicit, it must be approximated numerically using classical deterministic methods. This introduces approximation errors, but we assume they are negligible (see [\[Kri+19; BBD22\]](#) for the study of PDMPs approximated by discretization and [\[BM24\]](#) for an extension to piecewise diffusion Markov processes). The main issue is the computational cost of this approximation, especially for complex ODEs encountered in the modeling of large industrial systems. We do not focus here on the method for solving the flow, but consider it to be the main computational cost for simulating a PDMP trajectory.
- ◇ *Determining the time of the next jump starting from an initial state $x \in \mathcal{X}$.* On the one hand, it is necessary to determine whether it is a boundary jump or a spontaneous jump. In the first case, everything relies on solving the flow Ψ (see previous point). In the second case, a time has to be drawn according to an inhomogeneous Poisson process with intensity $t \mapsto \lambda_x(t) = \lambda(\Psi(x, t))$. The most frequent models using PDMPs do not provide direct access to the global intensity λ , but rather to marginal intensities $(\lambda^{(j)})_j$ associated with specific transitions. The global intensity is then the sum of the marginal intensities, this is the principle of superposition of Poisson processes (see [section 2.3.1](#)). This principle and the resolution of the flow Ψ give us access to λ_x . We then propose two methods for drawing a jump time with intensity λ_x : the *thinning* based on a rejection method specific to Poisson processes (see [section 2.3.2](#)) and the true jump method based on the numerical inversion of the c.d.f. of the jump time (see [section 2.3.3](#)).
- ◇ *Finally, determining the arrival state of the process after a jump from a departure state $x^- \in \mathcal{X}$.* This state is drawn according to a jump kernel Q that is rarely known explicitly. Indeed, as mentioned in the previous point, the most common models using PDMPs do not provide the local characteristics directly, but rather the different possible types of transitions and the associated probabilities or rates. Knowing Q is not difficult from a computational point of view, but it is very tedious (or impossible) to describe it exhaustively. Most often, we have access to $Q(\cdot | x^-)$ for a departure state $x^- = (z^-, v^-)$ belonging to the boundary $\partial\mathcal{X}_v^-$. In the case of spontaneous jumps, the kernel Q is a function of the marginal intensities $(\lambda^{(j)})_j$ described in the previous point according to the Poisson superposition principle (see [section 2.3.1](#)). We will not delve into the details of simulating jump destinations, as the methods are straightforward when the number of possible transitions is finite or countable, and very problem specific otherwise.

If these three aspects are well handled, then a PDMP trajectory can be generated as follows. Let $X_0 = x_0 \in \mathcal{X}$ be the initial state, at time $t \geq 0$:

(i) Compute the waiting time before reaching the boundary of the state space

$$t_\partial(X_t) \leftarrow \inf \{s > 0 : \Psi(X_t, s) \in \partial\mathcal{X}_V^-\}. \quad (2.63)$$

(ii) Draw a random time $T_\lambda(X_t)$ according to its survival function

$$s \mapsto \mathbb{P}(T_\lambda(X_t) > s) = \exp(-\Lambda(X_t, s)). \quad (2.64)$$

(iii) Follow the flow during time

$$T(X_t) \leftarrow \min(t_\partial(X_t), T_\lambda(X_t)), \quad (2.65)$$

and set

$$X_{t+T(X_t)}^- \leftarrow \Psi(X_t, T(X_t)). \quad (2.66)$$

(iv) Jump by drawing a destination state $Y \in \mathcal{X}$ according to the jump kernel

$$Y \sim Q(\cdot \mid X_{t+T(X_t)}^-), \quad (2.67)$$

and set

$$\{X_{t+T(X_t)}, t\} \leftarrow \{Y, t + T(X_t)\}, \quad (2.68)$$

then repeat steps (i) to (iv).

2.3.1 PDMP superposition

From the modeler's perspective, it is not always straightforward to formally separate the distribution of jump times from the distribution of jump destinations. Rather than characterizing the latter with an intensity λ and a kernel Q independent of the nature of the jump, it is sometimes more natural to consider the different events that can trigger a jump. The set of triggering events is an arbitrary categorization proposed by the modeler. For example, consider a PDMP modeling an industrial system with d_c components, where a jump occurs when a functional component fails or when a broken component is repaired. One modeling choice is to consider that there are only two possible triggering events at each jump: the failure or repair of a component. Another choice would be to consider that there are d_c possible triggering events: the change in status of the j -th component for $j \in \llbracket 1, d_c \rrbracket$. The modeling of industrial systems using PDMPs will be explored in several examples in [section 3.3.2](#) and in [section 5.3](#).

For any state $x \in \mathcal{X}$, we assume that the modeler has defined $C(x) = \left(c^{(j)}(x) \right)_{j=1}^{d_C(x)}$ a finite or countable set of cardinality $d_C(x)$ corresponding to the set of possible triggering events from $x \in \mathcal{X}$. It is reasonable to assume that this set does not change between jumps, since we can define a change as a boundary jump. Thus $d_C(x) = d_C(\Psi(x, t))$ for all $t \in [0, t_\theta(x))$. The idea is to choose $C(x)$ such that we can easily characterize the distribution of the next jump given the event that triggered it and not just given x . More precisely, we assume that for all $j \in \llbracket 1, d_C(x) \rrbracket$, we have a marginal jump intensity $\lambda^{(j)}$ specific to the triggering event $c^{(j)}(x)$, and a marginal jump kernel $Q^{(j)}$ with marginal density $q^{(j)}$ such that:

- ◊ the waiting time $T_{\lambda^{(j)}}(x)$ before the next jump, given its triggering event $c^{(j)}(x)$ and the current state of the process x , is distributed as the waiting time before the next jump of an inhomogeneous Poisson process with intensity $t \mapsto \lambda^{(j)}(\Psi(x, t))$,
- ◊ the destination of the next spontaneous jump, given its triggering event $c^{(j)}(x)$ and the waiting time $T_{\lambda^{(j)}}(x)$ before the jump from state x , is distributed according to the marginal kernel $Q^{(j)}(\cdot \mid \Psi(x, T_{\lambda^{(j)}}(x)))$.

A property of Poisson processes called *superposition* implies that the global intensity λ of the process is equal to the sum of its marginal intensities $\lambda^{(j)}$. Similarly, the global kernel Q is the mean of the marginal kernels $Q^{(j)}$ weighted by the weights $\lambda^{(j)}/\lambda$ corresponding for all $j \in \llbracket 1, d_C(x) \rrbracket$ to the probability that $c^{(j)}(x)$ is the triggering event of the jump. This gives us an alternative procedure for generating a PDMP trajectory in the case $d_C(x) < \infty$ for any $x \in \mathcal{X}$.

Alternative jump dynamics with multiple trigger events

Let $(X_t)_{t \geq 0}$ be a stochastic process with values in \mathcal{X} , piecewise deterministic trajectories given by Ψ and boundary jumps but with an alternative mechanics for spontaneous jumps.

- ◊ For any state $x^- \in \partial\mathcal{X}$, the distribution of the jump destination is given by a boundary kernel $Q^{(\partial)}(\cdot \mid x^-)$.
- ◊ For any state $x \in \mathcal{X} \setminus \{\partial\mathcal{X}\}$, assume we know $d_C(x) \in \mathbb{N}$ functions $\lambda^{(j)} : \mathcal{X} \rightarrow \mathbb{R}_+$ such that $t \mapsto \lambda^{(j)} \circ \Psi(x, t)$ is locally integrable on $[0, t_\theta(x))$. For any $x \in \mathcal{X} \setminus \{\partial\mathcal{X}\}$ and any $j \in \llbracket 1, d_C(x) \rrbracket$ let us also denote by $\Lambda^{(j)}$ the function satisfying $t \mapsto \Lambda^{(j)}(x, t) = \int_0^t \lambda^{(j)}(\Psi(x, u)) du$.
- ◊ For the same state $x \in \mathcal{X} \setminus \{\partial\mathcal{X}\}$, assume we also know $d_C(x)$ operators $Q^{(j)}$ such that $Q^{(j)}(\cdot \mid \Psi(x, t))$ is a Markovian kernel on \mathcal{X} with kernel density $q^{(j)}(\cdot \mid \Psi(x, t))$ (with respect to a dominant measure $\nu_{q^{(j)}}(\cdot \mid \Psi(x, t))$ on \mathcal{X}) for any time $t \in [0, t_\theta(x))$.

A trajectory of the process $(X_t)_t$ can be generated as follows. Let $X_0 = x_0 \in \mathcal{X}$ be the initial state, at time $t \geq 0$:

(i) Compute the waiting time before reaching the boundary of the state space

$$t_{\partial}(X_t) \leftarrow \inf\{s > 0 : \Psi(X_t, s) \in \partial\mathcal{X}_V\}. \quad (2.69)$$

(ii) For $j \in \llbracket 1, d_C(X_t) \rrbracket$, draw mutually independent random times $T_{\lambda^{(j)}}(X_t)$ according to their survival function

$$s \mapsto \mathbb{P}(T_{\lambda^{(j)}}(X_t) > s) = \exp\left[-\int_0^s \lambda^{(j)}(\Psi(X_t, u)) \, du\right], \quad (2.70)$$

and set

$$J^* \leftarrow \arg \min_{j \in \llbracket 1, d_C(X_t) \rrbracket} T_{\lambda^{(j)}}(X_t), \quad (2.71)$$

$$T_{\lambda}(X_t) \leftarrow T_{\lambda^{(J^*)}}(X_t). \quad (2.72)$$

(iii) Follow the flow during time

$$T(X_t) \leftarrow \min(t_{\partial}(X_t), T_{\lambda}(X_t)), \quad (2.73)$$

and set

$$X_{t+T(X_t)}^- \leftarrow \Psi(X_t, T(X_t)). \quad (2.74)$$

(iv) Jump by drawing a destination state $Y \in \mathcal{X}$ according to the jump kernel

$$Y \sim Q^{(\partial)}(\cdot \mid X_{t+T(X_t)}^-) \mathbb{1}_{T(X_t)=t_{\partial}(X_t)} + Q^{(J^*)}(\cdot \mid X_{t+T(X_t)}^-) \mathbb{1}_{T(X_t)<t_{\partial}(X_t)}, \quad (2.75)$$

and set

$$\{X_{t+T(X_t)}, t\} \leftarrow \{Y, t + T(X_t)\}, \quad (2.76)$$

then repeat steps (i) to (iv).

PROPOSITION 13 · PDMP superposition

The stochastic process $(X_t)_{t \geq 0}$ is a PDMP

(i) with jump intensity $\lambda : x \mapsto \sum_{j=1}^{d_C(x)} \lambda^{(j)}(x)$,

(ii) and jump kernel given for any $B \subset \mathcal{X}$ and any state $x \in \mathcal{X}$ by:

$$Q(B \mid x) = \mathbb{1}_{x \in \partial\mathcal{X}} Q^{(\partial)}(B \mid x) + \mathbb{1}_{x \in \mathcal{X}} \sum_{j=1}^{d_C(x)} Q^{(j)}(B \mid x) \frac{\lambda^{(j)}(x)}{\lambda(x)}. \quad (2.77)$$

Proof. (i) For any $x \in \mathcal{X}$, using independence of the $T_{\lambda^{(j)}}(x)$ for $j \in \llbracket 1, d_C(x) \rrbracket$ we have:

$$\begin{aligned} \mathbb{P}\left(\min_{j \in \llbracket 1, d_C(x) \rrbracket} \{T_{\lambda^{(j)}}(x)\} > s\right) &= \mathbb{P}\left(\bigcap_{j=1}^{d_C(x)} \{T_{\lambda^{(j)}} > s\}\right) = \prod_{j=1}^{d_C(x)} \exp\left[-\int_0^s \lambda^{(j)}(\Psi(x, u)) \, du\right] \\ &= \exp\left[-\int_0^s \sum_{j=1}^{d_C(x)} \lambda^{(j)}(\Psi(x, u)) \, du\right] \\ &= \mathbb{P}\left(T_{\sum_{j=1}^{d_C(x)} \lambda^{(j)}}(x) > s\right). \end{aligned} \quad (2.78)$$

Thus, we can identify and deduce:

$$\lambda(x) = \sum_{j=1}^{d_C(x)} \lambda^{(j)}(x). \quad (2.79)$$

(ii) Let us note Q the jump kernel of the process $(X_t)_t$ and $B \subset \mathcal{X}$. For any $x \in \partial\mathcal{X}$, we clearly have $Q(B \mid x) = Q^\partial(B \mid x)$. For any $x \in \mathcal{X} \setminus \{\partial\mathcal{X}\}$ such that $\lambda(x) > 0$, we have:

$$\begin{aligned} Q(B \mid x) &= \lim_{t \rightarrow 0^+} Q(B \mid \Psi(x, t)) = \lim_{t \rightarrow 0^+} \mathbb{P}(X_{T_\lambda(x)} \in B \mid T_\lambda(x) = t) \\ &= \lim_{t \rightarrow 0^+} \sum_{j=1}^{d_C(x)} \mathbb{P}(\{X_{T(x)} \in B\} \cap \{T_{\lambda^{(j)}}(x) = t\} \mid T_\lambda(x) = t) \\ &= \lim_{t \rightarrow 0^+} \sum_{j=1}^{d_C(x)} \mathbb{P}(X_{T(x)} \in B \mid T_{\lambda^{(j)}}(x) = T_\lambda(x) = t) \times \mathbb{P}(T_{\lambda^{(j)}}(x) = t \mid T_\lambda(x) = t) \\ &= \lim_{t \rightarrow 0^+} \sum_{j=1}^{d_C(x)} Q^{(j)}(B \mid \Psi(x, t)) \times \mathbb{P}(T_{\lambda^{(j)}}(x) = t \mid T_\lambda(x) = t). \end{aligned} \quad (2.80)$$

Remember that for any $x \in \mathcal{X}$ and any $j \in \llbracket 1, d_C(x) \rrbracket$, the waiting time $T_{\lambda^{(j)}}(x)$ admits the probability density function

$$s \mapsto \bar{p}_{\lambda^{(j)}}(s \mid x) = \lambda^{(j)}(x) \exp(-\Lambda^{(j)}(x, s)), \quad (2.81)$$

where $\Lambda^{(j)}(x, s) = \int_0^s \lambda^{(j)}(\Psi(x, t)) \, dt$ is the marginal integrated jump intensity.

$$\begin{aligned} \mathbb{P}(T_{\lambda^{(j)}}(x) = t \mid T_\lambda(x) = t) &= \frac{\mathbb{P}\left(\bigcap_{k \neq j} \{T_{\lambda^{(k)}}(x) > t\} \cap \{T_{\lambda^{(j)}}(x) = t\}\right)}{\bar{p}_\lambda(t \mid x)} \\ &= \frac{\bar{p}_{\lambda^{(j)}}(t \mid x)}{\bar{p}_\lambda(t \mid x)} \times \exp\left(-\sum_{k \neq j} \Lambda^{(k)}(x, t)\right) \\ &= \frac{\lambda^{(j)}(\Psi(x, t)) \exp(-\Lambda^{(j)}(x, t))}{\lambda(\Psi(x, t)) \exp(-\Lambda(x, t))} \times \exp[-(\Lambda(x, t) - \Lambda^{(j)}(x, t))] \\ &= \frac{\lambda^{(j)}(\Psi(x, t))}{\lambda(\Psi(x, t))}. \end{aligned} \quad (2.82)$$

By substituting into eq. (2.80), we get:

$$\begin{aligned}
 Q(B | x) &= \lim_{t \rightarrow 0^+} \sum_{j=1}^{d_C(x)} Q^{(j)}(B | \Psi(x, t)) \times \mathbb{P}(T_{\lambda^{(j)}}(x) = t | T_{\lambda}(x) = t) \\
 &= \sum_{j=1}^{d_C(x)} \lim_{t \rightarrow 0^+} Q^{(j)}(B | \Psi(x, t)) \times \frac{\lambda^{(j)}(\Psi(x, t))}{\lambda(\Psi(x, t))} \\
 &= \sum_{j=1}^{d_C(x)} Q^{(j)}(B | x) \times \frac{\lambda^{(j)}(x)}{\lambda(x)}.
 \end{aligned} \tag{2.83}$$

■

2.3.2 PDMP thinning

The superposition principle offers the possibility to decompose a PDMP with intensity λ into several marginal PDMPs with marginal intensity $\lambda^{(j)}$ such that λ is the sum of the $\lambda^{(j)}$, and therefore to generate the target process by combining the marginal processes. This is a phenomenon that works in both directions: generating a process with intensity λ and keeping the points with probability $\lambda^{(j)}/\lambda$ is equivalent to generating the process with intensity $\lambda^{(j)}$. This is known as Poisson marking or thinning.

THEOREM 3 · Thinning of an inhomogeneous Poisson process

Suppose we want to generate an inhomogeneous Poisson process with intensity λ . Let T_1, T_2, \dots be the inter-jump times of an other inhomogeneous Poisson process with intensity $\bar{\lambda}$ such that $\bar{\lambda}(t) \geq \lambda(t)$ for any $t \geq 0$. Then, the counting process $(N_t)_t$ formed by the remaining times when deleting each point T_i with probability $1 - \frac{\lambda(T_i)}{\bar{\lambda}(T_i)}$ is an inhomogeneous Poisson process with intensity λ .

The proof is given in the original article of 1979 by Lewis and Shedler ([LS79]) with an associated algorithm for the simulation of the process.

This is also one of the most popular methods for generating the jump times of a PDMP. In this case, however, only one point can be generated at a time (rather than a sequence from which elements are subsequently removed) because the intensity changes randomly after each jump. It is a sequential rejection method:

- (i) after each jump, from a state $x \in \mathcal{X}$, we draw a time $T_{\bar{\lambda}}(x)$ according to the survival function: $s \mapsto \exp\left[-\int_0^s \bar{\lambda}(\Psi(x, u)) du\right]$,
- (ii) if $T_{\bar{\lambda}}(x) > t_{\partial}(x)$, we jump at the boundary,
- (iii) else, we generate $U \sim \text{Unif}([0, 1])$ and accept $T_{\bar{\lambda}}(x)$ only if $\lambda(\Psi(x, T_{\bar{\lambda}}(x))) \geq U \bar{\lambda}(\Psi(x, T_{\bar{\lambda}}(x)))$,

- (iv) if this jump time is rejected, we follow the flow Ψ until time $T_\lambda(x)$ (no jump occurs before), then we restart the procedure from this new state.

The efficiency of the method depends on the average proportion of proposed jump times that are accepted, and therefore on the tightness of the upper bound $\bar{\lambda}$. The article [LTT18] analyses the efficiency of the method for several types of bounds.

In practice, the true computational cost lies in solving the flow and not in proposing jump times. Therefore, we propose [algorithm 4](#), an algorithm that aims to be economical in terms of the number of flow calls, at the cost of a potentially large number of generated random variables. The idea is to progress with a time step $h > 0$. At each iteration, starting from the current state $x \in \mathcal{X}$:

- (i) we solve the flow for a duration h or shorter if the boundary of the state space is met:

$$[\Psi(x, t)]_{t \in [0, h \wedge t_\partial(x)]},$$

- (ii) we determine $\bar{\lambda}$, a constant upperbound of λ on this interval,
- (iii) we generate a proposition \bar{T} with exponential distribution with parameter $\bar{\lambda}$,
- (iv) if the proposed time is greater than $t_\partial(x)$, we jump at the boundary,
- (v) if the proposed time is greater than h , we restart the procedure from state $\Psi(x, h)$,
- (vi) else, with probability $\frac{\lambda(\bar{T})}{\bar{\lambda}}$, the jump time is accepted and we jump from state $\Psi(x, \bar{T})$,
- (vii) and with probability $1 - \frac{\lambda(\bar{T})}{\bar{\lambda}}$, we do not jump and start over the procedure from state $\Psi(x, \bar{T})$ on interval $[0, h - \bar{T})$ (we go to the end of the time step before re-evaluating the flow over a new interval).

2.3.3 The inversion method with a twist

The other method, even more natural, for generating the jump times of an inhomogeneous Poisson process is the inversion method ([proposition 1](#)) of the c.d.f. of the inter-jump times. Indeed, let us note

$$F_x : s \in \mathbb{R}_+ \mapsto \mathbb{P}(T_\lambda(x) > s) = 1 - \exp\left[-\int_0^s \lambda(\Psi(x, t)) dt\right]. \quad (2.92)$$

We have:

$$F_x^{-1} : u \in [0, 1] \mapsto \inf\{s \geq 0 : F_x(s) \geq u\} = \inf\left\{s \geq 0 : \int_0^s \lambda(\Psi(x, t)) dt \geq -\log(1 - u)\right\}. \quad (2.93)$$

This leads to a “direct” simulation method. It is sufficient to solve:

$$T_\lambda^-(x) = F_x^{-1}(\text{Unif}([0, 1])) = \inf\{s \geq 0 : \Lambda(x, s) \geq \text{Exp}(1)\}. \quad (2.94)$$

If we know for all $\ell > 0$ the function $\Lambda^{-1}(x, \ell) = \inf\{s \geq 0 : \Lambda(x, s) \geq \ell\}$, the method is immediate. Otherwise, we must numerically invert Λ . This is why this method is usually less recommended. This inversion is not necessarily simple, and unlike the thinning method, it requires evaluating the integral Λ at several points rather than simply the intensity λ . It can therefore be very costly in terms of flow calls.

However, there are more refined implementations that can partially overcome these difficulties. We propose the following implementation. It consists of two sequences.

In the first sequence:

- (i) we draw once and for all $Y \sim \text{Exp}(1)$,
- (ii) we evaluate $\Psi(x, t)$ step by step on time intervals of length $h > 0$,
- (iii) if we meet the boundary, we jump when we reach it,
- (iv) by additivity of the integral, we update Λ by adding the integral of λ on the new time interval covered,
- (v) we stop the first sequence as soon as Λ exceeds Y .

This first sequence allows us to determine an interval $[L^-, L^+]$ containing Y , and therefore an interval $[\bar{t}, \bar{t} + h]$ of length h containing $T_\lambda^-(x)$. The function f defined by

$$f : s \in [\bar{t}, \bar{t} + h] \mapsto \Lambda(x, s) - Y, \quad (2.95)$$

is negative in \bar{t} , nonnegative in $\bar{t} + h$ and has a positive derivative $f'(s) = \lambda(\Psi(x, s))$. It therefore has a unique root in the interval $[\bar{s}, \bar{s} + h]$ that can be determined in a second sequence using a Newton-Raphson method [VC95]. At each iteration, a new point is proposed

$$s_{n+1} = \min\left(\max\left(\bar{s}, s_n - \frac{f(s_n)}{f'(s_n)}\right), \bar{s} + h\right). \quad (2.96)$$

The evaluation of the flow in the expression of f and f' is made much less costly by the fact that it is performed several times on the same small time interval.

ALGORITHM 4 · PDMP thinning

1 **Initialization** state $x \in \mathcal{X}$ and step size $h > 0$.

$$[y, \bar{h}, \bar{t}] \leftarrow [x, h \wedge t_\partial(y), 0]. \quad (2.84)$$

while a jump time T is not returned **do**

2 **Majoration step**

$$\bar{\lambda} = \sup_{t \in [0, \bar{h}]} \lambda(\Psi(y, t)). \quad (2.85)$$

Sampling step

$$\bar{T} \sim \text{Exp}(\bar{\lambda}). \quad (2.86)$$

if $\bar{T} \geq t_\partial(y)$ **then**

3 **Return**

$$[T, x_T] \leftarrow [\bar{t} + t_\partial(y), \Psi(y, t_\partial(y))]. \quad (2.87)$$

4 **else if** $\bar{T} \geq h$ **then**

5 **Update**

$$[y, \bar{h}, \bar{t}] \leftarrow [\Psi(y, h), h \wedge t_\partial(y), \bar{t} + h]. \quad (2.88)$$

6 **else**

7 **Accept-reject step**

$$U \sim \text{Unif}([0, 1]), \quad (2.89)$$

if $U \times \bar{\lambda} \leq \lambda(\Psi(y, \bar{T}))$ **then**

8 **Return**

$$[T, x_T] \leftarrow [\bar{t} + \bar{T}, \Psi(y, \bar{T})]. \quad (2.90)$$

else

9 **Update**

$$[y, \bar{h}, \bar{t}] \leftarrow [\Psi(y, \bar{T}), \bar{h} - \bar{T}, \bar{t} + \bar{T}]. \quad (2.91)$$

10 **end**

CHAPTER 3

Importance sampling of piecewise deterministic Markov processes

This chapter aims to present the fundamentals of importance sampling of PDMPs. This method requires the evaluation of the likelihood of a PDMP trajectory. We first give the definition of a dominant measure on the space of PDMP trajectories with respect to which they admit a probability density function. We then describe the process with optimal importance distribution. This optimal importance process is a piecewise deterministic process (not necessarily Markovian). It preserves the state space and flow of the original PDMP. The jump dynamics of the optimal importance process can be explicitly described. In the rare event estimation case, there is a Markovian reformulation of this optimal process. The expression of its jump intensity and jump kernel is then directly connected to the so-called “committor function” of the process. This function returns the probability that a trajectory realizes the event of interest given its current state. This committor function is not known but can be approximated. At the end of the chapter, we present an implementation of importance sampling based on a simple approximation of the committor function.

Content of this chapter

3.1	Distribution of a PDMP trajectory	66
3.1.1	Skeleton of a trajectory	66
3.1.2	Space of admissible trajectories	68
3.1.3	Likelihood of a PDMP trajectory	69
3.2	Importance sampling of PDMP	72
3.2.1	Admissible importance process	72
3.2.2	Optimal importance process	75
3.3	Rare event application	77
3.3.1	Committor function	77
3.3.2	Practical implementation	79

3.1 Distribution of a PDMP trajectory

Many statistical problems and methods (such as importance sampling) require the ability to compute the likelihood of an observation. Do we have such a notion for observed PDMP trajectories? In his PhD thesis [Gal19, Section 3.1 p72] (see also [Chr+19, Section 3.1]), Thomas Galtier proved that we do. He explained a dominant measure according to which a PDMP trajectory of fixed duration admits a probability density function. The essential results for constructing a measure space of trajectories and the explicit pdf are given here. The first step is to construct a relevant space in which PDMP trajectories live. This space will then be equipped with the appropriate σ -algebra and measure.

A PDMP trajectory is an application from \mathbb{R}_+ to \mathcal{X} . The Skorokhod space $\mathbf{D}(\mathbb{R}_+, \mathcal{X})$ of càdlàg functions from \mathbb{R}_+ to \mathcal{X} is a common choice to describe the set of all possible trajectories¹, but it is too general in our case. It is more convenient to restrict ourselves at least to the space of feasible trajectories on \mathcal{X} for a given deterministic flow Ψ . Recall that for any set $B \subseteq \mathbb{R}_+$, we denote $X_B = (X_t)_{t \in B}$. We also introduce the following notation: for any $s > 0$, X_s^- is the last state of the trajectory $X_{[0, s]}$ ².

3.1.1 Skeleton of a trajectory

For any $s \in \mathbb{R}_+ \cup \{+\infty\}$, let $x_{[0, s]}$ be a PDMP trajectory (of finite or infinite duration) on \mathcal{X} with deterministic flow Ψ . Let us denote by s_k the time of the k -th jump and t_k the waiting time between the k -th and the $(k + 1)$ -th jump. More formally, starting at time $s_0 = 0$, we have :

$$\begin{aligned} t_k &= t(x_{s_k}) = \inf\{t \geq 0 : \Psi(x_{s_k}, t) \neq x_{s_k+t}\}, \\ s_{k+1} &= s_k + t_k. \end{aligned} \tag{3.1}$$

The first line can be interpreted as follows: a jump has occurred when the state of the PDMP is no longer the one that would have been obtained by following the flow since the previous jump. We define κ , the companion counting process of the PDMP $(X_t)_t$. It associates to any time $s \geq 0$, the number of jumps performed by the PDMP before time s .

$$\kappa : s \in \mathbb{R}_+ \longmapsto \sup\{k \in \mathbb{N} : s_k < s\}. \tag{3.2}$$

Assuming non explosion of the PDMP (definition 16) we have $\kappa(s) < +\infty$ for any $s < \infty$.

¹The space of trajectories is more often referred to as the path space in the literature on stochastic processes. However, we wanted to avoid confusion with the reliability notion of path set, which will be introduced in chapter 5.

²In the case of a jump at time s , X_s^- refers to the starting state of the jump, and X_s to the ending state (X_s being the last state of $X_{[0, s]}$).

Converting a trajectory into its skeleton

Since the randomness of the process only intervenes at jump times, a PDMP trajectory can be precisely reconstructed from its pairs of jump times and jump destinations.

DEFINITION 18 · Skeleton of a trajectory

The skeleton of a PDMP trajectory is the list of the states of the trajectory after each jump and the list of the waiting times between the jumps. We note \mathfrak{S} the map that changes a trajectory $x_{[0,s]}$ of duration $s \in [0, \infty]$ in its skeleton.

$$\mathfrak{S} : x_{[0,s]} \mapsto \left(x_{s_k}, t_k \right)_{k=0}^{x(s)}. \quad (3.3)$$

The times $(t_k)_{k=0}^{x(s)-1}$ are those given in eq. (3.1). If the duration s is finite, we have $t_{x(s)} = s - s_{x(s)}$.

Let $(x_{s_k}, t_k)_{k=0}^n$ be a sequence in $(\mathcal{X} \times \mathbb{R}_+^*)^{n+1}$. We set $s_{k+1} = s_k + t_k$ for all $k \in \llbracket 0, n \rrbracket$. It is the skeleton of a PDMP trajectory with n jumps and deterministic flow Ψ only if the jumps did not occur after hitting the boundaries of the state space, i.e. for all $k \in \llbracket 0, n \rrbracket$:

$$t_k \leq t_\partial(x_{s_k}), \text{ i.e. } \Psi(x_{s_k}, t) \notin \partial\mathcal{X}_{v_{s_k}} \quad \forall t < t_k. \quad (3.4)$$

Converting a PDMP trajectory into its skeleton (or conversely converting a skeleton into a PDMP trajectory) requires the knowledge of the deterministic flow Ψ , but not of the jump intensity λ and jump kernel Q .

Mesurable space of skeletons

For any $n \in \mathbb{N}$, we define $\mathcal{S}^{(n)}$ the set of skeletons of PDMP trajectories with n jumps and deterministic flow Ψ :

$$\mathcal{S}^{(n)} = \left\{ \left(x_{s_k}, t_k \right)_{k \in \llbracket 0, n \rrbracket} \in (\mathcal{X} \times \mathbb{R}_+^*)^{n+1} \mid t_k \leq t_\partial(x_{s_k}), \forall k \in \llbracket 0, n \rrbracket \right\}. \quad (3.5)$$

A skeleton $(x_{s_k}, t_k)_{k=0}^n \in \mathcal{S}^{(n)}$ corresponds to a PDMP trajectory of duration $s_n + t_n$. The advantage of decomposing the space of skeletons according to the number of jumps n is that $\mathcal{S}^{(n)}$ can easily be equipped with a topology. Indeed, as seen in eq. (2.49), we can define a distance between two states of the PDMP. Suppose that we have a distance between two states of the time-augmented PDMP (where the position includes the elapsed time). Since two skeletons in

$\mathcal{S}^{(n)}$ have the same number of elements, it suffices to take:

$$\text{dist}_{\mathcal{S}^{(n)}}\left(\left[x_{s_k}^{(1)}, t_k^{(1)}\right]_{k=0}^n, \left[x_{s_k}^{(2)}, t_k^{(2)}\right]_{k=0}^n\right) = \sum_{k=0}^n \text{dist}_{\tilde{\mathcal{X}}}\left(\tilde{x}_{s_k}^{(1)}, \tilde{x}_{s_k}^{(2)}\right), \quad (3.6)$$

where $\tilde{x}_{s_k} = (x_{s_k}, s_k)$ is the time-augmented state and $\tilde{\mathcal{X}}$ the time-augmented space $\mathcal{X} \times \mathbb{R}_+$. Note that this distance only requires the two skeletons to be the same length, but does not require them to correspond to trajectories of the same duration or to start from the same point. Thereby, we can define $\mathcal{B}(\mathcal{S}^{(n)})$, the Borel σ -algebra on $\mathcal{S}^{(n)}$ equipped with $\text{dist}_{\mathcal{S}^{(n)}}$.

DEFINITION 19 · σ -algebra on the skeleton space

We define \mathcal{S} , the σ -algebra on the skeleton space generated by the union, for every possible numbers of jumps n , of the Borelians of every skeletons with n jumps :

$$\mathcal{S} = \sigma\left(\bigcup_{n \in \mathbb{N}} \mathcal{B}(\mathcal{S}^{(n)})\right). \quad (3.7)$$

3.1.2 Space of admissible trajectories

The application \mathfrak{S} is bijective. Its inverse is given for all $n \in \mathbb{N}$ (with $\kappa(t)$ given in eq. (3.2)) by :

$$\mathfrak{S}^{-1} : \left(x_{s_k}, t_k\right)_{k=0}^n \in \mathcal{S}^{(n)} \longmapsto \left[\Psi\left(x_{s_{\kappa(t)}}, t - s_{\kappa(t)}\right)\right]_{t \in [0, s_n + t_n]}. \quad (3.8)$$

So any skeleton of $\mathcal{S}^{(n)}$ corresponds to a unique PDMP trajectory of duration $s_n + t_n$. In practice, it is useful to know the states preceding the jumps without additional call to the flow. One can keep in memory an extended skeleton from a simulation: $(x_{s_k}^-, x_{s_k}, t_k)_k$ with $x_{s_{k+1}}^- = \Psi(x_{s_k}, t_k)$.

DEFINITION 20 · Measurable space of PDMP trajectories

We denote by \mathbf{T} the space of all feasible trajectories on \mathcal{X} with flow Ψ . It can be defined as the preimage for \mathfrak{S} of all the admissible skeletons.

$$\mathbf{T} := \bigcup_{n \in \mathbb{N}} \mathfrak{S}^{-1}(\mathcal{S}^{(n)}). \quad (3.9)$$

We also define \mathcal{T} the σ -algebra on the trajectory space :

$$\mathcal{T} = \mathfrak{S}^{-1}(\mathcal{S}). \quad (3.10)$$

Trajectories of fixed duration

For any $s \in \mathbb{R}_+$, we define $\mathcal{S}_s^{(n)}$ the set of skeletons of PDMP trajectories with n jumps, deterministic flow Ψ and of duration s :

$$\mathcal{S}_s^{(n)} = \left\{ (x_k, t_k)_{k \in \llbracket 0, n \rrbracket} \in \mathcal{S}^{(n)} \mid s_n + t_n = s \right\}. \quad (3.11)$$

We can therefore extend the previous definitions to the case of spaces of trajectories of finite duration $s > 0$:

$$\begin{aligned} \mathbf{T}_s &:= \bigcup_{n \in \mathbb{N}} \mathfrak{S}^{-1}(\mathcal{S}_s^{(n)}), \\ \mathcal{S}_s &= \sigma\left(\bigcup_{n \in \mathbb{N}} \mathcal{B}(\mathcal{S}_s^{(n)})\right), \\ \mathcal{T}_s &= \mathfrak{S}^{-1}(\mathcal{S}_s). \end{aligned} \quad (3.12)$$

Trajectories of random duration

Similarly, we can generalise to the case of trajectories of random duration. First, remark that $(\mathcal{T}_s)_{s \geq 0}$ is a filtration for the measurable space $(\mathbf{T}, \mathcal{T})$. Let τ be a stopping time with respect to $(\mathcal{T}_s)_{s \geq 0}$. We have:

$$\begin{aligned} \mathbf{T}_\tau &= \left\{ (x_t)_{t \in [0, \tau)} \in \mathbf{T} \right\}, \\ \mathcal{S}_\tau^{(n)} &= \mathcal{S}^{(n)} \cap \mathfrak{S}(\mathbf{T}_\tau), \\ \mathcal{S}_\tau &= \sigma\left(\bigcup_{n \in \mathbb{N}} \mathcal{B}(\mathcal{S}_\tau^{(n)})\right), \\ \mathcal{T}_\tau &= \mathfrak{S}^{-1}(\mathcal{S}_\tau). \end{aligned} \quad (3.13)$$

3.1.3 Likelihood of a PDMP trajectory

The space of skeletons and the space of trajectories are characterized by the state space \mathcal{X} and the deterministic flow Ψ . The distribution of a PDMP trajectory, on a given space and for a given flow, is characterized by the jump intensity λ and the jump kernel Q . Since the process is Markovian, the law of a PDMP trajectory can be expressed as the product of the distributions of the consecutive elements of its skeleton knowing the previous one. For any $A \subset \mathcal{X}$, $A' \subset \mathbb{R}_+$, we have:

$$\begin{aligned} \mathbb{P}\left[\left(X_{S_{k+1}}, T_k\right) \in A \times A' \mid \left(X_{S_k}, S_k\right) = \left(x_{s_k}, s_k\right)\right] &= \int_A \int_{A'} p_\lambda(t \mid x_{s_k}) \times q(x \mid \Psi(x_{s_k}, t)) \\ &\quad \times \nu_q(dx \mid \Psi(x_{s_k}, t)) \times \nu_\lambda(dt \mid x_{s_k}), \end{aligned} \quad (3.14)$$

where p_λ , given in eq. (2.47), is the pdf of a jump time with respect to the measure ν_λ , and q is the pdf of the jump kernel Q with respect to the measure ν_q .

Product measure for a n -jumps skeleton

It is clear from eq. (3.14) that a reference measure on $(\mathcal{S}^{(n)}, \mathcal{S})$ contains at least the product of measures ν_λ and ν_q among the various elements of the skeleton. Therefore, we define the following measure for skeletons with n jumps:

$$\nu_{\lambda,q}^{(n)}\left(d[x_{s_k}, t_k]_{k=0}^n\right) := \prod_{k=0}^{n-1} \nu_\lambda(dt_k | x_{s_k}) \times \nu_q(dx_{s_{k+1}} | \Psi(x_{s_k}, t_k)). \quad (3.15)$$

Let $x_{[0,s]}$ be an observed PDMP trajectory of duration $s > 0$. Suppose $x_{[0,s]}$ has n jumps. The corresponding skeleton is $\mathfrak{S}(x_{[0,s]}) = [x_{s_k}, t_k]_{k=0}^n \in \mathcal{S}_s^{(n)}$. The last state of the trajectory is $x_{s_n}^- = \Psi(x_{s_n}, t_n)$ and $t_n = s - s_n$. The probability distribution of T_n , the n -th and last element of the skeleton of a PDMP trajectory of finite duration s , is given, for any $A \in \mathbb{R}^+$, by:

$$\begin{aligned} \mathbb{P}[T_n \in A \mid (X_{S_n}, S_n) = (x_{s_n}, s_n)] &= \int_A \exp(-\Lambda(x_{s_n}, t)) \partial_{s-s_n}(dt) \\ &= \exp(-\Lambda(x_{s_n}, s - s_n)) \mathbb{1}_{\{s-s_n\} \subset A}. \end{aligned} \quad (3.16)$$

Here $\exp(-\Lambda(x_{s_n}, t))$ represents the probability that no other jump has occurred between the time s_n and the time $s_n + t$, the end date of the trajectory. The probability distribution of a skeleton with n jumps and duration $s > 0$ can be derived, given its initial state x_0 . For any $B \subset \mathcal{S}_s^{(n)}$, we have :

$$\begin{aligned} \mathbb{P}\left[\left(X_{S_k}, T_k\right)_{k=0}^n \in B \mid X_0 = x_0\right] &= \int_B \left(\prod_{k=0}^{n-1} p_\lambda(t_k | x_{s_k}) \times q(x_{s_{k+1}} | \Psi(x_{s_k}, t)) \right) \\ &\quad \times \nu_{\lambda,q}^{(n)}\left(d[x_{s_k}, t_k]_{k=0}^n\right) \times \exp(-\Lambda(x_{s_n}, t_n)) \times \partial_{s-s_n}(dt_n). \end{aligned} \quad (3.17)$$

More generally, for any $C \subset \mathcal{S}_\tau^{(n)}$ with τ an a.s. finite stopping time, we have the same result except for the final Dirac term:

$$\begin{aligned} \mathbb{P}\left[\left(X_{S_k}, T_k\right)_{k=0}^n \in C \mid X_0 = x_0\right] &= \int_C \left(\prod_{k=0}^{n-1} p_\lambda(t_k | x_{s_k}) \times q(x_{s_{k+1}} | \Psi(x_{s_k}, t)) \right) \\ &\quad \times \nu_{\lambda,q}^{(n)}\left(d[x_{s_k}, t_k]_{k=0}^n\right) \times \exp(-\Lambda(x_{s_n}, t_n)) \times \partial_{\tau-s_n}(dt_n). \end{aligned} \quad (3.18)$$

Reference measure and probability density function for PDMP trajectories

Based on eq. (3.18) and using the application \mathfrak{S} , we can establish a reference measure and a pdf for PDMP trajectories of finite duration³.

DEFINITION 21 · Dominant measure

Let τ be an almost surely finite stopping time with respect to the filtration $(\mathcal{F}_s)_{s \geq 0}$. The reference measure on the measurable space $(\mathbf{T}_\tau, \mathcal{T}_\tau)$ of PDMP trajectories of duration τ is defined as the product of:

- (i) the pullback measure, by the function \mathfrak{S} , of the product of the transitions measures between the elements of the skeleton,
- (ii) a Dirac measure to ensure that the last waiting time brings us to the expected end date τ .

$$\begin{aligned} \zeta_\tau(dx_{[0, \tau)}) &= \delta_{\tau - s_{\kappa(\tau)}}(dt_{\kappa(\tau)}) \times \nu_{\lambda, q}^{(x(\tau))}(\mathfrak{S}(dx_{[0, \tau)})), \\ &= \delta_{\tau - s_{\kappa(\tau)}}(dt_{\kappa(\tau)}) \prod_{k=0}^{\kappa(\tau)-1} \nu_\lambda(dt_k | x_{s_k}) \times \nu_q(dx_{s_{k+1}} | \Psi(x_{s_k}, t_k)). \end{aligned} \quad (3.19)$$

THEOREM 4 · Probability density function of a PDMP trajectory

Let τ be an almost surely finite stopping time with respect to the filtration $(\mathcal{F}_s)_{s \geq 0}$.

- (i) If $\sup_{x \in \mathcal{X}} \nu_q(\mathcal{X} | x) < +\infty$ then ζ_τ is a σ -finite measure.
- (ii) Let $x_{[0, \tau)}$ be a PDMP trajectory of duration τ with local characteristics (Ψ, λ, q) . We note $[x_{s_k}, t_k]_{k=0}^n = \mathfrak{S}(x_{[0, \tau)})$ its skeleton with $n = \kappa(\tau)$ its number of jumps. Then the probability density function of a PDMP trajectory with respect to the measure ζ_τ and evaluated on $x_{[0, \tau)}$ is given by:

$$\mathbf{p}(x_{[0, \tau)}) = \exp(-\Lambda(x_{s_n}, t_n)) \prod_{k=0}^{n-1} p_\lambda(t_k | x_{s_k}) q(x_{s_{k+1}} | x_{s_k}^-). \quad (3.20)$$

The proof is provided in [Chr+19, Theorem 3.3]. The initial state x_0 of the trajectory is not considered as random in theorem 4. It is still possible to assume that X_0 is randomly drawn from a pdf q_0 with respect to a measure ν_{q_0} on \mathcal{X} . One just has to multiply ζ by ν_{q_0} and \mathbf{p} by $q_0(x_0)$.

³It is a slight extension of Thomas Galtier's work [Gal19] who established them for PDMP trajectories of fixed duration.

3.2 Importance sampling of PDMP

We finally reach the central problem addressed in this thesis: efficiently estimating the probability that a trajectory of a piecewise deterministic Markov process hits a target region of its state space before a given time. As seen in [section 1.1.2](#), at a given level of precision, classical Monte Carlo methods consume a simulation budget inversely proportional to the rarity of the event whose probability we seek to estimate. We first present the general case of importance sampling of PDMPs for the estimation of the mean value of any function of a trajectory. We then discuss the specific case of rare events.

A state space \mathcal{X} and a triplet of local characteristics (Ψ, λ, Q) are assumed to be fixed once and for all. Let τ be an almost surely finite stopping time. We denote again by \mathbf{T} the space on which the PDMP trajectories live, and \mathbf{T}_τ the subspace of trajectories of duration τ . Finally, let \mathbf{p} denote the pdf of the distribution of a PDMP trajectory on \mathbf{T}_τ characterized by the jump intensity λ and jump kernel Q . We are interested in estimating \mathcal{I} , the mean value of a target function $\mathbf{I} : \mathbf{T}_\tau \rightarrow \mathbb{R}$ of a PDMP trajectory of duration τ . In the following, we denote $\mathbf{x} = x_{[0, \tau]} \in \mathbf{T}_\tau$, a PDMP trajectory of duration τ .

$$\text{(QoI)} \quad \mathcal{I} := \mathbb{E}_{\mathbf{X} \sim \mathbf{p}}[\mathbf{I}(\mathbf{X})] = \int_{\mathbf{T}_\tau} \mathbf{I}(\mathbf{x}) \mathbf{p}(\mathbf{x}) \zeta_\tau(\mathrm{d}\mathbf{x}). \quad (3.21)$$

The standard Monte Carlo estimator for \mathcal{I} is derived from an i.i.d. sample of trajectories from the distribution \mathbf{p} , given by:

$$\hat{\mathcal{I}}_n^{\text{CMC}} := \frac{1}{n} \sum_{k=1}^n \mathbf{I}(\mathbf{X}_k) \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \mathbb{E}_{\mathbf{p}}[\mathbf{I}(\mathbf{X})] = \mathcal{I} \quad \text{with} \quad \mathbf{X}_1, \dots, \mathbf{X}_n \stackrel{\text{i.i.d.}}{\sim} \mathbf{p}. \quad (3.22)$$

To estimate \mathcal{I} using importance sampling, it is necessary to be able to generate trajectories according to an alternative and admissible distribution with evaluable pdf \mathbf{g} . We would then have :

$$\hat{\mathcal{I}}_n^{\text{IS}} := \frac{1}{n} \sum_{k=1}^n \mathbf{I}(\mathbf{X}_k) \frac{\mathbf{p}(\mathbf{X}_k)}{\mathbf{g}(\mathbf{X}_k)} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \mathbb{E}_{\mathbf{g}} \left[\mathbf{I}(\mathbf{X}) \frac{\mathbf{p}(\mathbf{X})}{\mathbf{g}(\mathbf{X})} \right] = \mathcal{I} \quad \text{with} \quad \mathbf{X}_1, \dots, \mathbf{X}_n \stackrel{\text{i.i.d.}}{\sim} \mathbf{g}. \quad (3.23)$$

3.2.1 Admissible importance process

An admissible importance process must have an absolutely continuous density \mathbf{g} with respect to the measure ζ_τ . Therefore, a trajectory of this process should be reconstructible from its skeleton using the mapping \mathfrak{S}^{-1} . It follows that an admissible process is a piecewise deterministic process (PDP) living in the same state space and subjected to the same deterministic flow Ψ as the original process with distribution \mathbf{p} . A more detailed version of this reasoning is provided in [[Gal19](#), section 4.1]. However, there is no requirement for the admissible process to be

Markovian. The jump mechanism can rely on the entire past of the trajectory and not only of the current state.

Partial trajectories

Let $(X_s)_s$ be the original PDMP with local characteristics (Ψ, λ, Q) and consider the partial trajectory $s \mapsto X_{[0,s]}$. Since $(X_s)_s$ is Markovian, $(s \mapsto X_{[0,s]})$ is also a PDMP, but on \mathbf{T} . Its local characteristics are function of the past of the original PDMP trajectory.

(i) Its deterministic flow Ψ is given for any times s, t and any partial trajectory $x_{[0,s]}$ by:

$$\Psi(x_{[0,s]}, t) = (\tilde{x}_u)_{[0,s+t]}, \quad (3.24)$$

with $\tilde{x}_u = x_u$ for any $u < s$ and $\tilde{x}_u = \Psi(x_s, u - s)$ for any $u \geq s$.

(ii) Its jump intensity λ is given for any partial trajectory $x_{[0,s]}$ by:

$$\lambda(x_{[0,s]}) = \lambda(x_s^-). \quad (3.25)$$

(iii) Finally its jump kernel \mathbf{Q} admits a pdf \mathbf{q} given for any partial trajectories $x_{[0,s]}$ and $\tilde{x}_{[0,s]}$ by:

$$\mathbf{q}(\tilde{x}_{[0,s]} \mid x_{[0,s]}) = q(\tilde{x}_s \mid x_s^-) \mathbf{1}_{\tilde{x}_{[0,s]} = x_{[0,s]}}. \quad (3.26)$$

We now introduce a tool that will be useful in the following. The family of operators $\{\mathbf{P}_t\}_t$, described in [definition 22](#), can be interpreted as a semigroup for the process $(s \mapsto X_{[0,s]})$.

DEFINITION 22 · Trajectorial semigroup

For any bounded function $\varphi : \mathbf{T} \rightarrow \mathbb{R}$, any partial trajectory $x_{[0,s]}$ and any time $t \geq 0$, we define:

$$\mathbf{P}_t \varphi(x_{[0,s]}) := \mathbb{E}[\varphi(X_{[0,s+t]}) \mid X_{[0,s]} = x_{[0,s]}]. \quad (3.27)$$

For any $t \geq 0$, the operator \mathbf{P}_t characterizes the distribution of the partial trajectory after an additional time $t > 0$. It is worth noting that $\mathcal{I} = \mathbf{P}_\tau \mathbf{I}(x_0)$ (considering x_0 as a partial trajectory of duration 0). We also introduce the family of operators $\{\mathbf{P}_t^-\}_t$ that assumes that the trajectory jumps immediately before being extended.

DEFINITION 23 · Edge trajectorial semigroup

For any bounded function $\varphi : \mathbf{T} \rightarrow \mathbb{R}$, any partial trajectory $x_{[0,s]}$ and any time $t \geq 0$, we define:

$$\begin{aligned} \mathbf{P}_t^- \varphi(x_{[0,s]}) &= \mathbf{Q} \mathbf{P}_t \varphi(x_{[0,s]}) = \mathbb{E}_{X_{[0,s]} \sim \mathbf{q}(\cdot | x_{[0,s]})} [\mathbf{P}_t \varphi(X_{[0,s]})], \\ &= \int_{\mathbf{T}} \mathbb{E}[\varphi(X_{[0,s+t]}) \mid X_{[0,s]} = x_{[0,s]}] \mathbf{Q}(dx_{[0,s]} \mid x_{[0,s]}), \\ &= \int_{\mathcal{X}} \mathbb{E}[\varphi(X_{[0,s+t]}) \mid X_{[0,s]} = x_{[0,s]}, X_s = x_s] Q(dx_s \mid x_s^-). \end{aligned} \quad (3.28)$$

Support conditions for the importance process

Let $(Y_s)_s$ be the importance piecewise deterministic process and $(s \mapsto Y_{[0,s]})$ the associated partial trajectory. Since the deterministic flow of $(Y_s)_s$ is Ψ , the flow of $(s \mapsto Y_{[0,s]})$ is given by $\mathbf{\Psi}$. The jump mechanism of $(Y_s)_s$ can be characterized by a jump intensity λ' and a jump kernel \mathbf{Q}' with density \mathbf{q}' for the process $(s \mapsto Y_{[0,s]})$. For any partial trajectory $y_{[0,s]}$ and time $t \geq 0$, we note:

$$\begin{aligned} \Lambda'(y_{[0,s]}, t) &= \int_0^t \lambda'(\mathbf{\Psi}(y_{[0,s]}, u)) du, \\ p_{\lambda'}(t \mid y_{[0,s]}) &= \lambda'(\mathbf{\Psi}(y_{[0,s]}, t))^{\mathbf{1}_{t < t_0(y_s)}} \exp(-\Lambda'(y_{[0,s]}, t)). \end{aligned} \quad (3.29)$$

Let $y_{[0,\tau]} \in \mathbf{T}_\tau$ be a trajectory of the importance process with n jumps. The pdf \mathbf{g} of $y_{[0,\tau]}$ with respect to ζ_τ is thus given by:

$$\mathbf{g}(y_{[0,\tau]}) = \exp(-\Lambda'(y_{[0,s_n]}, t_n)) \prod_{k=0}^{n-1} p_{\lambda'}(t_k \mid y_{[0,s_k]}) \mathbf{q}'(y_{[0,s_{k+1}]} \mid y_{[0,s_k]}). \quad (3.30)$$

The pdf \mathbf{g} characterizes an admissible distribution for importance sampling if and only if:

$$\mathbf{I}(y_{[0,\tau]}) \times \mathbf{p}(y_{[0,\tau]}) \neq 0 \implies \mathbf{g}(y_{[0,\tau]}) \neq 0 \quad \text{for } \zeta_\tau\text{-almost all } y_{[0,\tau]} \in \mathbf{T}_\tau. \quad (3.31)$$

The proposition [Gal19, Property 6.] provides conditions on λ' and \mathbf{q}' for the admissibility of the process.

PROPOSITION 14 · Admissible importance process

We have $\mathbf{I}(\mathbf{y})\mathbf{p}(\mathbf{y}) \neq 0 \implies \mathbf{g}(\mathbf{y}) \neq 0$ for ζ_τ -almost all $\mathbf{y} \in \mathbf{T}_\tau$ if and only if for any times $s > 0$ and $t \in [0, t_\delta(\gamma_s))$ such that $s + t < \tau$, we have:

$$\begin{aligned} q(\gamma_s | \gamma_s^-) \times \mathbf{P}_{\tau-s} \mathbf{I}(\gamma_{[0,s]}) \neq 0 &\implies \mathbf{q}'(\gamma_{[0,s]} | \gamma_{[0,s]}) \neq 0, \\ \lambda(\Psi(\gamma_s, t)) \times \mathbf{P}_{\tau-s-t} \mathbf{I}(\gamma_{[0,s+t]}) \neq 0 &\implies \lambda'(\Psi(\gamma_{[0,s]}, t)) \neq 0. \end{aligned} \quad (3.32)$$

This condition can be interpreted rather straightforwardly. It is possible to trigger jumps from and to states that are impossible for the distribution \mathbf{p} , provided that the expected value of \mathbf{I} given the past is nonzero before and after the jump. This implies that at any time $s \in [0, \tau)$, all transitions are still possible, independently of λ and q , as long as the event $\{\mathbf{I}(Y_{[0,\tau]}) \neq 0 \mid Y_{[0,s]}\}$ has non-zero probability. However, it can be challenging to verify this condition in practice depending on the target function \mathbf{I} . We instead choose to impose:

$$\begin{aligned} q(\gamma_s | \gamma_s^-) > 0 &\implies \mathbf{q}'(\gamma_{[0,s]} | \gamma_{[0,s]}) > 0, \\ \lambda(\Psi(\gamma_s, t)) > 0 &\implies \lambda'(\Psi(\gamma_{[0,s]}, t)) > 0. \end{aligned} \quad (3.33)$$

3.2.2 Optimal importance process

We call “optimal importance distribution”, the importance distribution that produces the IS estimator with the smallest variance. As we know, the optimal distribution admits the pdf:

$$\mathbf{g}^* : \mathbf{x} \in \mathbf{T}_\tau \mapsto \frac{1}{\mathcal{I}} \times |\mathbf{I}(\mathbf{x})| \times \mathbf{p}(\mathbf{x}). \quad (3.34)$$

Based on the foregoing, this is the distribution of a PDP on \mathcal{X} with flow Ψ . Its jump mechanism remains to be specified. It is known that it can be characterized using the optimal jump intensity λ^* and the optimal jump kernel density \mathbf{Q}^* of the associated partial trajectory.

The following theorem contains one of the main results of Thomas Galtier’s thesis [Gal19], and on which all my contributions are based. It provides the explicit expression for the optimal jump intensity λ^* and the density \mathbf{q}^* of the optimal jump kernel \mathbf{Q}^* .

THEOREM 5 · Optimal jump intensity and optimal jump kernel

Let \mathbf{p} be a PDMP distribution on a state space \mathcal{X} with local characteristics (Ψ, λ, Q) . The optimal importance distribution for the estimation of $\mathcal{I} = \mathbb{E}_{\mathbf{X} \sim \mathbf{p}}[\mathbf{I}(\mathbf{X})]$ is the distribution \mathbf{g}^* of a PDP with the same flow Ψ , an optimal jump intensity λ^* and an optimal jump kernel density \mathbf{q}^* . These optimal intensity and kernel density are given, for any partial

trajectory $x_{[0,s]}$, by:

$$\begin{aligned}\lambda^*(x_{[0,s]}) &= \lambda(x_s^-) \times \frac{\mathbf{P}_{\tau-s}^- |\mathbf{I}|(x_{[0,s]})}{\mathbf{P}_{\tau-s}^- |\mathbf{I}|(x_{[0,s]})}, \\ \mathbf{q}^*(x_{[0,s]} | x_{[0,s]}) &= q(x_s | x_s^-) \times \frac{\mathbf{P}_{\tau-s}^- |\mathbf{I}|(x_{[0,s]})}{\mathbf{P}_{\tau-s}^- |\mathbf{I}|(x_{[0,s]})}.\end{aligned}\tag{3.35}$$

The proof is given in [Chr+19, Theorem 4.3 and 4.4] or alternatively in [Gal19, Theorem 17 and 18]. We propose the following interpretation:

- (i) If the expected value of $|\mathbf{I}|$ is twice by jumping now than by not jumping, then the optimal importance jump intensity λ^* is twice the nominal jump intensity λ on the current state.
- (ii) If the expected value of $|\mathbf{I}|$ is twice by jumping to a specific state than by jumping randomly, then the optimal importance jump kernel density \mathbf{q}^* is twice the nominal jump kernel density q on this specific state.

Markovian importance process

The optimal importance process is generally not a PDMP because the optimal jump intensity and optimal jump kernel can depend on all the states of the partial trajectory. This dependence comes from the fact that the target function \mathbf{I} itself can depend on all the states of the trajectory. However, when \mathbf{I} only depends on the last state of the trajectory, the importance process can be made Markovian and thus, a PDMP.

For any bounded function $\varphi : \mathcal{X} \rightarrow \mathbb{R}$, any states x^-, x , and any time $t < \tau$, we define:

$$\begin{aligned}P_t \varphi(x) &= \mathbb{E}[\varphi(X_t) | X_0 = x], \\ P_t^- \varphi(x^-) &= Q P_t \varphi(x^-) = \mathbb{E}_{X^- \sim q(\cdot | x^-)}[P_t \varphi(X)].\end{aligned}\tag{3.36}$$

The collection $\{P_t\}_t$ is the semigroup of the PDMP of distribution \mathbf{p} . Let us assume that there exists a function $I : \mathcal{X} \rightarrow \mathbb{R}$ such that for any trajectory $x_{[0,\tau]} \in \mathbf{T}_\tau$ we have: $|\mathbf{I}(x_{[0,\tau]})| = |I(x_\tau^-)|$. For any partial trajectory $x_{[0,s]}$ we thus have:

$$\mathbf{P}_{\tau-s}^- |\mathbf{I}|(x_{[0,s]}) = \mathbb{E}[|I(X_\tau)| | X_s = x_s] = P_{\tau-s} I(x_s).\tag{3.37}$$

This last quantity depends on the state x_s but also on the elapsed time s . Now let us assume that the position space \mathcal{Z} includes time (or let us consider the time-augmented process instead of the nominal PDMP). In that case, the optimal importance process is Markovian. The optimal jump intensity and optimal jump kernel density can be reexpressed as functions of the current state

instead of the partial trajectory. For any partial trajectory $x_{[0,s]}$, we have:

$$\begin{aligned}\lambda^*(x_{[0,s]}) &= \lambda^*((x_s^-, s)) := \lambda(x_s^-) \times \frac{P_{\tau-s}^- |I|(x_s^-)}{P_{\tau-s}^- |I|(x_s^-)}, \\ \mathbf{q}^*(x_{[0,s]} | x_{[0,s]}) &= q^*(x_s | x_s^-, s) := q((x_s, s) | (x_s^-, s)) \times \frac{P_{\tau-s}^- |I|(x_s)}{P_{\tau-s}^- |I|(x_s^-)}.\end{aligned}\tag{3.38}$$

3.3 Rare event application

The contributions of this thesis focus on the case where the target function \mathbf{I} is the indicator of a rare event achievable by a PDMP trajectory. We consider this event to occur when the PDMP trajectory reaches a target region $\mathcal{F} \subset \mathcal{X}$ of its state space before a fixed time s_{\max} . From an applied perspective, the PDMP models an industrial system, and the target region \mathcal{F} determines the set of possible states corresponding to a critical failure of the system. The goal is to estimate the probability of this critical failure.

3.3.1 Committor function

For any subset $\mathcal{F} \subset \mathcal{X}$ of the state space, and any maximal duration $s_{\max} \geq 0$, we define the following stopping time⁴:

$$\tau := \inf\{t \geq 0 : X_t \in \mathcal{F}\} \wedge s_{\max}.\tag{3.39}$$

We also define \mathbf{F} as the set of trajectories that pass through \mathcal{F} before time s_{\max} :

$$\mathbf{F} := \{x_{[0,\tau]} \in \mathbf{T}_\tau \mid \exists t \in [0, s_{\max}] : x_t \in \mathcal{F}\} = \{x_{[0,\tau]} \in \mathbf{T}_\tau \mid \tau < s_{\max}\}.\tag{3.40}$$

Let I and \mathbf{I} be the two following functions:

$$\begin{aligned}I : x \in \mathcal{X} &\mapsto \mathbb{1}_{x \in \mathcal{F}}, \\ \mathbf{I} : x_{[0,\tau]} \in \mathbf{T}_\tau &\mapsto \mathbb{1}_{x_{[0,\tau]} \in \mathbf{F}} = \mathbb{1}_{\tau < s_{\max}} = I(x_\tau^-).\end{aligned}\tag{3.41}$$

We recall the notation $\mathbf{x} = x_{[0,\tau]} \in \mathbf{T}_\tau$. Our objective is to estimate:

$$\mathcal{I} := \mathbb{E}_{\mathbf{X}-\mathbf{p}}[\mathbf{I}(\mathbf{X})] = \mathbb{P}_{\mathbf{X}-\mathbf{p}}(\mathbf{X} \in \mathbf{F}),\tag{3.42}$$

where \mathbf{p} is the pdf of a nominal PDMP distribution on \mathbf{T}_τ .

The optimal importance distribution for the estimation of \mathcal{I} with importance sampling, has been described in [section 3.2.2](#). In the case of rare events, this optimal distribution can be linked to a concept from molecular dynamics: the committor function.

⁴Everything that follows remains valid by taking $\tau = s_{\max}$ a.s. provided that \mathcal{F} is an absorbing region.

Committor function in molecular dynamics

A classical sampling problem in computational physics and chemistry is simulating systems in spaces with free-energy barriers, i.e., composed of highly attractive regions separated by regions of nearly zero attraction. This is a dynamic view of a multimodal distribution. Metastability in the transition path theory is a fundamental example. A metastable region for a stochastic process is an "attractive" part of the state space where the process stays for a very long time on average before escaping. The focus is then on the probability of reaching a metastable region B before another metastable region A . In this context, the committor function associates to any state x , the probability that the stochastic process starting from x reaches region B before region A . For a fixed state x_0 very close to A , the probability to reach B before A is very small and, therefore, challenging to estimate using Monte Carlo methods. This is a domain in which methods like AMS (Adaptive Multilevel Splitting, see Section 1.2.2) are highly popular. It is well-known that the committor function serves as the optimal importance function (also referred to as the reaction coordinate in the molecular dynamics community) to use in the diffusive case [BK16; CGR19; MPP23].

Optimal importance sampling and committor function

In our context, we define the committor function as the probability that a PDMP trajectory with distribution \mathbf{p} reaches the region \mathcal{F} before the time s_{\max} given its current state.

DEFINITION 24 · Committor function

For any state $x \in \mathcal{X}$ and any time $s \geq 0$, we define:

$$\xi^*((x, s)) := P_{\tau-s} I(x) = \mathbb{P}_{\mathbf{X}-\mathbf{p}}(X_\tau \in \mathcal{F} \mid X_s = x). \quad (3.43)$$

This definition can be related to the one used in molecular dynamics for the time-augmented process. The set A corresponds to all the states $(x, s) \in \mathcal{X} \times \mathbb{R}_+$ such that $s \geq s_{\max}$, and the set B corresponds to $\mathcal{F} \times \mathbb{R}_+$. We then have:

$$\begin{aligned} \tau &= \inf\{t \geq 0 : (X_t, t) \in A \cup B\}, \\ \xi^*((x, s)) &= \mathbb{P}((X_\tau, \tau) \in B \mid X_s = x). \end{aligned} \quad (3.44)$$

We also define the variant called "edge committor function" as the probability that a trajectory of the PDMP with distribution \mathbf{p} reaches the region \mathcal{F} before the time s_{\max} given its current state and that it jumps immediately from this state.

DEFINITION 25 · Edge committor function

For any state $x^- \in \mathcal{X}$ and any time $s \geq 0$, we define:

$$\xi^{*-}((x^-, s)) := P_{\tau-s}^- I(x^-) = \mathbb{E}_{X-Q(\cdot|x^-)}[\xi^*((X, s))]. \quad (3.45)$$

We deduce a reformulation of the optimal jump intensity and the optimal jump kernel in rare event simulation for PDMPs.

PROPOSITION 15 · Optimality with committor function

For any states $x^-, x \in \mathcal{X}$ and any time $s \geq 0$, we have:

$$\begin{aligned} \lambda^*((x^-, s)) &= \lambda(x^-) \times \frac{\xi^{*-}((x^-, s))}{\xi^*((x^-, s))}, \\ q^*((x, s) | (x^-, s)) &= q(x | x^-) \times \frac{\xi^*((x, s))}{\xi^{*-}((x^-, s))}. \end{aligned} \quad (3.46)$$

This result can be interpreted as follows:

- (i) If the probability of reaching \mathcal{F} before the time s_{\max} is twice as high by jumping immediately rather than by jumping later, then λ^* is twice as high as λ on the current state.
- (ii) If the probability of reaching \mathcal{F} before the time s_{\max} is twice as high by jumping to a specific state x , given that we jump from a state x^- , than by jumping randomly, then q^* must be twice as high as q between x^- and x .

The full knowledge of the committor function would make it possible to generate trajectories according to the optimal distribution \mathbf{g}^* , and therefore to estimate the probability of interest \mathcal{I} with a single draw.

3.3.2 Practical implementation

In practice, the committor function is never known. Its behavior, however, is quite intuitive. It is often possible to propose an approximation of it based on some knowledge, even if limited, of the phenomenon modeled by the PDMP. This approximation should be simple yet flexible, allowing for refinement during iterations. We present an example based on the methodology proposed by [Chr+19].

Adaptive methodology

Suppose we have an approximation ξ_θ of the committor function, parameterized by a vector $\theta \in \mathbb{R}^{d_\theta}$. For any value of the parameter θ and any pair of states $x^-, x \in \mathcal{X}$, we set:

$$\begin{aligned}\xi_\theta^-(x^-) &= \mathbb{E}_{X \sim Q(\cdot|x^-)}[\xi_\theta(X)], \\ \lambda_\theta(x) &= \lambda(x) \times \frac{\xi_\theta^-(x)}{\xi_\theta^-(x^-)}, \\ q_\theta(x | x^-) &= q(x | x^-) \times \frac{\xi_\theta(x)}{\xi_\theta^-(x^-)}.\end{aligned}\tag{3.47}$$

For each value of θ , there corresponds a distribution with pdf \mathbf{g}_θ on \mathbf{T}_τ . For any trajectory $\mathbf{x} \in \mathbf{T}_\tau$, we have:

$$\mathbf{g}_\theta(\mathbf{x}) = \exp\left(-\int_0^{t_n} \lambda_\theta(\Psi(x_{s_n}, t)) dt\right) \prod_{k=0}^{n-1} p_{\lambda_\theta}(t_k | x_{s_k}) q_\theta(x_{s_{k+1}} | x_{s_k}^-).\tag{3.48}$$

We then have a family of importance distributions $(\mathbf{g}_\theta)_{\theta \in \Theta}$. The authors of [Chr+19] propose to sequentially determine a “good” value of the parameter θ by cross entropy. The quantity of interest \mathcal{I} can then be estimated by importance sampling using the distribution parameterized by the final value provided by the cross entropy procedure. We have described this procedure in [section I.3.2](#).

The heated room system

The method is applied to the test case of the heated room from the documentation of the tool PyCATSH00. This is the main case study of Thomas Galtier’s thesis [Gal19] and the article [Chr+19].

EXAMPLE 7 · The heated room

The temperature of a room must be maintained between a value z_{\min} and a value z_{\max} . As the room is not perfectly insulated, it is cooled by transmission of the low outside temperature z_{out} . It is equipped with three identical heaters in passive redundancy. Each heater can be active, inactive or broken, but only one can be active at a time (one is enough to heat the room in any circumstances). An active heater switches OFF automatically when the temperature reaches z_{\max} , and one inactive heater switches ON automatically when the temperature reaches z_{\min} . Heaters can break spontaneously (at a given failure rate λ_{fail}), or on demand when they are switched ON (with a given probability q_θ). A broken heater can be repaired spontaneously (at a given repair rate λ_{rep}).

The dynamics of this system are well described by a PDMP $(X_t)_t = (Z_t, V_t)_t$. At any time $t \geq 0$, the temperature in the room corresponds to the position Z_t and the status of the different heaters

corresponds to the regime $V_t = \left(V_t^{(j)}\right)_{j=1}^3 \in \{-1, 0, 1\}^3$ (with -1 for broken, 0 for inactive and 1 for active). The temperature evolves according to a deterministic flow depending on the status of the heaters:

$$\frac{dZ_t}{dt} = w_{\text{out}} \times (z_{\text{out}} - Z_t) + w_{\text{heat}} \times \mathbb{1}_{\max(V_t)=1}, \quad (3.49)$$

where w_{out} is the rate of heat transmission between the room and the outside, and w_{heat} is the heating power of each heaters. One active heater is enough to heat the room to temperature z_{max} if and only if $w_{\text{heat}} > w_{\text{out}}(z_{\text{max}} - z_{\text{out}})$.

There are jumps at the boundaries of $[z_{\text{min}}, z_{\text{max}}]$. If the temperature reaches z_{max} , then the active radiator becomes inactive with probability 1 . If the temperature falls below z_{min} , an attempt is made to activate a non-broken radiator. Each attempt results in either the failure on demand of the solicited radiator with probability q_{b} , or its activation (with probability $1 - q_{\text{b}}$). The activation process stops as soon as a radiator is active or when all radiators are broken.

For all $j \in \llbracket 1, 3 \rrbracket$, the heater j has a marginal jump intensity $\lambda^{(j)}$ that depends on its status $v^{(j)}$ and the temperature z . For any $x = (z, v)$ with $v = \left(v^{(j)}\right)_{j=1}^3$, we have:

$$\lambda^{(j)}(x) \mapsto \lambda_{\text{fail}}(z) \mathbb{1}_{v^{(j)} > -1} + \lambda_{\text{rep}}(z) \mathbb{1}_{v^{(j)} = -1}. \quad (3.50)$$

The jump intensity λ corresponds to the sum of marginal jump intensities:

$$\lambda : x \mapsto \sum_{j=1}^3 \lambda^{(j)}(x) = \lambda_{\text{fail}}(z) \sum_{j=1}^3 \mathbb{1}_{v^{(j)} > -1} + \lambda_{\text{rep}}(z) \sum_{j=1}^3 \mathbb{1}_{v^{(j)} = -1}. \quad (3.51)$$

A spontaneous jump from state x corresponds to a change in the status of heater j with probability $\lambda^{(j)}(x)/\lambda(x)$. An activation attempt is made after each jump if the temperature is less than or equal to z_{min} . The system parameters are given in [table 3.1](#).

Numerical results

We are interested in the probability that the room temperature falls below a threshold $z_{\mathcal{F}} < z_{\text{min}}$ before a given time s_{max} . This can only happen if all three radiators remain failed, simultaneously, for a long enough time. It is clear that the states closest to the failure region $\mathcal{F} = \{(z, v) \in \mathcal{X} : z \leq z_{\mathcal{F}}\}$ are those for which multiple radiators are broken. The authors of [\[Chr+19\]](#) propose the following approximation of the committor function:

$$\xi_{\mathcal{G}}((z, v)) = \exp \left[\left(\theta_1 \mathbb{1}_{z > z_{\text{min}}} + \theta_2 \mathbb{1}_{z \leq z_{\text{min}}} \right) \left(\sum_{i=1}^3 \mathbb{1}_{v_i = -1} \right)^2 \right]. \quad (3.52)$$

This approximation is therefore a function of the number of failed radiators. It is parameterized by two real numbers θ_1 and θ_2 . Each one allows to control how much the failures should be

TABLE 3.1 **Parameters** of the heated room. Values taken from [Chr+19].

Parameter	Value	Description
z_{\min}	0.5	Minimal temperature before switching ON a heater
z_{\max}	5.5	Maximal temperature before switching OFF a heater
z_{out}	-1.5	Outside temperature
w_{out}	0.1	Rate of heat transition with exterior
w_{heat}	5	Heating power of each heater
q_{∂}	0.01	Probability of failure on demand
$\lambda_{\text{fail}}(z)$	$0.0021 + 0.00015z$	Failure rate of each heater
$\lambda_{\text{rep}}(z)$	0.2	Failure rate of each heater
s_{\max}	100	Duration of the mission
z_0	7.5	Initial temperature in the room
$z_{\mathcal{F}}$	0	Threshold of critical failure

avored and the repairs hindered. The first parameter θ_1 is used when the transition does not involve the activation of a radiator and therefore corresponds to spontaneous failures and repairs. The second parameter θ_2 is used when the transition requires the activation of a radiator and interacts with the probability of failure on demand.

We now present the experiment they carried out and the results they obtained. The parameters of the flow, the failure and repair rates, and the probability of failure on demand are those given in table 3.1. The sample size used at each iteration of the cross-entropy procedure is such that each sample contains $n_{\text{CE}} = 100$ failing trajectories. Three cross-entropy iterations were performed (with respective sample sizes of 1970, 126, 127). The method is initialized with $\theta^{(0)} = (0.5, 0.5)$. The cross-entropy procedure returned at the third iteration the parameter value $\theta^{(3)} \approx (0.915, 1.197)$. Importance sampling (IS) with distribution $\mathbf{g}_{\theta^{(3)}}$ is compared to a classic Monte Carlo (CMC) method in table 3.2 for different sample sizes n .

TABLE 3.2 **Comparison of the classic Monte Carlo (CMC) method with an importance sampling method for different sample sizes n . Results obtained by the authors of [Chr+19].**

	n	$\hat{\mathcal{I}} \times 10^5$	$\hat{\sigma}^2/n$	95% $\widehat{\text{CI}} \times 10^5$
CMC	10^6	0.4	4.00×10^{-12}	[0.01, 0.79]
	10^7	1.3	1.28×10^{-12}	[1.07, 1.51]
IS	10^3	1.28	4.37×10^{-13}	[1.15, 1.41]
	10^4	1.273	5.07×10^{-14}	[1.228, 1.317]
	10^5	1.289	5.07×10^{-15}	[1.275, 1.303]
	10^6	1.288	5.05×10^{-16}	[1.283, 1.292]

CONTRIBUTIONS

CHAPTER 4

Adaptive importance sampling strategy

The optimality conditions of an importance distribution for importance sampling of PDMPs were recalled in [chapter 3](#). We are interested in the case of rare event simulation realized by PDMPs. The optimal importance distribution is then characterized by an optimal jump intensity and optimal jump kernel, which are given by the committor function of the process. This committor function associates to each possible state of a PDMP trajectory the probability that it realizes the event of interest. We first introduce the general form of a family of approximations of the committor function. We then detail an adaptive importance sampling procedure with recycling to jointly determine a good candidate within the family of approximations, and to estimate the probability of interest. We prove the asymptotic optimality and normality of the method under easily verifiable and interpretable conditions. Finally, we discuss implementation choices of the method.

Content of this chapter

4.1	Towards committor function approximation	88
4.1.1	Reminder on importance sampling of PDMP	88
4.1.2	Problem parameterization	89
4.2	General parametric form of the importance function	91
4.2.1	Parametric form of the IF	91
4.2.2	Proximity scores	93
4.2.3	Basis functions	94
4.3	Adaptive importance sampling with recycling scheme	95
4.3.1	Cross entropy with recycling scheme	95
4.3.2	Asymptotic properties	97
4.3.3	Implementation guidelines	103

4.1 Towards committor function approximation

Before delving into our contributions, here is a summary of the concepts you will need from the [chapter 3](#).

4.1.1 Reminder on importance sampling of PDMP

Let \mathcal{X} be the state space of a PDMP with local characteristics (Ψ, λ, Q) . Let also \mathcal{F} be a target region of \mathcal{X} and $s_{\max} > 0$ the maximum considered duration of a PDMP trajectory. For any PDMP trajectory (X_t) , we define the stopping time:

$$\tau := \inf\{t \geq 0 : X_t \in \mathcal{F}\} \wedge s_{\max}. \quad (4.1)$$

We denote by \mathbf{T}_τ the space of PDP trajectories of duration τ with flow Ψ on \mathcal{X} . Let also \mathbf{p} be the pdf of a reference distribution of a PDMP trajectory (characterized by (Ψ, λ, Q)) in \mathbf{T}_τ . For the following, we note $\mathbf{x} = x_{[0,\tau]} \in \mathbf{T}_\tau$ a PDMP trajectory. Finally, let \mathbf{F} be the set of trajectories that reach \mathcal{F} before time s_{\max} :

$$\mathbf{F} := \{x_{[0,\tau]} \in \mathbf{T}_\tau \mid \exists t \in [0, s_{\max}) : x_t \in \mathcal{F}\} = \{\mathbf{x} \in \mathbf{T}_\tau \mid \tau < s_{\max}\}. \quad (4.2)$$

We aim to estimate the integral

$$\mathcal{I} := \mathbb{E}_{\mathbf{p}}[\mathbf{I}(\mathbf{X})], \quad (4.3)$$

in the rare event case

$$\mathbf{I} : \mathbf{x} \in \mathbf{T}_\tau \mapsto \mathbf{1}_{\mathbf{x} \in \mathbf{F}} = \mathbf{1}_{\tau < s_{\max}} = \mathbf{1}_{x_\tau \in \mathcal{F}}. \quad (4.4)$$

Thus $\mathcal{I} = \mathbb{P}_{\mathbf{X}-\mathbf{p}}(\mathbf{X} \in \mathbf{F})$. It is the probability that a trajectory of a PDMP with pdf \mathbf{p} reaches a region \mathcal{F} of its state space before a specified time s_{\max} . As previously discussed, on a given state space \mathcal{X} and under a fixed deterministic flow Ψ , the distribution of a PDMP is characterized by its jump intensity λ and jump kernel Q (or by the probability density function q of the kernel). The optimal distribution for importance sampling (yielding an estimator with zero variance of the probability of interest) is characterized by an optimal jump intensity λ^* and an optimal jump kernel Q^* of pdf q^* .

Committor function

The forms of the optimal jump intensity and kernel are known and admit explicit expressions based on the committor function of the process. Recall that the committor function represents the probability that the trajectory realizes the event of interest before the time s_{\max} given its current state. Additionally, we define the edge committor function, which represents the probability that the trajectory realizes the event of interest before the time s_{\max} given that the

process is about to jump from its current state.

$$\xi^*(x, s) = \mathbb{E}_{\mathbf{p}}[\mathbf{I}(\mathbf{X}) \mid X_s = x] = \mathbb{P}_{\mathbf{p}}(\mathbf{X} \in \mathbf{F} \mid X_s = x), \quad (4.5)$$

$$\xi^{*-}(x^-, s) = \mathbb{E}_{X \sim Q(\cdot \mid x^-)}[\xi^*(X, s)] = \int_{\mathcal{X}} \xi^*(x, s) q(x \mid x^-) \nu_q(dx \mid x^-). \quad (4.6)$$

The expressions for the optimal jump intensity and optimal jump kernel are given by:

$$\lambda^*((x, s)) = \lambda(x) \frac{\xi^{*-}(x, s)}{\xi^*(x, s)}, \quad (4.7)$$

$$q^*((x, s) \mid (x^-, s)) = q(x \mid x^-) \frac{\xi^*(x, s)}{\xi^{*-}(x^-, s)}. \quad (4.8)$$

Hence, knowing the committor function ξ^* is sufficient to completely characterize the optimal importance distribution. Note that we even have $\mathcal{I} = \xi^*(x_0, 0)$. However, the committor function is not explicitly known. The idea, as discussed in the preceding chapter, is to replace the committor function ξ^* with an approximation. Instead of searching for new jump intensity and new jump kernel among the set of all possible ones, we considerably reduce the size of the problem by focusing on finding a good approximation of the committor function. The interpretation of such a function with respect to the rare event is much clearer than that of an alternative intensity and kernel, making it also more intuitive to determine good candidates.

4.1.2 Problem parameterization

A good approximation of the committor function should, in some sense, quantify the proximity between the current state of the trajectory and the region of interest \mathcal{F} . A function fulfilling this role is sometimes referred to in the literature as an “importance function” (IF) [Vilo7; BS20; Lou+17], “reaction coordinate” (RC) [MPP23; Bit+23; Kri13; BHo5], or “collective variable” (CV) [Hol+24; FKH13; Che21; Bha22]. We are interested in this function not for what it would teach us about the process but solely as fuel for an importance sampling method. Hence, we will preferentially use the term IF in the following¹. Saying this also implies that we are not concerned with the faithfulness of an IF to the true committor function ξ^* but rather with the performance of the importance distribution that can be associated with it.

¹Furthermore, RC and CV are terms with a stronger “molecular dynamics” connotation, which is not the application we have in mind.

Parametric family of importance distributions

In practice, we consider a parametric family of importance functions (IFs) denoted by $(\xi_\theta)_{\theta \in \Theta}$, where $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$ is a vector parameter. Instead of learning an IF that closely mimics the behavior of ξ^* , we select the best corresponding importance distribution using a cross-entropy method (see [section 1.3.2](#)).

The distribution \mathbf{g}_θ corresponding to an IF ξ_θ is characterized by the jump intensity λ_θ and the jump kernel q_θ , obtained by replacing ξ^* with ξ_θ in [eqs. \(4.6\) to \(4.8\)](#). For any states $x, x^- \in \mathcal{X}$, and any time $s, t \geq 0$, we have:

$$\xi_\theta^-(x^-, s) = \mathbb{E}_{X^- \sim Q(\cdot | x^-)} [\xi_\theta(X, s)] = \int_{\mathcal{X}} \xi_\theta(x, s) q(x | x^-) \nu_q(dx | x^-) \quad (4.9)$$

$$\lambda_\theta(x) = \lambda(x) \frac{\xi_\theta^-(x, s)}{\xi_\theta(x, s)}, \quad (4.10)$$

$$q_\theta(x | x^-) = q(x | x^-) \frac{\xi_\theta(x, s)}{\xi_\theta^-(x^-, s)}, \quad (4.11)$$

$$p_{\lambda_\theta}(t | x) = [\lambda_\theta(\Psi(x, t))]^{1_{t < t_\theta(x)}} \exp\left[-\int_0^t \lambda_\theta(\Psi(x, u)) du\right]. \quad (4.12)$$

For a given family of importance functions $(\xi_\theta)_{\theta \in \Theta}$, we deduce the family of importance distributions $(\mathbf{g}_\theta)_{\theta \in \Theta}$, characterized by the p.d.f. given in [eq. \(4.13\)](#). For any trajectory $\mathbf{x} \in \mathbf{T}_\tau$ with n jumps and whose skeleton is $[x_{s_k}, t_k]_{k=0}^n$, we have:

$$\mathbf{g}_\theta(\mathbf{x}) = \exp\left(-\int_0^{t_n} \lambda_\theta(\Psi(x_n, t)) dt\right) \prod_{k=0}^n p_{\lambda_\theta}(t_k | x_{s_k}) \times q_\theta(x_{s_{k+1}} | \Psi(x_{s_k}, t_k)). \quad (4.13)$$

Practical considerations

For most of the phenomena modelled by PDMPs, the number of possible transitions is finite from any departure state. The integral ξ_θ^- is then a simple finite sum. In particular, jumping to a new state often modifies the regime only and not the position (which only moves with the flow). The committor function then only appears in the expressions for the optimal jump intensity [eq. \(4.7\)](#) and the optimal jump kernel [eq. \(4.8\)](#) as ratios evaluated at identical times and positions but distinct regimes, which removes some of the position and time dependence. Since what happens in a given regime is deterministic, achieving the event of interest depends more on the regimes to which the process jumps than on the position and time. It remains relevant, and perhaps more reasonable, to restrict ourselves to IFs that depend only on the regime and not on the position or time.

Since the committor function only appears as ratios in [eq. \(4.7\)](#) and [eq. \(4.8\)](#), we only need a good approximation up to a multiplicative constant. Therefore, the IFs do not need to be

normalized and interpreted as probabilities.

4.2 General parametric form of the importance function

In this section, we propose a general form of importance function that depends only on the PDMP regime. A possible equivalent of the committor function but restricted to the regime could be the following function:

$$v \in \mathcal{V} \mapsto \int_0^{s_{\max}} \mathbb{E}_{\mathbf{p}}[\xi^*(X_s, s) \mid V_s = v] ds. \quad (4.14)$$

It is clear that this function takes its largest values when the target region \mathcal{F} is accessible by staying in regime v for a sufficiently long time. The integrand then corresponds roughly to the probability of not leaving such a regime until \mathcal{F} is reached.

DEFINITION 26 · Critical regimes

We define the set $\mathcal{V}_{\mathcal{F}}$ of critical regimes that lead to reach the region of interest \mathcal{F} .

$$\mathcal{V}_{\mathcal{F}} := \{v \in \mathcal{V} \mid \exists z \in \mathcal{Z} : x = (z, v) \in \mathcal{F}\}. \quad (4.15)$$

It may happen that some states in \mathcal{F} are not accessible to a trajectory \mathbf{X} of distribution \mathbf{p} . In this case, it is more cautious to consider the following definition (although more difficult to verify in practice):

$$\mathcal{V}_{\mathcal{F}}^* := \{v \in \mathcal{V} \mid \exists t \in [0, s_{\max}] : \mathbb{P}_{\mathbf{p}}(X_t \in \mathcal{F} \mid V_t = v) > 0\}. \quad (4.16)$$

It is necessary to be in $\mathcal{V}_{\mathcal{F}}$ to end up in \mathcal{F} . Therefore, the committor function is large for a given regime v if the set $\mathcal{V}_{\mathcal{F}}$ can be reached from v in a small number of not too unlikely jumps. We now consider reaching $\mathcal{V}_{\mathcal{F}}$ as the main obstacle to overcome in order to realize the rare event. We are thus looking for IFs that can quantify the proximity of a regime to the set $\mathcal{V}_{\mathcal{F}}$ ².

4.2.1 Parametric form of the IF

Since we are looking for ξ_{ϑ} only among IFs depending solely on the regime, we have only a countable (and most often finite) number of values to choose from. In the case where the number of regimes $d_{\mathcal{V}} = \text{Card}(\mathcal{V})$ is finite, it suffices to define $\xi_{\vartheta}(x) = \theta_v$ for all $x = (z, v)$ with $d_{\Theta} = d_{\mathcal{V}}$ to express the set of IFs of interest. Unfortunately, the number of regimes $d_{\mathcal{V}}$ is

²Note that $\mathcal{V}_{\mathcal{F}}$ does not necessarily represent a minority of existing regimes. When \mathcal{F} corresponds to the critical failure of large industrial systems (a case that we will address in [chapter 5](#)), most regimes belong to $\mathcal{V}_{\mathcal{F}}$. Indeed, there are many more possible configurations in which the system does not work than configurations in which it works correctly. However, these are regimes that are rarely visited or for durations that are too short to reach \mathcal{F} .

often too large in interesting cases for such a family of IFs to be explorable. We need to restrict ourselves to families of IFs parameterized by vectors of smaller dimension. While the family $(\xi_\theta)_{\theta \in \Theta}$ may then lack flexibility, it becomes easier to determine one of its best representatives. We want the form of our IF to allow easy adjustment of the dimension d_Θ of its parameter vector θ . The key idea of our approach is to introduce a minimal amount of information about the PDMP to guide the IF. This a priori information (i.e., before any simulation) takes the form of a score function quantifying the proximity between a given regime and the region \mathcal{V}_F . The IF we propose is a parametric transformation of this proximity score.

The proximity score can be very rudimentary. Once again, the goal is not to construct the most faithful approximation of the committor function. Only the performance of the importance distribution associated with the IF matters. Introducing a priori information about the behavior of the PDMP helps to avoid over-parameterizing a family of IFs. We no longer seek the best function of the regime (a nearly impossible task without specifying links between the different regimes), but the best function of the score (a unidimensional variable in the interval $[0, 1]$).

DEFINITION 27 · Parametric form of the IF

For any state $x = (z, v) \in \mathcal{X}$ and any time $s \geq 0$ we introduce:

$$\xi_\theta(x, s) = \exp \left[\sum_{j=1}^{d_\Theta} \theta_j \times \phi_j(\beta(v)) \right]. \quad (4.17)$$

This form is defined by:

- (i) the vector parameter $\theta \in \Theta \subset \mathbb{R}^{d_\Theta}$ with dimension $d_\Theta > 0$,
- (ii) the proximity score $\beta : \mathcal{V} \mapsto [0, 1]$,
- (iii) a family of basis functions $(\phi_j)_{j=1}^{d_\Theta}$ with $\phi_j : [0, 1] \mapsto \mathbb{R}$ for any $j \in \llbracket 1, d_\Theta \rrbracket$.

The flexibility of the IF is determined by the dimension of the vector θ , as this determines the number of basis functions used to transform the proximity score β .

Note, that the edge committor function ξ_θ^{*-} depends on the jump kernel q , which is a function of the complete state of the process and not just its regime. Except for special cases (where the destination of a jump depends only on the departure regime), the approximation ξ_θ^- is therefore not solely a function of the regime, even when ξ_θ is. On the other hand, the edge committor function does not depend on the elapsed time. Therefore, we will allow ourselves the notational abuse of using $\xi_\theta(x)$ and $\xi_\theta^-(x)$ in the following rather than $\xi_\theta(x, s)$ and $\xi_\theta^-(x, s)$.

Eligibility

The exponential ensures that $\xi_\theta > 0$ and thus that λ_θ and q_θ are non-negative (thus valid jump intensities and kernels). It is then observed that q_θ remains normalized to 1 (independently of whether q is normalized):

$$\int_{\mathcal{X}} q_\theta(x | x^-) \nu_q(dx | x^-) = \frac{\int_{\mathcal{X}} q(x | x^-) \xi_\theta(x) \nu_q(dx | x^-)}{\int_{\mathcal{X}} q(x' | x^-) \xi_\theta(x') \nu_q(dx' | x^-)} = 1. \quad (4.18)$$

Another desirable property of the IF is to prevent the explosion of the importance process. It suffices to observe that $\inf_x \xi_\theta(x) > 0$ and $\sup_x \xi_\theta(x) < \infty$ when the parameter θ , the score function β , and the family of basis functions (ϕ_j) are fixed. We then note that the ratio $\frac{\xi_\theta^-}{\xi_\theta}$ is upper-bounded, which ensures that λ_θ is bounded if λ is (see [proposition 10](#)).

Including the reference distribution \mathbf{p} within the family of importance distributions $(\mathbf{g}_\theta)_{\theta \in \Theta}$ is a useful precaution (ensuring that one can always find an importance distribution that performs no worse than classical Monte Carlo). Let $\mathbf{0}_{d_\Theta} = (0, \dots, 0) \in \mathbb{R}^{d_\Theta}$. We have $\xi_{\mathbf{0}_{d_\Theta}}(x) = 1$ for all $x \in \mathcal{X}$, hence $\lambda_{\mathbf{0}_{d_\Theta}} = \lambda$, $q_{\mathbf{0}_{d_\Theta}} = q$, and thus $\mathbf{g}_{\mathbf{0}_{d_\Theta}} = \mathbf{p}$.

4.2.2 Proximity scores

As we have said, it is not reasonable to explore all the possible values that can be assigned to each regime. Nevertheless, we will see in the [chapters 5](#) and [6](#) that it is possible to assign an initial value (even a very naive one) to the different regimes based on the practitioner's a priori idea of their proximity to \mathcal{V}_F . For each regime $v \in \mathcal{V}$, we denote $\beta(v) \in [0, 1]$ as its proximity score to the set \mathcal{V}_F . The higher the score $\beta(v)$, the more we assume that v is "close" to \mathcal{V}_F . We impose the following condition: $\beta(v) = 1 \iff v \in \mathcal{V}_F$. This proximity score is transformed by our IF, whose parameters will be calibrated during the simulation and optimization phases of the cross-entropy method, to associate the most relevant possible value to each regime.

This is one of the main messages of this thesis. As tempting as it may be, using totally agnostic distributions (today conjugated with deep learning methods in generative models) to do multimodal importance sampling in high dimension is not always the most relevant approach. Very good results can be obtained by introducing a minimal amount of information into a parametric model. We will see that this is a small price to pay in the context of PDMPs because the relevant information can be extracted without an in-depth knowledge of the problem's specificities. Our approach aims to clearly separate the information that enters the model, completely characterized by the proximity score β , from the rest of our methodology, which is truly "generalist".

Let us now focus on the form of our IF. The proximity score β is assumed to be fixed and so are the scores $\beta(v)$ for all $v \in \mathcal{V}$. They define the maximum degree of expressiveness of our IF (up to

d_ν different values if all score values $\beta(v)$ are distinct). It remains to determine how to express a wide variety of functions of $\beta \in [0, 1]$ with a limited number of parameters.

4.2.3 Basis functions

We first write $\xi_\theta = \exp(-U_\theta)$ where U_θ is a function of β , which is zero when θ is identically zero³. The exponential ensures both the positivity of the IF (without restricting the sign of U_θ) and that the IF is equal to 1 for $\theta = \mathbf{0}_{d_\Theta}$. Furthermore, it facilitates optimization by further convexifying the problem⁴. Finally, we construct U_θ by drawing inspiration from truncated projection methods on L^2 for the approximation of functions from $[0, 1]$ to \mathbb{R} . The function $U_\theta = -\sum_{j=1}^{d_\Theta} \theta_j \times \phi_j(\beta)$ is a finite linear combination of basis functions ϕ_j in β with θ_j as coefficients.

There is no shortage of options for the basis functions ϕ_j (starting with the set of orthonormal bases of $L^2([0, 1])$). We have been guided by the two following principles:

- (i) There must always exist combinations of parameters θ allowing the IF to be strictly increasing in β . This is the expected behavior of the IF if the proximity score β has not been too poorly chosen.
- (ii) Increasing the dimension of the parameter vector should generally allow for a refined approximation. The IF should be able to coincide with any step function of β at the points $(\beta(v))_{v \in \mathcal{V}}$ when the dimension d_Θ is large enough.

We propose and discuss two families of basis functions: monomials and piecewise linear functions.

Polynomial

For any $j \in \llbracket 1, d_\Theta \rrbracket$ and any $\beta \in [0, 1]$ we define:

$$\phi_j(\beta) = \beta^j. \quad (4.19)$$

The IF is then simply the exponential of a polynomial of degree d_Θ . Other polynomial bases than monomials can be used, such as Bernstein polynomials: $\phi_j(\beta) = \binom{d_\Theta}{j} \beta^j (1 - \beta)^{d_\Theta - j}$. By associating a polynomial of degree d_Θ with each parameter θ_j , we avoid the situation where the IF is driven by only a few θ_j as the degree increases. Increasing the degree allows for the

³This form inevitably recalls what a Boltzmann-Gibbs distribution with the potential $U_\theta(\beta) = -\sum_{j=1}^{d_\Theta} \theta_j \times \phi_j(\beta)$ would be, but this is not a path we pursue here.

⁴Let $\phi_j = (\phi_j)_{j=1}^{d_\Theta}$ be different from the zero vector. The Hessian of ξ_θ at θ is given by $\xi_\theta \times \phi \cdot \phi^T$, which is positive definite, while the Hessian of U_θ at θ is zero.

approximation of any function in $L^2([0, 1])$ without the issues encountered in interpolation (i.e., U_β is not constrained to pass through specific points but is adjusted to minimize a loss).

High-degree polynomials are particularly effective for approximating multimodal functions. They are likely useful for correcting poorly chosen proximity score β and allow great global freedom to U_β , including the freedom to not increase everywhere in β . However, they are less suitable for finely approximating the slope of “more monotonic” functions and offer less local freedom to U_β .

Piecewise linear functions

Let $\mathbf{c} = (c_0, \dots, c_{d_\Theta}) \in [0, 1]^{d_\Theta}$ such that $0 = c_0 < c_1 < \dots < c_{d_\Theta} = 1$. For any $j \in \llbracket 1, d_\Theta \rrbracket$, and any $\beta \in [0, 1]$ we define:

$$\phi_j(\beta) = (\beta - c_j) \mathbb{1}_{\beta \geq c_j}. \quad (4.20)$$

At a fixed β , the supports $[c_j, +\infty)$ for $j \in \llbracket 1, d_\Theta \rrbracket$ are overlapping to allow for the continuity of U_β with respect to β . This avoids threshold effects that would occur at points c_j with a support of the form $\{\beta \in [c_j, c_{j+1})\}$. Here, unlike polynomial-based IFs, the ϕ_j have a local influence and allow for fine-tuning U_β when d_Θ is high.

It can be noted that the $\beta(v)$ values are not necessarily evenly distributed over $[0, 1]$. Thus, we do not recommend to choose the $(c_j)_j$ evenly spaced. A first simple alternative is to choose \mathbf{c} such that each interval $[c_j, c_{j+1}]$ contains the same number of elements of $(\beta(v))_{v \in \mathcal{V}}$. We suggest another option: building d_Θ classes on $[0, 1]$ using a 1D k -means method.

4.3 Adaptive importance sampling with recycling scheme

Let us consider, for the remainder of this chapter, a family of proximity scores $(\beta_v)_{v \in \mathcal{V}}$ (whose construction is addressed in [chapter 5](#) and [chapter 6](#)) and a family of basis functions $(\phi_j)_j$. Consequently, we have a family of importance functions (IFs) $(\xi_\theta)_{\theta \in \Theta}$ (as defined in [definition 27](#)), along with an associated family of importance distributions $(\mathbf{g}_\theta)_{\theta \in \Theta}$ (characterized by a p.d.f. whose expression is given in [eq. \(4.13\)](#)). We now present the adaptive importance sampling method upon which we rely to both sequentially determine a suitable candidate importance distribution and to estimate the probability of the event of interest. We also examine its asymptotic properties.

4.3.1 Cross entropy with recycling scheme

We rely on a cross-entropy method as outlined in [section 1.3.2](#) to sequentially improve the importance distribution. We briefly recall the principle of the method:

- ◊ We seek the distribution \mathbf{g}_θ that minimizes the forward Kullback-Leibler (KL) divergence with the optimal distribution \mathbf{g}^* within the family $(\mathbf{g}_\theta)_{\theta \in \Theta}$.

$$\theta^* \in \arg \min_{\theta \in \Theta} \mathcal{D}_{\text{KL}}(\mathbf{g}^* \parallel \mathbf{g}_\theta). \quad (4.21)$$

- ◊ It can be demonstrated (see eq. (1.55)) that this is equivalent to determining:

$$\theta^* \in \arg \max_{\theta \in \Theta} \mathbb{E}_{\mathbf{p}}[\mathbf{I}(\mathbf{X}) \log \mathbf{g}_\theta(\mathbf{X})]. \quad (4.22)$$

- ◊ We do not directly maximize this expectation but rather an iterative importance sampling approximation. Each iteration ℓ is decomposed into two steps: first, we simulate a sample $\mathbf{X}_1, \dots, \mathbf{X}_{n_\ell}$ according to a distribution $\mathbf{g}_{\theta^{(\ell)}}$, and then we determine the parameter for the next iteration $\theta^{(\ell+1)}$ that maximizes an estimate of $\mathbb{E}_{\mathbf{p}}[\mathbf{I}(\mathbf{X}) \log \mathbf{g}_\theta(\mathbf{X})]$ based on the sample $\mathbf{X}_1, \dots, \mathbf{X}_{n_\ell}$.
- ◊ After several iterations, assuming we have obtained a good importance distribution, we then simulate a final sample according to this distribution to construct the importance sampling estimator of the probability of interest.

We summarize this version of the method in [algorithm 5](#). Implementation details such as initialization, termination criterion, or the size of each sample will be discussed in [section 4.3.3](#).

Recycling scheme

The method described in [algorithm 5](#) is simple to implement, analyze, and converges quickly in practice⁵. However, it has some drawbacks. Only the last drawn sample contributes to the estimation of the probability of interest. We do not take advantage of the realizations generated throughout the iterations. The importance distribution also cannot be updated too frequently because the cross entropy loss has to be estimated with a large sample size at each iteration.

The alternative is to recycle realizations from previous iterations at each optimization step and at the final estimation step. Recycling an observation $\mathbf{X}_k^{(\ell)}$ generated at iteration ℓ simply requires associating it with the proper likelihood ratio $\mathbf{p}/\mathbf{g}_{\theta^{(\ell)}}$ (with $\mathbf{g}_{\theta^{(\ell)}}$ being the distribution that generated the observation). This leads us to [algorithm 6](#).

⁵A good importance distribution is usually obtained in 4 or 5 iterations.

ALGORITHM 5 · Vanilla cross entropy for rare events

1 **Initialization** with parameter $\theta^{(1)} \in \Theta$.

2 **for** $\ell = 1, \dots, L - 1$ **do**

3 **Sampling step**

$$\mathbf{X}_1^{(\ell)}, \dots, \mathbf{X}_{n_\ell}^{(\ell)} \stackrel{\text{i.i.d.}}{\sim} \mathbf{g}_{\theta^{(\ell)}}. \quad (4.23)$$

4 **Optimization step**

$$\theta^{(\ell+1)} \in \arg \min_{\theta \in \Theta} - \sum_{k=1}^{n_\ell} \mathbf{I}(\mathbf{X}_k^{(\ell)}) \frac{\mathbf{P}(\mathbf{X}_k^{(\ell)})}{\mathbf{g}_{\theta^{(\ell)}}(\mathbf{X}_k^{(\ell)})} \log \mathbf{g}_\theta(\mathbf{X}_k^{(\ell)}). \quad (4.24)$$

5 **end**

6 **Final sampling step**

$$\mathbf{X}_1^{(L)}, \dots, \mathbf{X}_{n_L}^{(L)} \stackrel{\text{i.i.d.}}{\sim} \mathbf{g}_{\theta^{(L)}} \quad (4.25)$$

7 **Estimation step**

$$\hat{\mathcal{I}}^{\text{CE}} = \frac{1}{n_L} \sum_{k=1}^{n_L} \mathbf{I}(\mathbf{X}_k^{(L)}) \frac{\mathbf{P}(\mathbf{X}_k^{(L)})}{\mathbf{g}_{\theta^{(L)}}(\mathbf{X}_k^{(L)})}. \quad (4.26)$$

4.3.2 Asymptotic properties

Using a scheme to recycle past samples necessarily complicates the study of the asymptotic properties of the method. The data used are indeed no longer all from the same distribution (since it changes between iterations), and above all are not independent since the distribution of the data at a given iteration is constructed from the data generated in previous iterations.

The recycling scheme presented above corresponds to the one found in the article [DP18]. In this article, the authors rely on a martingale property of the cross-entropy loss estimator and of the final estimator of the quantity of interest to demonstrate the consistency of the algorithm [DP18, Theorem 2] and its asymptotic normality [DP18, Theorem 3]. We restate below the conditions of their theorem in the context of our work with our notations⁶.

⁶These results are valid for any family of importance distributions $(\mathbf{g}_\theta)_{\theta \in \Theta}$ satisfying the conditions of theorems 6 and 7. We are therefore not restricted to the case of families of distributions constructed from a family of IFs, which is simply the approach we propose but not the only possible one.

ALGORITHM 6 · Cross entropy with recycling scheme

1 **Initialization** with parameter $\theta^{(1)} \in \Theta$.

2 **for** $\ell = 1, \dots, L - 1$ **do**

3 **Sampling step**

$$\mathbf{X}_1^{(\ell)}, \dots, \mathbf{X}_{n_\ell}^{(\ell)} \stackrel{\text{i.i.d.}}{\sim} \mathbf{g}_{\theta^{(\ell)}}. \quad (4.27)$$

4 **Optimization step**

$$\theta^{(\ell+1)} \in \arg \min_{\theta \in \Theta} - \sum_{m=1}^{\ell} \sum_{k=1}^{n_m} \mathbf{I}(\mathbf{X}_k^{(m)}) \frac{\mathbf{p}(\mathbf{X}_k^{(m)})}{\mathbf{g}_{\theta^{(m)}}(\mathbf{X}_k^{(m)})} \log \mathbf{g}_{\theta}(\mathbf{X}_k^{(m)}). \quad (4.28)$$

5 **end**

6 **Final sampling step**

$$\mathbf{X}_1^{(L)}, \dots, \mathbf{X}_{n_L}^{(L)} \stackrel{\text{i.i.d.}}{\sim} \mathbf{g}_{\theta^{(L)}} \quad (4.29)$$

7 **Estimation step**

$$\hat{\mathcal{I}}^{\text{RCE}} = \frac{1}{\sum_{\ell=1}^L n_\ell} \sum_{\ell=1}^L \sum_{k=1}^{n_\ell} \mathbf{I}(\mathbf{X}_k^{(\ell)}) \frac{\mathbf{p}(\mathbf{X}_k^{(\ell)})}{\mathbf{g}_{\theta^{(\ell)}}(\mathbf{X}_k^{(\ell)})}. \quad (4.30)$$

THEOREM 6 · Theorem 2 from [DP18]

Assume that:

- ◇ $\Theta \subset \mathbb{R}^{d_\theta}$ is a compact set.
- ◇ For any $\mathbf{x} \in \text{supp}(\mathbf{I}_p)$, the integrand of the cross entropy loss $\theta \mapsto -\mathbf{I}(\mathbf{x}) \log \mathbf{g}_\theta(\mathbf{x})$ is a continuous function of θ .
- ◇ The cross entropy loss is always finite:

$$\mathbb{E}_p \left[\mathbf{I}(\mathbf{X}) \sup_{\theta \in \Theta} \{-\log \mathbf{g}_\theta(\mathbf{X})\} \right] < \infty. \quad (4.31)$$

- ◇ θ^* is the unique minimizer of the cross entropy loss, i.e. for any $\theta \in \Theta \setminus \{\theta^*\}$, we have:

$$\mathbb{E}_p [-\mathbf{I}(\mathbf{X}) \log \mathbf{g}_{\theta^*}(\mathbf{X})] < \mathbb{E}_p [-\mathbf{I}(\mathbf{X}) \log \mathbf{g}_\theta(\mathbf{X})]. \quad (4.32)$$

- ◇ The following quantity is bounded:

$$\sup_{\theta \in \Theta} \mathbb{E}_{\mathbf{P}} \left[\mathbf{I}(\mathbf{X}) \frac{\mathbf{p}(\mathbf{X})}{\mathbf{g}_{\theta}(\mathbf{X})} \left(\sup_{\theta' \in \Theta} \{-\log \mathbf{g}_{\theta'}(\mathbf{X})\} \right)^2 \right] < \infty. \quad (4.33)$$

- ◇ The size of every sample is positive: $n_{\ell} \geq 1$ for any iteration $\ell \in \llbracket 1, L \rrbracket$.

Then, under these conditions, the [algorithm 6](#) is consistent:

$$\theta^{(L)} \xrightarrow[L \rightarrow \infty]{\text{a.s.}} \theta^*. \quad (4.34)$$

THEOREM 7 · Theorem 3 from [\[DP18\]](#)

Under the assumptions of [theorem 6](#), and with the additionnal assumption that there exists $\eta > 0$ such that:

$$\sup_{\theta \in \Theta} \mathbb{E}_{\mathbf{P}} \left[\mathbf{I}(\mathbf{X})^{2+\eta} \left(\frac{\mathbf{p}(\mathbf{X})}{\mathbf{g}_{\theta}(\mathbf{X})} \right)^{1+\eta} \right] < \infty, \quad (4.35)$$

and by denoting $N_L = \sum_{\ell=1}^L n_{\ell}$, we have:

$$\sqrt{N_L} (\hat{\mathcal{T}}^{\text{RCE}} - \mathcal{I}) \xrightarrow[N_L \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left(0, \mathbb{E}_{\mathbf{P}} \left[\mathbf{I}(\mathbf{X})^2 \frac{\mathbf{p}(\mathbf{X})}{\mathbf{g}_{\theta^*}(\mathbf{X})} \right] - \mathcal{I}^2 \right). \quad (4.36)$$

These conditions, especially [eqs. \(4.31\)](#), [\(4.33\)](#) and [\(4.35\)](#), are unfortunately challenging to interpret and verify in practice.

A proof for the PDMP case

We propose explicit conditions on the PDMP and the family of IFs $(\xi_{\theta})_{\theta \in \Theta}$ that are easier to interpret and verify and that imply the satisfaction of the assumptions of [theorems 6](#) and [7](#).

ASSUMPTION 1 · Assumptions on the PDMP

The PDMP of distribution \mathbf{p} on \mathbf{T}_{τ} with states in \mathcal{X} , flow Ψ , intensity λ and kernel Q verifies the following conditions:

- ◇ There exist $\lambda_{\min}, \lambda_{\max} > 0$ such that for any $x \in \text{supp}(\lambda)$, we have $\lambda_{\min} \leq \lambda(x) \leq \lambda_{\max}$.
- ◇ There exist $q_{\min}, q_{\max} > 0$ such that for any $x^- \in \mathcal{X}$ and $x \in \text{supp} q(\cdot | x^-)$, we have $q_{\min} \leq q(x | x^-) \leq q_{\max}$.

- ◇ For any $\varepsilon > 0$, let $\mathcal{A}_\varepsilon = \{x' \in \mathcal{X} \mid t_\theta(x') \geq \varepsilon\}$. There exists $\bar{\varepsilon} > 0$ such that $Q(\mathcal{A}_{\bar{\varepsilon}} \mid x) = 1$ for any $x = (z, v) \in \partial\mathcal{X}_v$.

ASSUMPTION 2 · Assumptions on the IF family $(\xi_\theta)_{\theta \in \Theta}$

- ◇ The function $\mathbf{x} \mapsto |\mathbf{I}|$ is bounded.
- ◇ Θ is a compact subset of \mathbb{R}^{d_θ} for some $d_\theta > 0$.
- ◇ $\theta^* \in \Theta$ is the unique minimizer of $\theta \mapsto -\mathbb{E}_p[\mathbf{I}(\mathbf{X}) \log \mathbf{g}_\theta(\mathbf{X})]$.
- ◇ There exist $\xi_{\min}, \xi_{\max} > 0$ such that for any $\theta \in \Theta$ and $x \in \mathcal{X}$, $\xi_{\min} \leq \xi_\theta(x) \leq \xi_{\max}$.

We simply need the optimization problem to be well-posed, the functions λ_θ and q_θ to be bounded from above and below on their support, and ensure that a jump at the boundary cannot occur arbitrarily close to the previous one.

We can now introduce our theorem.

THEOREM 8 · Asymptotic optimality

Assume that one of the two following conditions is satisfied:

- (i) $n_\ell > 0$ for any $\ell \in \mathbb{N}^*$ and $L \rightarrow \infty$,
- (ii) $L < \infty$ is fixed, $n_{L-1} \rightarrow \infty$ and $n_L/n_{L-1} \rightarrow \infty$.

Under [assumption 1](#) and [assumption 2](#), we have consistency:

$$\theta^{(L)} \xrightarrow[L \rightarrow \infty]{\text{a.s.}} \theta^*, \quad (4.37)$$

and asymptotic normality:

$$\sqrt{N_L}(\hat{\mathcal{I}}^{\text{RCE}} - \mathcal{I}) \xrightarrow[N_L \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma_{\mathbf{g}_{\theta^*}}^2), \quad (4.38)$$

with total sample size $N_L = \sum_{\ell=1}^L n_\ell$ and asymptotic variance $\sigma_{\mathbf{g}_{\theta^*}}^2$ given by:

$$\mathbb{E}_p \left[\mathbf{I}(\mathbf{X})^2 \frac{\mathbf{p}(\mathbf{X})}{\mathbf{g}_{\theta^*}(\mathbf{X})} \right] - \mathcal{I}^2. \quad (4.39)$$

The asymptotics can therefore be taken either in the number of iterations L or in the size of the last two samples n_{L-1} and n_L . These are two different yet specific ways to make the total number

of simulated trajectories tend towards infinity. The first case, more standard, is proven in the paragraph below⁷. In the second case, with a fixed number of iterations L , if the number of simulated trajectories at the second to last iteration n_{L-1} tends to infinity, then we minimize $\theta \mapsto -\mathbb{E}_{\mathbf{p}}[\mathbf{I}(\mathbf{X}) \log \mathbf{g}_{\theta}(\mathbf{X})]$ which gives θ^* . Thus at the last iteration L , the trajectories are generated according to \mathbf{g}_{θ^*} . It is then sufficient that n_L/n_{L-1} tends to infinity for the proportion of trajectories drawn according to \mathbf{g}_{θ^*} to converge to one.

We have defined the duration of a trajectory as bounded in the problem statement : $\tau \leq s_{\max}$. The [theorem 8](#) can be proved under a less restrictive assumption, the duration τ just needs to admit exponential moments of all orders:

$$\mathbb{E}[e^{c\tau}] \leq +\infty \text{ for any } c > 0. \quad (4.40)$$

However, this condition is more difficult for practitioners to interpret and verify.

Proof. The results come directly from showing that we verify the hypotheses of Theorems 2 and 3 from [\[DP18\]](#). We know by [assumption 2](#) that Θ is compact and that we have $-\mathbb{E}_{\mathbf{p}}[\mathbf{I}(\mathbf{X}) \log \mathbf{g}_{\theta}(\mathbf{X})] > -\mathbb{E}_{\mathbf{p}}[\mathbf{I}(\mathbf{X}) \log \mathbf{g}_{\theta^*}(\mathbf{X})]$ if $\theta \neq \theta^*$. Moreover, since ξ_{θ} is uniformly bounded from above and below, the continuity of the application $\theta \mapsto \xi_{\theta}(v)$ for any $x = (z, v) \in \mathcal{X}$ implies the continuity of $\theta \mapsto \mathbf{I}(\mathbf{x}) \mathbf{p}(\mathbf{x}) \log \mathbf{g}_{\theta}(\mathbf{x})$ for any $\mathbf{x} \in \mathcal{T}_{\tau}$. To obtain the convergence of the sequence $(\theta^{(L)})_{L>0}$, it remains to show that:

$$-\mathbb{E}_{\mathbf{p}} \left[\mathbf{I}(\mathbf{X}) \sup_{\theta \in \Theta} \log \mathbf{g}_{\theta}(\mathbf{X}) \right] < \infty, \quad (4.41)$$

$$\sup_{\theta \in \Theta} -\mathbb{E}_{\mathbf{p}} \left[\mathbf{I}(\mathbf{X}) \frac{\mathbf{p}(\mathbf{X})}{\mathbf{g}_{\theta}(\mathbf{X})} \left(\sup_{\theta' \in \Theta} \log \mathbf{g}_{\theta'}(\mathbf{X}) \right)^2 \right] < \infty. \quad (4.42)$$

And to get the asymptotic normality of the estimator $\hat{\mathcal{T}}^{\text{rcE}}$, we have to prove that there exists $\gamma > 0$ such that:

$$\sup_{\theta \in \Theta} \mathbb{E}_{\mathbf{p}} \left[\mathbf{I}(\mathbf{X})^{2+\gamma} \left(\frac{\mathbf{p}(\mathbf{X})}{\mathbf{g}_{\theta}(\mathbf{X})} \right)^{1+\gamma} \right] < \infty. \quad (4.43)$$

From the definitions [eqs. \(4.10\)](#) and [\(4.11\)](#) of λ_{θ} and q_{θ} , then from [assumption 1](#) and [assumption 2](#), we obtain that for any $x \in \text{supp}(\lambda)$:

$$\lambda_{\min} \xi_{\min} / \xi_{\max} \leq \lambda_{\theta}(x) \leq \lambda_{\max} \xi_{\max} / \xi_{\min}, \quad (4.44)$$

and for any $x^- \in \mathcal{X}$ and any $x \in \text{supp } q(\cdot | x^-)$:

$$q_{\min} \times \xi_{\min} / \xi_{\max} \leq q(x | x^-) \leq q_{\max} \times \xi_{\max} / \xi_{\min}. \quad (4.45)$$

⁷This proof does not assume that the IF ξ_{θ} takes the form given in [definition 27](#), nor that it is a function of the regime only.

Then from the definition eq. (4.13) of the density of a PDMP trajectory $\mathbf{x} \in \mathbb{T}_\tau$ with $\kappa(\tau)$ jumps, there exist $c_{\min}, c_{\max} > 0$ such that:

$$\begin{aligned} \sup_{\theta \in \Theta} \mathbf{g}_\theta(\mathbf{x}) &\leq \prod_{k=0}^{\kappa(\tau)-1} \max\left(\lambda_{\max} \frac{\xi_{\max}}{\xi_{\min}}, 1\right) q_{\max} \frac{\xi_{\max}}{\xi_{\min}} \leq (c_{\max})^{\kappa(\tau)-1} \\ \inf_{\theta \in \Theta} \mathbf{g}_\theta(\mathbf{X}) &\geq \exp\left[-\lambda_{\max} \frac{\xi_{\max}}{\xi_{\min}} s_{\max}\right] \prod_{k=0}^{\kappa(\tau)-1} \min\left(\lambda_{\min} \frac{\xi_{\min}}{\xi_{\max}}, 1\right) \exp\left[-\lambda_{\max} \frac{\xi_{\max}}{\xi_{\min}} s_{\max}\right] q_{\min} \frac{\xi_{\min}}{\xi_{\max}} \\ &\geq (c_{\min})^{\kappa(\tau)} \end{aligned} \quad (4.46)$$

These results also hold for \mathbf{p} since $\mathbf{p} = \mathbf{g}_0$. Using eq. (4.46) we see that conditions eqs. (4.41) to (4.43) are ensured by the following one: for any positive constant $c > 0$, $\mathbb{E}_{\mathbf{p}}[c^{\kappa(\tau)}] < +\infty$.

We have $\kappa(\tau) = \kappa_\lambda(\tau) + \kappa_\partial(\tau)$ with $\kappa_\lambda(\tau)$ the number of spontaneous jumps with jump rate λ and $\kappa_\partial(\tau)$ the number of jumps at boundaries. At most, the process reaches the state space boundary “almost immediately” after each spontaneous jump, and a time $\bar{\varepsilon}$ after reaching another boundary. So $\kappa_\partial(\tau) \leq \kappa_\lambda(\tau) + s_{\max}/\bar{\varepsilon}$, and thus $\kappa(\tau) \leq 2\kappa_\lambda(\tau) + s_{\max}/\bar{\varepsilon}$. We just need to prove that $\mathbb{E}_{\mathbf{p}}[c^{\kappa_\lambda(\tau)}] < +\infty$ for any positive constant c .

We define $\tilde{\mathbf{X}}$ as a jump process analogous to a PDMP but with some jumps rejected and not taking place. It is characterized by its flow Ψ , its constant jump intensity λ_{\max} and its jump kernel \tilde{Q} defined as follows:

$$\tilde{Q}(dx | x^-) = \begin{cases} Q(dx | x^-) \frac{\lambda(x^-)}{\lambda_{\max}} + \left(1 - \frac{\lambda(x^-)}{\lambda_{\max}}\right) \delta_{x^-}(dx) & \text{if } x^- \notin \partial\mathcal{X}, \\ Q(dx | x^-) & \text{otherwise.} \end{cases} \quad (4.47)$$

Thus a part of the spontaneous jumps are “rejected” because the process remains on the same state at each jump with probability $\left(1 - \frac{\lambda(x^-)}{\lambda_{\max}}\right)$. Following theorem 5.5 from [Dav84], we notice that the generator of this process is the same as that of the PDMP \mathbf{X} . Indeed by denoting \mathcal{A} the generator of the PDMP \mathbf{X} and $\tilde{\mathcal{A}}$ the generator of the process $\tilde{\mathbf{X}}$, for any state $x^- \in \mathcal{X}$ and φ a function of the domain of the extended generator $\tilde{\mathcal{A}}$ (see theoretical detail in [Dav84] and our brief presentation of the PDMP generator in section 2.2.2):

$$\begin{aligned} \tilde{\mathcal{A}}\varphi(x^-) &= \mathfrak{F}\varphi(x) + \lambda_{\max}(x^-) \int_{\mathcal{X}} [\varphi(x) - \varphi(x^-)] \tilde{Q}(dx | x^-), \\ &= \mathfrak{F}\varphi(x) + \lambda(x^-) \int_{\mathcal{X}} [\varphi(x) - \varphi(x^-)] Q(dx | x^-) + \left(1 - \frac{\lambda(x^-)}{\lambda_{\max}}\right) [\varphi(x^-) - \varphi(x^-)], \\ &= \mathcal{A}\varphi(x^-). \end{aligned} \quad (4.48)$$

Since the generator characterizes the distribution of the process, the trajectories of the PDMP \mathbf{X} and of the jump process $\tilde{\mathbf{X}}$ are identically distributed. In particular, the number of jumps $\tilde{\kappa}(\tau)$ of the process $\tilde{\mathbf{X}}$ follows the same distribution than $\kappa(\tau)$, and $\tilde{\kappa}'_{\lambda_{\max}}(\tau)$ follows the same distribution than $\kappa_\lambda(\tau)$. Let us note $\tilde{\kappa}'_{\lambda_{\max}}(\tau)$ the number of proposed jumps for the process $\tilde{\mathbf{X}}$ with jump intensity λ_{\max} including the rejected ones. The number of proposed jumps in the time interval $[0, \tau)$ therefore does not depend on the states occupied by the process $\tilde{\mathbf{X}}$ and follows a Poisson distribution with intensity $\lambda_{\max} \times \tau$. It is also

straightforward to see that $\tilde{\kappa}'_{\lambda_{\max}}(\tau) \geq \tilde{\kappa}_{\lambda_{\max}}(\tau)$. Thus, for any $c > 0$, we have:

$$\begin{aligned} \mathbb{E}_{\mathbf{p}}[c^{\tilde{\kappa}_i(\tau)}] &= \mathbb{E}[c^{\tilde{\kappa}_{\max}(\tau)}] \leq \mathbb{E}[c^{\tilde{\kappa}'_{\max}(\tau)}] = \mathbb{E}[\mathbb{E}[c^{\text{Pois}(\lambda_{\max} \times \tau)} \mid \tau]] \\ &= \mathbb{E}\left[\sum_{k=0}^{\infty} c^k \frac{(\lambda_{\max} \times \tau)^k}{k!} e^{-\lambda_{\max} \times \tau}\right] = \mathbb{E}[e^{\lambda_{\max} \tau (c-1)}] < +\infty. \end{aligned} \quad (4.49)$$

The last inequality comes from the assumption [eq. \(4.40\)](#), which is clearly met in the case $\tau \leq s_{\max}$. This completes the proof of the theorem. ■

Asymptotic confidence intervals

If the convergence of the parameter $\theta^{(L)}$ to θ^* is reassuring, it is primarily the asymptotic normality of the estimator that will serve us in practice. Indeed, it opens the door to asymptotic confidence intervals allowing us to quantify the uncertainty around our estimate.

For a number of iterations L and a total sample size $N_L = \sum_{\ell=1}^L n_{\ell}$, and under the assumptions of [theorem 8](#), we propose in [eq. \(4.50\)](#) a consistent estimator $\hat{\sigma}_{\mathbf{g}^*}^2$ of the asymptotic variance $\sigma_{\mathbf{g}^*}^2$ given in [eq. \(4.39\)](#):

$$\hat{\sigma}_{\mathbf{g}^*}^2 = \frac{1}{N_L} \sum_{\ell=1}^L \sum_{k=1}^{n_{\ell}} \mathbf{I}(\mathbf{X}_k^{(\ell)})^2 \left[\frac{\mathbf{p}(\mathbf{X}_k^{(\ell)})}{\mathbf{g}^{\theta^{(\ell)}}(\mathbf{X}_k^{(\ell)})} \right]^2 - (\hat{\mathcal{I}}^{\text{RCE}})^2. \quad (4.50)$$

We thus have asymptotic validity of the following confidence interval through an application of Slutsky's lemma:

THEOREM 9 · Asymptotic confidence interval

Let $\alpha \in [0, 1]$ be a confidence level and Φ^{-1} be the quantile function of the $\mathcal{N}(0, 1)$ distribution. Then, as $N_L \rightarrow \infty$, we have:

$$\mathbb{P}(\mathcal{I} \in [\hat{\mathcal{I}}^{\text{RCE}} - \hat{C}_{\alpha, L}; \hat{\mathcal{I}}^{\text{RCE}} + \hat{C}_{\alpha, L}]) \longrightarrow 1 - \alpha, \quad (4.51)$$

with $\hat{C}_{\alpha, L} := \Phi^{-1}(1 - \frac{\alpha}{2}) \sqrt{\frac{\hat{\sigma}_{\mathbf{g}^*}^2}{N_L}}$.

4.3.3 Implementation guidelines

The previous subsection aimed to present the adaptive importance sampling algorithm with recycling scheme and its properties in the most general way. Some aspects of the method were not mentioned (initialization, sampling policy, optimization procedure). Essential for an efficient

implementation, these aspects are more specific to the considered problem, and the practitioner should have control over them.

Initialization of the algorithm

It is known that the choice of the initial distribution of an AIS method is crucial. A common option that has the advantage of being without a priori is to take $\mathbf{g}_{\theta^{(0)}} = \mathbf{p}$, but it is not suitable to deal with rare events. Let us recall that we would like to maximize $\mathbb{E}_{\mathbf{p}}[\mathbf{I}(\mathbf{X}) \log \mathbf{g}_{\theta}(\mathbf{X})]$ but we minimize in practice an empirical approximation eq. (4.28). With a poorly chosen initial auxiliary distribution, the minimizer of the approximation could be too far from the true minimizer. It is then difficult to find the right track over the iterations and the final result of the procedure depends therefore a lot on this initial choice. Determining an initial importance distribution that both frequently realizes the event of interest and produces reasonably stable likelihood ratios is generally too ambitious. It is often better to “help without pushing too hard” initially, even if it consumes a large number of simulations to obtain the first realizations of the event of interest.

The multilevel approach described in section 1.3.2 is a classic and effective alternative in this case. Levels can here be defined by the maximum proximity scores reached by the trajectory. A generated trajectory then no longer needs to realize the event of interest to contribute to the objective function but only needs to exceed a given threshold. However, this may not necessarily solve the initialization problem. The initial state of the trajectory usually has a small proximity score (typically $\beta_{v_0} = 0$), and there is no guarantee that it is likely to exit this state within a time less than s_{\max} . This is the case, for example, when simulating highly reliable industrial systems where no component fails over long periods.

We propose the following heuristic, which works very well in our test cases. The user sets a time $\tilde{s} < s_{\max}$ and a probability $\tilde{p} \in [0, 1]$. We determine the initial parameter $\theta^{(0)}$ with the smallest norm such that the probability of the first jump occurring before time \tilde{s} is greater than or equal to \tilde{p} .

$$\begin{aligned} \Theta_{\text{init}} &:= \left\{ \theta \in \Theta : \mathbb{P}_{T \sim p_{\theta}(\cdot | x_0)}(T \leq \tilde{s}) \geq \tilde{p} \right\}, \\ \theta^{(1)} &\in \arg \min_{\theta \in \Theta_{\text{init}}} \|\theta\|. \end{aligned} \tag{4.52}$$

We recommend the use of a Newton-Raphson method to determine the root of the function $\theta \mapsto \mathbb{P}_{T \sim p_{\theta}(\cdot | x_0)}(T \leq \tilde{s}) - \tilde{p}$ (its derivative is explicitly known). Thus no preliminary sampling is required.

Sampling size policy and stopping criterion

Given a fixed simulation budget, we would like to choose n_{ℓ} the sample size at iteration ℓ , and L the total number of iterations. Let us recall that $N_L = \sum_{\ell=1}^L n_{\ell}$ is the total number of trajectories to be drawn.

If we do not opt for a recycling scheme, it is imperative to set n_ℓ large enough (at the very least $n_\ell = 100$) even if it means doing few iterations, because we cannot rely on past samples to approximate the true objective function in eq. (4.24). On the other hand with a recycling scheme, we would ideally like to choose n_ℓ as small as possible in order to perform as many iterations as possible, thus as many minimizations as possible, and to give ourselves the best chances to get close to \mathbf{g}^* .

In practice, it all depends on the optimization method used to solve eqs. (4.24) and (4.28). If it is expensive, we cannot afford too many iterations. Moreover, the cost of calling the function to be minimized depends linearly on the number of terms in the sum, so the minimization will be more and more expensive with each iteration (using a recycling scheme). To get an idea, if we do not want the cost dedicated to the optimization to exceed the cost dedicated to the simulation, the number of iterations L should be smaller than $2c_S/c_O - 1$ with c_S the computational cost of simulating one trajectory and c_O the computational cost of maximizing eq. (4.24) for a single trajectory.

For a fixed simulation budget N_L , we propose to determine n_ℓ at iteration ℓ as the minimum between the remaining budget $N_L - N_{\ell-1}$ and the (random) smallest number of trajectories to be drawn such that n_{CE} of them belong to \mathbf{F} (with for example $n_{\text{CE}} = 10$ in the case of a total budget $N_L = 10^4$).

Numerical optimization

It is necessary at each iteration to call an optimization program to solve eq. (4.28). Let us point out that with a large number of iterations and few new trajectories at each sample, it is not useful to determine the true minimizer (which would imply using sophisticated methods adapted to non-convex problems). We only need to improve our instrumental distribution a little bit at each iteration.

We employed the BFGS method [Dai02] that can be found in many toolboxes. We used the function `minimize(·, method=BFGS)` from the `scipy.optimize` toolbox in Python [Vir+20]. A classical stopping criterion of a BFGS method is to obtain a sufficiently small gradient (i.e. close to zero). The default threshold in `scipy.optimize` is 10^{-5} . It is better to drastically lower this threshold to 10^{-20} for example because since the set of possible trajectories is a very high dimensional space, the p.d.f. of the trajectories are very small and the gradient of the objective function is small.

The BFGS method performs better when given the explicit gradient of the function to be minimized rather than letting it approximate it by finite differences. The only quantity depending on θ in the objective function is the log-likelihood: $\mathbf{x} \mapsto \log \mathbf{g}_\theta(\mathbf{x})$. Let $\mathbf{x} \in \mathbf{T}_\tau$ be a PDMP trajectory of duration τ , with $n = \kappa(\tau)$ jumps and skeleton $(x_{t_k}, t_k)_{k=0}^n$. The gradient of $\log \mathbf{g}_\theta(\mathbf{x})$ with respect to θ is given by:

$$\begin{aligned} \nabla_{\theta} \log \mathbf{g}_{\theta}(\mathbf{x}) &= \sum_{k=0}^{n-1} \left[\nabla_{\theta} \log p_{\lambda_{\theta}}(t_k | x_{x_k}) + \nabla_{\theta} \log q_{\theta}(x_{x_{k+1}} | \Psi(x_{x_k}, t_k)) \right] \\ &\quad - \int_0^{t_n} \lambda(\Psi(x_{x_n}, u)) \nabla_{\theta} r_{\theta}(\Psi(x_{x_n}, u), \Psi(x_{x_n}, u)) \, du, \end{aligned} \quad (4.53)$$

with, for any states $x^{-}, x \in \mathcal{X}$, and any time $t > 0$:

$$r_{\theta}(x^{-}, x) := \xi_{\theta}^{-}(x^{-}) / \xi_{\theta}(x) \quad (4.54)$$

$$\nabla_{\theta} r_{\theta}(x^{-}, x) = \int_{\mathcal{X}} \frac{\xi_{\theta}(x) \nabla_{\theta} \xi_{\theta}(x^{+}) - \xi_{\theta}(x^{+}) \nabla_{\theta} \xi_{\theta}(x)}{\xi_{\theta}(x)^2} q(x^{+} | x^{-}) \nu_{\theta}(dx^{+} | x^{-}), \quad (4.55)$$

$$\nabla_{\theta} \log r_{\theta}(x^{-}, x) = \frac{\nabla_{\theta} r_{\theta}(x^{-}, x)}{r_{\theta}(x^{-}, x)}, \quad (4.56)$$

$$\begin{aligned} \nabla_{\theta} \log p_{\lambda_{\theta}}(t | x) &= \mathbb{1}_{t < t_{\theta}(x)} \nabla_{\theta} \log r_{\theta}(\Psi(x, t), \Psi(x, t)) \\ &\quad - \int_0^t \lambda(\Psi(x, u)) \nabla_{\theta} r_{\theta}(\Psi(x, u), \Psi(x, u)) \, du, \end{aligned} \quad (4.57)$$

$$\nabla_{\theta} \log q_{\theta}(x | x^{-}) = -\nabla_{\theta} \log r_{\theta}(x^{-}, x). \quad (4.58)$$

Finally, the importance function ξ_{θ} we proposed in [definition 27](#) admits a very simple gradient for any $x = (z, v) \in \mathcal{X}$:

$$\nabla_{\theta} \xi_{\theta}(x) = \left(\phi_j(\beta(v)) \right)_{j=1}^{d_{\Theta}} \times \xi_{\theta}(x). \quad (4.59)$$

If the number of possible transitions from any state $x \in \mathcal{X}$ is finite, then the above quantities simplify, and the computation of $\nabla_{\theta} \log \mathbf{g}_{\theta}$ reduces to vector operations that can be implemented efficiently.

CHAPTER 5

Importance function for multicomponent systems

An adaptive importance sampling strategy has been proposed in [chapter 4](#) for rare event simulation of PDMPs. The importance distribution is built upon an approximation of the committor function of the process. This approximation, called importance function, is a parametric transformation of a proximity score β between any regime of the PDMP and the set $\mathcal{V}_{\mathcal{F}}$ of regimes allowing to access the critical region \mathcal{F} . In this chapter, we consider PDMPs modeling multi-component industrial systems. The rare event of interest is the critical failure of the system. This critical failure can only occur when certain key combinations of components remain in a failed state for a sufficiently long time. Fault tree analysis provides tools to rigorously define and determine these combinations of components. We propose a proximity score based on the notion of minimal path sets, which corresponds to the combinations of components that prevent the system failure. Our adaptive importance sampling method is tested with this proximity score on toy series/parallel systems then on a more complex industrial test case: the spent fuel pool.

Content of this chapter

5.1	A first approach for multicomponent systems	108
5.1.1	Multicomponent systems in reliability analysis	108
5.1.2	Coherent systems	110
5.1.3	Broken Components Rate approach	112
5.2	Proximity scores based on fault tree analysis	114
5.2.1	The limits of structure blindness	114
5.2.2	Minimal cut sets and minimal path sets	118
5.2.3	Proximity score and implementation	120
5.3	Numerical experiments	123
5.3.1	Series and parallel systems	123
5.3.2	The spent fuel pool	125

5.1 A first approach for multicomponent systems

The previous chapter established the connection between a general parametric family of importance functions $(\xi_\theta)_{\theta \in \Theta}$ (viewed as a family of approximations of the committor function ξ^* of the process) and an adaptive importance sampling method with a recycling scheme of past samples for the estimation of rare event probabilities. On the one hand, we provided guarantees on the asymptotic behavior of the estimator for any family of IFs. On the other hand, we proposed a particular (yet very general) form of IF. For any $\theta \in \Theta$ and any $x \in \mathcal{X}$, we have set:

$$\xi_\theta(x) = \exp \left[\sum_{j=1}^{d_\Theta} \theta_j \times \phi_j(\beta(v)) \right]. \quad (5.1)$$

We clarified the role of proximity scores β and basis functions $(\phi_j)_j$ (specifying several families of examples for the latter) in [section 4.2.1](#). As a reminder, the score $\beta(v)$ of a regime $v \in \mathcal{V}$ should quantify its proximity to the set of critical regimes¹ $\mathcal{V}_\mathcal{F}$. The IF, although a function of the process state, only depends on the regime of the process in practice. It is a transformation parameterized by θ of the regime's proximity score.

5.1.1 Multicomponent systems in reliability analysis

This chapter focuses on the main case study of the thesis: the reliability analysis of multicomponent industrial systems. It is mainly based on our paper [[Che+24a](#)]. Here, \mathcal{F} represents the region of critical system failure. The major contribution of this work is to propose families of proximity scores that enable the construction of IFs suitable for approximating the committor functions of high-dimensional hybrid systems.

In the context of reliability analysis, a multicomponent industrial system refers to a complex system composed of multiple interconnected elements or subsystems, called components. Each component can be an equipment, a machine, an electronic device, etc. It contributes in one way or another, and most often in interaction with the other components, to accomplish the overall function of the system. Interactions may include flows of material, energy, information, or control between components. For example, a component may receive inputs from another component and produce outputs that are then used by other components. A component can be in different operating modes, for example: active, inactive, active in a degraded mode, failed, etc. This operating mode is called the component's status. The status can be automatically changed via control and feedback mechanisms or following random events such as failures or repairs.

¹A critical regime is a regime leading to the failure region \mathcal{F} .

Critical failure and PDMP interpretation

The reliability of an industrial system is its ability to perform its overall function for a given period of time under specified conditions. Critical failure is declared when this function is no longer fulfilled. It generally corresponds to the exceeding of critical thresholds by key continuous quantities of the system (e.g., temperature, pressure, or liquid level in a tank). If the system is well designed, this can only happen if the system stays long enough in a very degraded state, i.e., after the failure of vital groups of components and before their repair. This falls within the framework of PDMPs. Component failures and repairs are viewed as random one-time events, while the evolution of continuous variables is governed by deterministic differential equations derived from physical laws. Each component has its own reliability characteristics that may continuously depend on continuous physical variables (for example, an immersed component may be weakened by excessive pressure or temperature of a liquid). The differential equations governing the evolution of physical variables are, in turn, parameterized by the status of the components (water, for example, will flow through the system at a different rate if a valve remains stuck in the open or closed position).

The high reliability of these systems is explained on the one hand by their high level of redundancy: the system can be reconfigured using several identical components to ensure its operation while waiting for the repair or replacement of broken components. On the other hand, the average waiting time before the failure of a component is generally considerably larger than the average waiting time before its repair.

Static and binary interpretation

Our task is to determine how to quantify the proximity of a regime $v \in \mathcal{V}$, partially or fully determined by the status of the components, to the set \mathcal{V}_F . We assume throughout this chapter that the broken² or unbroken status of each component of the system is sufficient to determine whether the regime v belongs to \mathcal{V}_F . Therefore, we adopt a binary and static view of the state of a system and its components.

- (i) Regardless of the variety of possible statuses of a component, it only admits one significant degraded mode: failure. Its status can therefore be represented by a Boolean variable equal to 0 if it is failed and 1 otherwise.
- (ii) The system is said to work or function in a given state $x = (z, v) \in \mathcal{X}$ if and only if $v \notin \mathcal{V}_F$ (and not just as long as $x \notin \mathcal{F}$). And as we have said, belonging to \mathcal{V}_F depends on which components are failed.

In the static point of view, we consider the final regime of the trajectory and we declare that the trajectory has failed if that regime belongs to \mathcal{V}_F (without taking account the time during which

²“broken” and “failed” are used interchangeably.

the position evolves to \mathcal{F} when its regime belongs to $\mathcal{V}_{\mathcal{F}}$ as well as the possibility that the system is repaired during this time).

5.1.2 Coherent systems

The proximity scores we propose are based in part on a so-called “coherence” assumption of the system.

DEFINITION 28 · Coherent system

A system is said to be coherent if it verifies the following conditions:

- (i) it works when none of its components is broken,
- (ii) it does not work when all its components are broken,
- (iii) if it does not work in a given regime then it does not work with the additional failure of a component,
- (iv) conversely if it works in a given regime then it still works if broken components are repaired,

Interpretation and counterexample

It is not a big assumption to suppose that there always exists at least one configuration in which the system works and at least one in which it does not. There is no guarantee that these configurations correspond respectively to the one where all components are failed and the one where none are. Indeed, and sometimes counterintuitively, there are systems for which the malfunction of an additional component at the right time can prevent critical failure. Here is a simple example of a non-coherent system.

EXAMPLE 8 · Controlled Mixing System with Intersection Priority

- ◇ The system comprises two production lines, L_A and L_B , intersecting at a junction.
- ◇ Each line transports a distinct product, denoted as P_A and P_B , respectively.
- ◇ The system has to ensure that P_A and P_B do not mix at the intersection.
- ◇ A valve V_B , is positioned on production line L_B to block the flow of product P_B until product P_A has traversed the intersection.

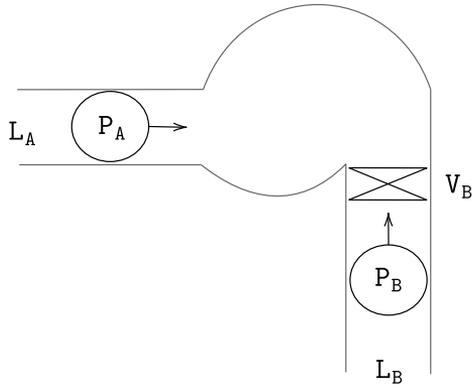


FIGURE 5.1 Representation of the controlled mixing system.

However, if valve V_B malfunctions and remains open, allowing P_B to flow prematurely, it is preferred for production line L_A to also malfunction. This additional malfunction prevents component P_A from reaching the intersection, thereby avoiding mixing with P_B . This is a non-coherent behavior because an additional malfunction, which typically leads to system failure, actually prevents the occurrence of a critical failure scenario, thus defying the expected outcome.

Most well-designed systems do not suffer from this issue, but it is difficult to verify the coherence of a complex system. This coherence assumption can be formulated mathematically using some basic notions of Boolean algebra.

Formal definition with Boolean algebra

Let d_c be the number of components in the system.

- (i) For $j \in \llbracket 1, d_c \rrbracket$, the Boolean variable $c_j(v) \in \{0, 1\}$ represents the status of the j -th component in regime $v \in \mathcal{V}$, with $c_j(v) = 0$ if the j -th component is failed and $c_j(v) = 1$ if it is operational.
- (ii) The Boolean vector $\mathbf{c}(v) = (c_1(v), \dots, c_{d_c}(v)) \in \{0, 1\}^{d_c}$ denote the status of all d_c components in regime $v \in \mathcal{V}$.
- (iii) For any Boolean variable $c \in \{0, 1\}$, we denote by $\bar{c} = 1 - c$ its Boolean opposite, and for any Boolean vector $\mathbf{c} \in \{0, 1\}^d$, we note $\bar{\mathbf{c}} = (\bar{c}_j)_{j=1}^d$.

DEFINITION 29 · Structure function

We call the structure function of the system the Boolean function^a $\varphi_{\mathcal{F}}$ which associates 1 to the regimes for which the system works and 0 to the other ones. For any regime $v \in \mathcal{V}$ and denoting by $\mathbf{c}(v) \in \{0, 1\}^{d_c}$ the status of the components in that regime, we define :

$$\varphi_{\mathcal{F}} : \mathbf{c}(v) \mapsto 1 - \mathbf{1}_{v \in \mathcal{V}_{\mathcal{F}}}. \quad (5.2)$$

^aA function from $\{0, 1\}^d$ to $\{0, 1\}$ with $d \in \mathbb{N}^*$.

Let $\mathbf{c}' = (c'_1, \dots, c'_{d_c}) \in \{0, 1\}^{d_c}$, we write $\mathbf{c} \leq \mathbf{c}'$ if and only if $c_j \leq c'_j$ for all $j \in \llbracket 1, d_c \rrbracket$. Let us recall that a Boolean function φ is non-decreasing if and only if $\varphi(\mathbf{c}) \leq \varphi(\mathbf{c}') \iff \mathbf{c} \leq \mathbf{c}'$ for all $\mathbf{c}, \mathbf{c}' \in \{0, 1\}^{d_c}$.

PROPOSITION 16 · Coherent system

A system with structure function $\varphi_{\mathcal{F}}$ is coherent (i.e. verifies the conditions of [definition 28](#)) if and only if :

- (i) $\mathbf{c}(v) = \mathbf{1}_{d_c} \implies \varphi_{\mathcal{F}}(\mathbf{c}(v)) = 1$,
- (ii) $\mathbf{c}(v) = \mathbf{0}_{d_c} \implies \varphi_{\mathcal{F}}(\mathbf{c}(v)) = 0$,
- (iii) $\varphi_{\mathcal{F}}$ is non-decreasing.

5.1.3 Broken Components Rate approach

The first approach we propose does not strictly require the assumption of system coherence. However, this assumption ensures that the more components are failed in the system, the more degraded it is and the more likely it is to be close to a critical regime (belonging to $\mathcal{V}_{\mathcal{F}}$). A very natural idea is therefore to construct an IF based on the quantity of failed components in the system. For any regime $v \in \mathcal{V}$, we denote $b(v)$ the number of failed components in regime v .

$$b : v \in \mathcal{V} \mapsto \sum_{j=1}^{d_c} 1 - c_j(v) = \sum_{j=1}^{d_c} \overline{c_j(v)}. \quad (5.3)$$

DEFINITION 30 · Broken Components Rate proximity score

We define the proximity score β^{BCR} as the square of the proportion of broken components in the system:

$$\beta^{\text{BCR}} : v \in \mathcal{V} \mapsto \left(\frac{b(v)}{d_c} \right)^2 = \left(\frac{1}{d_c} \sum_{j=1}^{d_c} c_j(v) \right)^2. \quad (5.4)$$

We also note ξ_θ^{BCR} the IF defined for any $\theta \in \Theta$ and basis functions $(\phi_j)_j$ by:

$$\xi_\theta^{\text{BCR}}((z, v)) = \exp \left[\sum_{j=1}^{d_\Theta} \theta_j \times \phi_j(\beta^{\text{BCR}}(v)) \right]. \quad (5.5)$$

Associated with a parameter θ of dimension 1 and a polynomial or piecewise linear basis function, we obtain an importance function (IF) of the form $\xi_\theta^{\text{BCR}}(x) = \exp \left[\theta_1 \times \left(\frac{b(v)}{d_c} \right)^2 \right]$. We then virtually obtain the IF proposed by [Chr+19] that we have described in section 3.3.2 (see eq. (3.52)). Our general form of IF based on the score β^{BCR} with θ of flexible dimension allows us to generalize the IF proposed by [Chr+19].

Justification

The proportion of broken components keeps $\beta^{\text{BCR}} \in [0, 1]$. Thus, for any $v \in \mathcal{V}$, $\beta^{\text{BCR}}(v) = 0$ only in the least degraded state (no failed components), and $\beta^{\text{BCR}}(v) = 1 \Rightarrow v \in \mathcal{V}_F$. It is also more interpretable than the raw number of failed components in the system, which provides little information when not normalized by the total number of components.

We kept the square from the IF of [Chr+19]. Their justification is the following. During a trajectory leading to failure, either the successive failure times occur very close together, or many components remain failed for a long time. These two scenarios are highly unlikely, unless the bias of the IF in favor of failures and against repairs increases with the number of broken components. Without the square and with a one-dimensional parameter $\theta \in \mathbb{R}$, we would have:

$$\frac{\exp \left(\frac{\theta}{d_c} (b(v) + 1) \right)}{\exp \left(\frac{\theta}{d_c} b(v) \right)} = \exp \left(\frac{\theta}{d_c} \right). \quad (5.6)$$

To avoid constant ratios (in $b(v)$), it is therefore necessary to choose β as a function that grows in $b(v)$ faster than a linear function (hence the square).

The argument proposed in [Chr+19] is actually more general. If the event $\{\mathbf{X} \in \mathbf{F}\}$ is rare, then the process is more likely to move away from \mathcal{F} at each jump than to move closer to it. If the value of the committor function is large at the current state of the trajectory, then it is expected

that with high probability it will be lpower after the next jump. Let $x = (z, v) \in \mathcal{X}$ be a state with $v \notin \mathcal{V}_{\mathcal{F}}$ and let $s \geq 0$ be a time instant, we denote $T(x)$ as the waiting time before the next jump from state x . Since $v \notin \mathcal{V}_{\mathcal{F}}$, the event $\{\mathbf{X} \in \mathbf{F}\}$ can only occur if $T(x) \leq s_{\max} - s$.

$$\begin{aligned} \xi^*(x, s) &= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\mathbf{X} \in \mathbf{F}} \mid X_{s+T(x)}\right] \mid X_s = x, T(x) \leq s_{\max} - s\right] \times \mathbb{P}(T(x) \leq s_{\max} - s), \\ &= \mathbb{E}\left[\xi^*(X_{s+T(x)}, s + T(x)) \mid X_s = x, T(x) \leq s_{\max} - s\right] \times \mathbb{P}(T(x) \leq s_{\max} - s), \\ &= \mathbb{E}\left[\xi^{*-}\left(\Psi(x, T(x)), s + T(x)\right) \mid X_s = x, T(x) \leq s_{\max} - s\right] \times \mathbb{P}(T(x) \leq s_{\max} - s). \end{aligned} \tag{5.7}$$

We observe that the expected value of the committor function after the next jump is its current value (in the case where the jump occurs). Therefore, there is a compensation phenomenon; after the next jump, the committor function will take:

- ◊ a slightly smaller value with high probability,
- ◊ or a much larger value with small probability.

The committor function increases the gaps in order to take into account the probability of moving backward (away from \mathcal{F}). Ideally we expect the proximity score of a good importance function (IF) to tighten up the smaller values and keep out the larger ones.

This was a crucial aspect in the article [Chr+19], but one that we will not really have to worry about with our IF. The proximity score only needs to correctly order the states. The gaps will be correctly adjusted during the adaptive procedure if the dimension of θ is large enough.

5.2 Proximity scores based on fault tree analysis

The IF $\xi_{\theta}^{\text{B}^{\text{CR}}}$ constructed from the proximity score $\beta^{\text{B}^{\text{CR}}}$ is a relevant generalization of the IF proposed in [Chr+19] for simple industrial systems. As we will see, the score $\beta^{\text{B}^{\text{CR}}}$ is not necessarily suitable for complex systems. One of the main contributions of this thesis is a new proximity score, as relevant as $\beta^{\text{B}^{\text{CR}}}$ on simple systems, and able to take advantage of the specificities of complex systems.

5.2.1 The limits of structure blindness

It should be noted that while we have $\beta^{\text{B}^{\text{CR}}}(v) = 1 \Rightarrow v \in \mathcal{V}_{\mathcal{F}}$, the reverse implication is not true. For most systems, the critical failure is reachable with only a portion of its components broken. The proximity score $\beta^{\text{B}^{\text{CR}}}$ is more suitable for some systems than for others. We introduce several industrial systems that will serve us both as theoretical examples and test cases for the implementation of our method.

Series and parallel systems

Series and parallel systems form two opposite poles among multi-component systems. They are also basic building blocks. Any coherent industrial system (see [definition 28](#)) can be decomposed into subsystems of series and parallel types.

EXAMPLE 9 · Series and parallel systems

A series system is a configuration of components in which the failure of any one component is sufficient to cause system failure (see [Figure 5.2](#)).

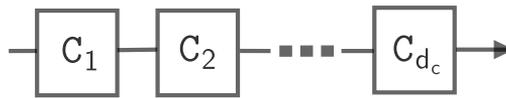


FIGURE 5.2 **Series system** with d_c components.

A parallel system is a configuration of components in which the failure of all components is necessary to cause system failure (see [Figure 5.3](#)).

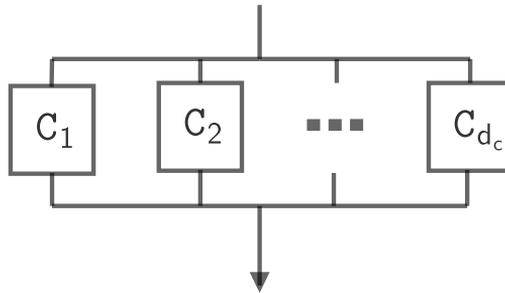


FIGURE 5.3 **Parallel system** with d_c components.

We will once again turn to Boolean algebra to shed light on these notions. To avoid any ambiguity, the operator $+$ will always be used in its traditional sense and never in its Boolean algebra sense (where $1 + 1 = 1$). We introduce the operators \vee and \wedge . For any $c, c' \in \{0, 1\}$, we note $c \vee c' = \max(c, c')$ and $c \wedge c' = \min(c, c')$. For any $J \subset \llbracket 1, d_c \rrbracket$, we also note :

$$\bigvee_{j \in J} c_j = \max\{c_j, j \in J\} \quad \text{and} \quad \bigwedge_{j \in J} c_j = \min\{c_j, j \in J\}. \quad (5.8)$$

From the structure function $\varphi_{\mathcal{F}}$ perspective, a system with d_c components is:

- (i) a series system if: $\varphi_{\mathcal{F}}(\mathbf{c}) = \bigwedge_{j=1}^{d_c} c_j$ for any $\mathbf{c} \in \{0, 1\}^{d_c}$,

(ii) a parallel system if: $\varphi_{\mathcal{F}}(\mathbf{c}) = \bigvee_{j=1}^{d_c} c_j$ for any $\mathbf{c} \in \{0, 1\}^{d_c}$.

The proximity score β^{BCR} is very well suited for parallel systems, since each component must be broken to achieve the system failure, and we indeed have $\beta^{\text{BCR}}(v) = 1 \Leftrightarrow v \in \mathcal{V}_{\mathcal{F}}$. The heated room system described in [section 3.3.2](#) is a good example of a parallel system. Failure is only possible when all the heaters are broken. This is why an importance function (IF) close to that provided by β^{BCR} gives such good results in [\[Chr+19\]](#). In the case of series systems, however, we have $\beta^{\text{BCR}}(v) > 0 \Leftrightarrow v \in \mathcal{V}_{\mathcal{F}}$, since a single broken component can cause the failure, regardless of the total number of components in the system. Of course, dealing with a full series system is not a difficulty in practice (and rarely constitutes an example of a highly reliable system). This example mainly serves to show that β^{BCR} does not always reflect a true degree of damage to the system.

Spent fuel pool system

We now introduce an example of a complex industrial system. We propose a redesigned version of the system presented in [\[CHS16\]](#). It is a simplification of the real operation of the storage pools of water for spent fuel from nuclear reactors.

EXAMPLE 10 · The spent fuel pool system (SFP)

The spent nuclear fuel is stored in the bottom of the pool. The water in the pool cools the fuel and provides protection from radiation. Conversely, the fuel heats the water in the pool, which will eventually boil, evaporate and allow the fuel to damage the structure and contaminate the outside environment. All the components of the system (shown in [fig. 5.4](#)) are designed to keep the water in the pool cold enough to prevent it from evaporating. System failure is declared when the water level in the pool falls below a critical level.

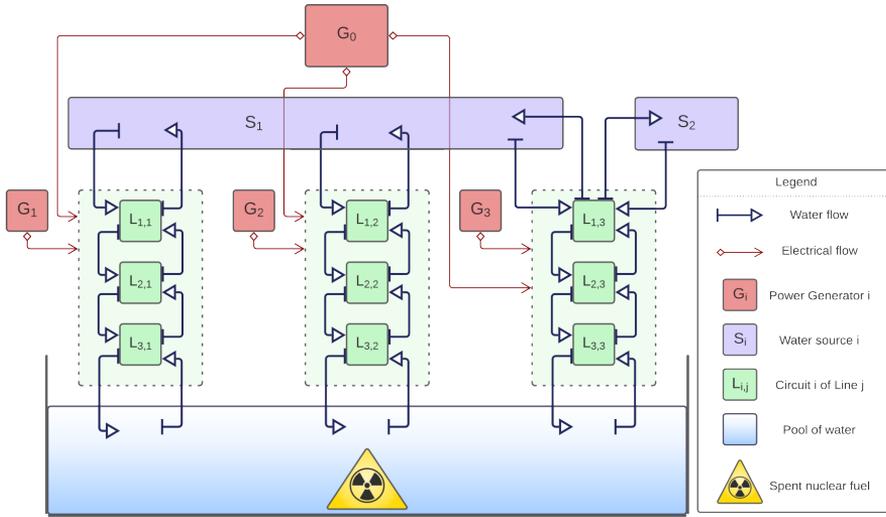


FIGURE 5.4 Representation of the spent fuel pool. The temperature of an external cold water source S_1 is transferred to the pool by means of three sealed circuits connected by heat exchangers $L_{1,1}$, $L_{2,1}$ and $L_{3,1}$ forming a line L_1 . The system has a general power supply G_0 . In the event of a problem with one of these components, the system is equipped with two other lines L_2 and L_3 identical to L_1 , an emergency diesel generator for each line G_1 , G_2 and G_3 , and a second outside water source S_2 accessible only to the third line L_3 .

The state evolution of this system can be described by a PDMP. The position corresponds to the water temperature and the water level in the pool. The regime is the combination of the status (active, inactive, broken) of the 15 components. The critical region \mathcal{F} is defined by all the states of \mathcal{X} in which the water level (second coordinate of the position) is lower or equal to a critical threshold $z_{\min}^{(2)}$. The system data allowing to compute the flow, and the different failure and repair rates of each component allowing to build the jump intensities and kernels are given in section 5.3.2.

We notice that these positions are not accessible for the flow in any given regime. This system admits a non-trivial structure function $\varphi_{\mathcal{F}}$ indicating the regimes belonging to $\mathcal{V}_{\mathcal{F}}$. The complexity of $\varphi_{\mathcal{F}}$ has a direct impact on the relevance of the score $\beta^{\text{B.C.R.}}$. For example, we can see that:

- (i) the system can fail with the failure of a very small number of components, for example the accesses to the two water sources S_1 and S_2 ,
- (ii) the system can work with a large number of failed components, for example when the combination $(G_1, G_2, G_3, S_2, L_{1,2}, L_{2,2}, L_{3,2}, L_{1,3}, L_{2,3}, L_{3,3})$ is failed.

5.2.2 Minimal cut sets and minimal path sets

The previous examples show that there are groups of components whose simultaneous malfunction leads to system failure. Such groups are called system “cut sets”.

DEFINITION 31 · Cut set and minimal cut set

A group of components forms a cut set if and only if the joint failure of these components causes the system failure. This cut set is said minimal if none of its strict subsets is another cut set. In this case we abbreviate MCS (minimal cut set).

For highly reliable complex systems, the complete failure of a cut set is always a rare event. Two scenarios can lead to the failure of a cut set:

- (i) the system has experienced a very large number of failures, statistically resulting in the failure of all the components in a cut set,
- (ii) or the system has only experienced a small number of failures, but these have precisely targeted the components in the same cut set.

The second condition of the system coherence assumption (see [definition 28](#)) implies that the set of all system components forms a cut set. This cut set is minimal only in the case of a parallel system. The proximity score $\beta^{\text{B CR}}$ marks the progression of the trajectory along this specific cut set. The associated IF $\xi_{\theta}^{\text{B CR}}$ seeks to maximize the number of broken components along the trajectory. It therefore focuses on the first scenario. By breaking components at random, it always ends up breaking an entire cut set. It does not consider the role played by each component within the system. In the real world, however, the second scenario represents a more plausible risk. It is less unlikely that few failures target the right components rather than failures could affect a large number of components without repairs intervening. We therefore wish to present another proximity score, better able to realize the second scenario. It is clear that this one can no longer be blind to the role played by the component in the system. The failure of a specific component in a specific system configuration has to be encouraged to a greater or lesser extent depending on how it interacts with the other components.

From minimal cut set to minimal path set

As mentioned before, if a regime belongs to $\mathcal{V}_{\mathcal{F}}$, then the failed components in this regime forms a cut set. The necessary information to evaluate the level of degradation of the system is contained in the respective status of each MCS (Minimal Cut Set). A good proximity score with respect to $\mathcal{V}_{\mathcal{F}}$ must therefore quantify how close each MCS is to being completely failed. Each MCS can be seen as a gauge to be filled and the proximity score monitors the state of these gauges.

The challenge is to capture the useful information provided by the state of these gauges into a single score variable. A first idea would be to only look at the state of the most advanced gauge. For example, by choosing as the proximity score the highest proportion of failed components among all MCSs. This amounts to considering each MCS as a subsystem and returning the maximum of $\beta^{\text{B}^{\text{CR}}}$ among all these subsystems. The problem is that the score does not progress if components fail outside of the most damaged MCS. However, we will see that these failures also contribute to bringing us closer to the failure.

A change of perspective gives us the key. If there are groups of components, called cut sets, whose malfunction guarantees critical system failure, the inverse also exists. There are groups of components, called “path sets”, whose proper functioning prevents system failure.

DEFINITION 32 · Path set and minimal path set

A group of components forms a path set if and only if the joint operation of these components prevents the system failure. This path set is said minimal if none of its strict subsets is an other path set. In this case we abbreviate MPS (minimal path set).

A regime in which all the components of at least one MPS are functioning does not belong to $\mathcal{V}_{\mathcal{F}}$. Conversely, if the regime belongs to $\mathcal{V}_{\mathcal{F}}$, then each MPS has at least one failed component. Thus, rather than monitoring the gauges corresponding to each MCS, we can simply monitor a single global gauge: the proportion of MPSs with at least one failed component.

Formal definition of MPS

Still using Boolean algebra, we can propose a more formal definition of MPS and MCS. Combining Theorem 1.14 and Theorem 1.21 from [CH11], we obtain the following decomposition:

THEOREM 10 · Complete disjunctive normal form

If φ is a non-decreasing Boolean function, then it admits a unique decomposition (except for the numbering of the terms) of the form:

$$\varphi(\mathbf{c}) = \bigvee_{i=1}^{d_{\varphi}} \bigwedge_{j \in J_i} c_j \quad \text{for all } \mathbf{c} \in \{0, 1\}^{d_{\mathbf{c}}}, \quad (5.9)$$

where d_{φ} is an integer and $J_1, \dots, J_{d_{\varphi}} \subset \llbracket 1, d_{\mathbf{c}} \rrbracket$ are such that $J_i \not\subset J_{i'}$ for all $i, i' \in \llbracket 1, d_{\varphi} \rrbracket$ with $i \neq i'$.

We assume from now on that the system is coherent, so we can apply [theorem 10](#) to its structure function $\varphi_{\mathcal{F}}$. There are then d_{MPS} sets of components $J_1, \dots, J_{d_{\text{MPS}}} \subset \llbracket 1, d_{\mathbf{c}} \rrbracket$ such that none of them is contained in another and such that the operation of all the components of a set causes

the operation of the system independently of the status of the components in the other sets. These sets are the unique minimal paths sets of the system. By switching to the complementary with $\bar{c} = 1 - c$ for any $c \in \{0, 1\}$, we can notice that:

$$\varphi_{\mathcal{F}}(\mathbf{c}(v)) = \bigvee_{i=1}^{d_{\text{MPS}}} \bigwedge_{j \in I_i} c_j(v) \iff \overline{\varphi_{\mathcal{F}}(\mathbf{c}(v))} = \bigwedge_{i=1}^{d_{\text{MPS}}} \bigvee_{j \in I_i} \overline{c_j(v)}. \quad (5.10)$$

Therefore the minimal paths sets can be alternatively defined such that the failure of at least one component in each set causes the system failure.

Formal definition of MCS

Let us first note that since $\varphi_{\mathcal{F}}$ is a non-decreasing Boolean function, the function $\mathbf{c}(v) \mapsto \overline{\varphi_{\mathcal{F}}(\overline{\mathbf{c}(v)})}$ is also non-decreasing (it is called the dual function of $\varphi_{\mathcal{F}}$). Indeed: let $\mathbf{c}(v) \leq \mathbf{c}'(v) \in \{0, 1\}^{d_c}$, we have $\overline{\mathbf{c}(v)} \geq \overline{\mathbf{c}'(v)}$ so $\varphi_{\mathcal{F}}(\overline{\mathbf{c}(v)}) \geq \varphi_{\mathcal{F}}(\overline{\mathbf{c}'(v)})$ and finally $\overline{\varphi_{\mathcal{F}}(\overline{\mathbf{c}(v)})} \leq \overline{\varphi_{\mathcal{F}}(\overline{\mathbf{c}'(v)})}$. We can therefore apply [theorem 10](#) to it and get a set of d_{MCS} lists of indices $\tilde{J}_1, \dots, \tilde{J}_{d_{\text{MCS}}} \subset \llbracket 1, d_c \rrbracket$ with $\tilde{J}_i \not\subset \tilde{J}_{i'}$ for any $i, i' \in \llbracket 1, d_{\text{MCS}} \rrbracket$ with $i \neq i'$ such that:

$$\overline{\varphi_{\mathcal{F}}(\overline{\mathbf{c}(v)})} = \bigvee_{i=1}^{d_{\text{MCS}}} \bigwedge_{j \in \tilde{J}_i} c_j(v) \quad \text{i.e.} \quad 1 - \varphi_{\mathcal{F}}(\mathbf{c}(v)) = \overline{\varphi_{\mathcal{F}}(\overline{\mathbf{c}(v)})} = \bigvee_{i=1}^{d_{\text{MCS}}} \bigwedge_{j \in \tilde{J}_i} \overline{c_j(v)}. \quad (5.11)$$

These d_{MCS} sets of components $\tilde{J}_1, \dots, \tilde{J}_{d_{\text{MCS}}}$ are the unique minimal cuts sets of the system. If all the components of a MCS are broken then this causes (from the static point of view) the system failure. And again by switching to the complementary, if at least one component per MCS works, the system failure is prevented:

$$\overline{\varphi_{\mathcal{F}}(\mathbf{c}(v))} = \bigvee_{i=1}^{d_{\text{MCS}}} \bigwedge_{j \in \tilde{J}_i} \overline{c_j(v)} \iff \varphi_{\mathcal{F}}(\mathbf{c}(v)) = \bigwedge_{i=1}^{d_{\text{MCS}}} \bigvee_{j \in \tilde{J}_i} c_j(v). \quad (5.12)$$

5.2.3 Proximity score and implementation

We propose the following proximity score based on the proportion of MPSs with at least one broken component.

DEFINITION 33 · MPS-based proximity score

We define the proximity score β^{MPS} as the square of the proportion of MPSs with at least one broken component. Let $J_1, \dots, J_{d_{\text{MPS}}} \subseteq \llbracket 1, d_c \rrbracket$ be the d_{MPS} MPSs of the system.

$$\beta^{\text{MPS}} : v \in \mathcal{V} \mapsto \left(\frac{1}{d_c} \sum_{i=1}^{d_{\text{MPS}}} \bigvee_{j \in J_i} \overline{c_j(v)} \right)^2. \quad (5.13)$$

We also note $\xi_{\theta}^{\text{MPS}}$ the IF defined for any $\theta \in \Theta$ and basis functions $(\phi_j)_j$ by:

$$\xi_{\theta}^{\text{MPS}}((z, v)) = \exp \left[\sum_{j=1}^{d_{\Theta}} \theta_j \times \phi_j(\beta^{\text{MPS}}(v)) \right]. \quad (5.14)$$

We still have $\beta^{\text{MPS}} \in [0, 1]$. In contrast to β^{BCR} , we have $\beta^{\text{MPS}}(v) = 1 \Leftrightarrow v \in \mathcal{V}_{\mathcal{F}}$. The square plays here the same role as for β^{BCR} . Implementing this proximity score requires the MPS of the system to be determined before the simulation phase.

MPS/MCS on test cases

A series system only works if all its components are functioning and fails as soon as a component is broken. A series system with d_c components therefore has a unique MPS containing all the components of the system, and d_c MCSs, one per component.

Conversely, a parallel system works as long as at least one component is functioning and fails as soon as all the components are broken. A parallel system with d_c components therefore has d_c MPSs, one per component, and a unique MCS containing all the components of the system. We can deduce that for parallel systems, we have $\beta^{\text{BCR}} = \beta^{\text{MPS}}$ and therefore $\xi_{\theta}^{\text{BCR}} = \xi_{\theta}^{\text{MPS}}$.

The SFP example can be represented as a series/parallel diagram (see [fig. 5.5](#)) facilitating its decomposition into MPSs and MCSs. The MPSs correspond to all vertical combinations and the MCSs to all horizontal combinations.

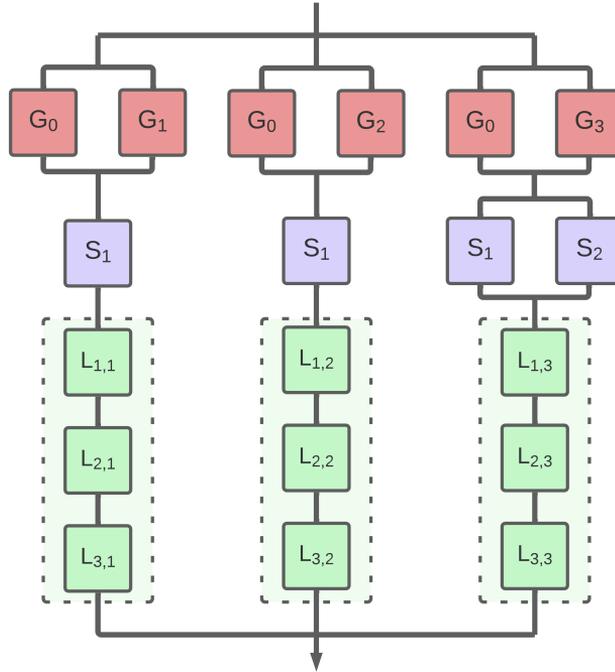


FIGURE 5.5 **Series/Parallel diagram of the spent fuel pool of Figure 5.4.**

There are 8 MPS and 69 MCS in the SFP system. Here is the comprehensive list of its MPSs:

- | | |
|--|---|
| (i) $\{G_0, S_1, L_{1,1}, L_{2,1}, L_{3,1}\},$ | (v) $\{G_0, S_1, L_{1,3}, L_{2,3}, L_{3,3}\},$ |
| (ii) $\{G_1, S_1, L_{1,1}, L_{2,1}, L_{3,1}\},$ | (vi) $\{G_3, S_1, L_{1,3}, L_{2,3}, L_{3,3}\},$ |
| (iii) $\{G_0, S_1, L_{1,2}, L_{2,2}, L_{3,2}\},$ | (vii) $\{G_0, S_2, L_{1,3}, L_{2,3}, L_{3,3}\},$ |
| (iv) $\{G_2, S_1, L_{1,2}, L_{2,2}, L_{3,2}\},$ | (viii) $\{G_3, S_2, L_{1,3}, L_{2,3}, L_{3,3}\}.$ |

Fault tree analysis

Listing all the MPS or MCS of a system is an NP-hard problem. This task can be done by hand on the SFP system but it becomes impractical in the case of very large, highly redundant systems. The literature presents more methods to determine the MCS than the MPS of a system but the two problems are strictly equivalent. The search for the MCS of the system belongs to a field called *fault tree analysis* (FTA). Fault trees are the most common representation of systems in the static Boolean approach, but FTA is not restricted to the fault tree representation. This is an old but still active field in the industrial and academic communities. A complete and relatively recent

survey of the field is proposed in [RS15]. In particular, the interested reader can refer to [RS15, Table 1] for a list of methods (with or without dedicated modelling and calculation tools) that can be used to determine MCSs.

New approaches offer other perspectives on the decomposition of the structure function, for example based on the differential logic calculus [Rus+22; Bri+21] or quantum methods [SD24]. We should also mention [RA21], which tackles the derivation of MCSs from the list of MPSs. FTA also offers tools to construct quantitative measures of system reliability such as importance indices for the components or approximations of the probability of failure [Rau93; Lee+85; Ćepi1]. The construction of importance functions proposed in [TPZ16] to apply a RESTART method to hybrid systems is very close to our philosophy (see also [BS20] for the same idea with a splitting method for dynamic fault trees).

5.3 Numerical experiments

In this section we present the results obtained with the AIS method described in chapter 4, for the two families of IF $\xi_{\theta}^{\text{BCR}}$ and $\xi_{\theta}^{\text{MPS}}$. We compare the performances of our method to a CMC (classical Monte Carlo) method first on series and parallel systems (see example 9), and then on the spent fuel pool system described in example 10.

5.3.1 Series and parallel systems

We study series and parallel systems with d_c components. We set $\mathcal{V} = \{0, 1\}^{d_c}$. The regime of the system is $v = (v^{(1)}, \dots, v^{(d_c)})$ where for $j \in \{1, \dots, d_c\}$, the status of the j -th component $v^{(j)} = 1$ if the component is active and 0 if it is broken. The regime at time $t \geq 0$ thus corresponds to the current status of each component: $V_t = (V_t^{(1)}, \dots, V_t^{(d_c)})$. Therefore in the case of series systems $V_t \in \mathcal{V}_{\mathcal{F}}$ if there is $j \in \llbracket 1, d_c \rrbracket$ such that $V_t^{(j)} = 0$, and in the case of parallel systems $V_t \in \mathcal{V}_{\mathcal{F}}$ if $V_t^{(j)} = 0$ for any $j \in \llbracket 1, d_c \rrbracket$.

Flow and jump intensity

The system failure is reached either as soon as the process has spent a total time larger than $z_{\max}^{(1)}$ in $\mathcal{V}_{\mathcal{F}}$ (global grace period), or when it remains in $\mathcal{V}_{\mathcal{F}}$ a time larger than $z_{\max}^{(2)}$ without leaving it (local grace period). We note $Z_t = (Z_t^{(1)}, Z_t^{(2)}, Z_t^{(3)})$ the position of the process at time $t \geq 0$ with $Z_t^{(1)}$ the total time spent in $\mathcal{V}_{\mathcal{F}}$ during the entire trajectory, $Z_t^{(2)}$ the elapsed time since the entry in $\mathcal{V}_{\mathcal{F}}$ if the process is there and 0 otherwise, and finally $Z_t^{(3)} = t$ the total elapsed time. For an initial time $t_0 > 0$ and a departure state $X_{t_0} = (Z_{t_0}, V_{t_0})$, the flow of the PDMP is given by

$\Psi(\cdot | X_{t_0}) : h \mapsto (Z_{t_0+h}, V_{t_0+h})$ with:

$$\begin{aligned} Z_{t_0+h}^{(1)} &= Z_{t_0}^{(1)} \mathbb{1}_{V_{t_0} \notin \mathcal{V}_{\mathcal{F}}} + (Z_{t_0}^{(1)} + h) \mathbb{1}_{V_{t_0} \in \mathcal{V}_{\mathcal{F}}}, \\ Z_{t_0+h}^{(2)} &= (Z_{t_0}^{(2)} + h) \mathbb{1}_{V_{t_0} \in \mathcal{V}_{\mathcal{F}}}, \\ Z_{t_0+h}^{(3)} &= t_0 + h. \end{aligned} \tag{5.15}$$

Under distribution \mathbf{p} , for $j \in \llbracket 1, d_c \rrbracket$, the j -th component has a marginal jump rate $\lambda^{(j)}$ that depends on its status (in other words it has a constant failure rate and a constant repair rate). From the state x^- , the next jump occurs at a random time of jump intensity $\lambda(x^-) = \sum_{j=1}^{d_c} \lambda^{(j)}(x^-)$. At each jump from state x^- , only one component is randomly selected with probability $\lambda^{(j)}(x^-)/\lambda(x^-)$ for $j \in \llbracket 1, d_c \rrbracket$ and it then changes status. The marginal jump rates of each system component are given in [table 5.1](#). Note that we have $\xi_{\theta}^{\text{B CR}} = \xi_{\theta}^{\text{MPS}}$ for parallel systems since a parallel system has d_c MPS containing each 1 component.

TABLE 5.1 Marginal jump intensity of each component for the **series** and **parallel** systems.

Component	$\lambda^{(j)}$ for Series system	$\lambda^{(j)}$ for Parallel system
c_1	$1 \times 10^{-9} \mathbb{1}_{v^{(1)}=0} + 1 \times 10^{-6} \mathbb{1}_{v^{(1)}=1}$	$6 \times 10^{-5} \mathbb{1}_{v^{(1)}=0} + 1 \cdot 10^{-4} \mathbb{1}_{v^{(1)}=1}$
c_2	$5 \times 10^{-9} \mathbb{1}_{v^{(2)}=0} + 5 \times 10^{-6} \mathbb{1}_{v^{(2)}=1}$	$2 \times 10^{-4} \mathbb{1}_{v^{(2)}=0} + 5 \times 10^{-4} \mathbb{1}_{v^{(2)}=1}$
c_3	$5 \cdot 10^{-9} \mathbb{1}_{v^{(3)}=0} + 1 \times 10^{-6} \mathbb{1}_{v^{(3)}=1}$	$2 \times 10^{-4} \mathbb{1}_{v^{(3)}=0} + 1 \times 10^{-3} \mathbb{1}_{v^{(3)}=1}$
c_4	$1 \times 10^{-9} \mathbb{1}_{v^{(4)}=0} + 5 \times 10^{-6} \mathbb{1}_{v^{(4)}=1}$	$6 \times 10^{-5} \mathbb{1}_{v^{(4)}=0} + 5 \times 10^{-4} \mathbb{1}_{v^{(4)}=1}$
c_5	$8 \times 10^{-9} \mathbb{1}_{v^{(5)}=0} + 6 \times 10^{-6} \mathbb{1}_{v^{(5)}=1}$	$5 \times 10^{-4} \mathbb{1}_{v^{(5)}=0} + 8 \times 10^{-4} \mathbb{1}_{v^{(5)}=1}$

Results for series and parallel systems

We compare on a series and on a parallel system the performance of a CMC method to our AIS method with the two families of IF $\xi_{\theta}^{\text{B CR}}$ and $\xi_{\theta}^{\text{MPS}}$. The sample size of the CMC method is ranging from 10^5 to 10^7 . The sample size of the AIS method is ranging from 10^3 to 10^4 . For both the series and parallel systems, we generated trajectories of duration $s_{\max} = 1500$ with global grace period $z_{\max}^{(1)} = 75$ and local grace period $z_{\min}^{(2)} = 50$. Each system has five components. The jump parameters of the two systems are described in [table 5.1](#). The results obtained on the series system, resp. parallel system, are described in [table 5.2](#), resp. in [table 5.3](#).

The AIS method is initialized here with the procedure proposed in [section 4.3.3](#). We choose the smallest θ (in norm) such that the probability that at least one component failure occurs before the end of the simulation is larger than $1/3$. We set $d_{\theta} = 5$. At each iteration, we generate trajectories until we have $n_{\text{CE}} = 10$ failures before updating θ for $N_L = 10^3$ and $n_{\text{CE}} = 50$ for $N_L = 10^4$. We stop when the total budget N_L is reached. The estimated probability \widehat{L}_{N_L} is given

TABLE 5.2 **Results on the series system case** (with jump rates of the five components from table 5.1).

Method	N_L	$\hat{\mathcal{I}}_{N_L}$	C.o.v $\frac{\hat{\sigma}_{N_L}}{\hat{\mathcal{I}}_{N_L}}$	95% confidence interval
CMC	10^5	4×10^{-5}	158	$[8.02 \times 10^{-7}; 7.92 \times 10^{-5}]$
	10^6	2.9×10^{-5}	186	$[1.84 \times 10^{-5}; 3.96 \times 10^{-5}]$
	10^7	2.7×10^{-5}	192	$[2.38 \times 10^{-5}; 3.02 \times 10^{-5}]$
IS with $\xi_{\theta}^{\text{BCR}}$	10^3	2.96×10^{-5}	0.76	$[2.82 \times 10^{-5}; 3.09 \times 10^{-5}]$
	10^4	2.89×10^{-5}	0.78	$[2.84 \times 10^{-5}; 2.93 \times 10^{-5}]$
IS with $\xi_{\theta}^{\text{MPS}}$	10^3	2.82×10^{-5}	0.80	$[2.68 \times 10^{-5}; 2.96 \times 10^{-5}]$
	10^4	2.91×10^{-5}	0.77	$[2.87 \times 10^{-5}; 2.95 \times 10^{-5}]$

by eq. (4.30), the estimated asymptotic standard deviation $\hat{\sigma}_{N_L}$ is the square root of the variance estimator given in eq. (4.50) and the confidence interval formula is given by theorem 9. For the CMC method, we simply generate N_L trajectories and we count the proportion of faulty trajectories.

The AIS method performs better than the CMC method in all configurations. The estimated probabilities are of the same order and the confidence intervals produced by the AIS method for a given sample size are of comparable length to the confidence intervals produced by the CMC method for a sample size 10^4 larger. It can be seen that the best performance is obtained on the series system despite a slightly smaller failure probability. This result is not surprising since the failed trajectories of a series system generally contain few jumps and thus produce likelihood ratios that are easier to stabilize. The $\xi_{\theta}^{\text{BCR}}$ and $\xi_{\theta}^{\text{MPS}}$ seem to have the same effectiveness here. Each of the two systems presents a different challenge for importance sampling. The series system requires multi-modal importance distribution since the failure can come from any component. On the other hand, the importance distribution for a parallel system must produce sequences where all components fail that are plausible from the perspective of jump times.

5.3.2 The spent fuel pool

The roles of the components $(c_j)_{j=1}^{d_c=15}$ are described in fig. 5.4. We set $\mathcal{V} = \{-1, 0, 1\}^{d_c}$. The regime of the system is $v = (v^{(1)}, \dots, v^{(d_c)})$ where for $j \in \llbracket 1, d_c \rrbracket$, the status of the j -th component $v^{(j)} = 1$ if the component is active, 0 if it is inactive and -1 if it is broken. The regime $V_t = (V_t^{(1)}, \dots, V_t^{(d_c)})$ of the process at time $t \geq 0$ corresponds to the current status of each component. Recall that we have $V_t \in \mathcal{V}_{\mathcal{F}}$ if at time $t \geq 0$ all MPS are damaged or equivalently if at least one MCS has all its components broken.

TABLE 5.3 **Results on the parallel system case** (with jump rates of the five components from table 5.1 in appendix).

Method	N_L	\hat{T}_{N_L}	C.o.v $\frac{\hat{\sigma}_{N_L}}{\hat{T}_{N_L}}$	95% confidence interval
CMC	10^5	8×10^{-5}	111.75	$[2.46 \times 10^{-5}; 1.35 \times 10^{-4}]$
	10^6	6.7×10^{-5}	112.54	$[5.10 \times 10^{-5}; 8.30 \times 10^{-5}]$
	10^7	6.73×10^{-5}	121.84	$[6.22 \times 10^{-5}; 7.24 \times 10^{-5}]$
IS with $\xi_{\theta}^{\text{BCR}} = \xi_{\theta}^{\text{MPS}}$	10^3	4.85×10^{-5}	3.65	$[3.76 \times 10^{-5}; 5.94 \times 10^{-5}]$
	10^4	5.80×10^{-5}	4.36	$[5.29 \times 10^{-5}; 6.31 \times 10^{-5}]$

Flow and jump intensity

We note $Z_t = (Z_t^{(1)}, Z_t^{(2)}, Z_t^{(3)})$ the position of the process at time $t \geq 0$ with $Z_t^{(1)}$ the temperature of the water in the pool in °C, $Z_t^{(2)}$ the water level in the pool in meters (m) and $Z_t^{(3)} = t$ the total elapsed time. The evolution of these variables is described by the system of ordinary differential equations:

$$\begin{aligned}
 \frac{dZ_t^{(1)}}{dt} &= \mathbb{1}_{Z_t^{(1)} < 100} \times \frac{r + \rho CK (z_S^{(1)} - Z_t^{(1)}) \mathbb{1}_{V_t \neq V_x}}{\rho CA Z_t^{(2)}}, \\
 \frac{dZ_t^{(2)}}{dt} &= -\mathbb{1}_{Z_t^{(1)} = 100} \times \frac{r}{\rho CA \ell}, \\
 \frac{dZ_t^{(3)}}{dt} &= 1,
 \end{aligned} \tag{5.16}$$

where the physical parameters are given in the table 5.4 (values taken from [CHS16]).

Under distribution \mathbf{p} , each component $v^{(j)}$ has a marginal jump rate $\lambda^{(j)}$ which depends on its status and on the values of the physical variables of the system. The jump intensity of the PDMP in a state $x \in \mathcal{X}$ is the sum of the jump rates of the components in state x : $\lambda(x) = \sum_{j=1}^{d_c} \lambda^{(j)}(x)$. At each jump from state x^- , a component c_j is randomly selected with probability $\lambda^{(j)}(x^-) / \lambda(x^-)$ and changes status (it is repaired if it was broken, and fails otherwise). The system automatically reconfigures itself by enabling or disabling components so that exactly 1 MPS has all its components active if possible (be careful not to confuse inactive component $v^{(j)} = 0$ and broken component $v^{(j)} = -1$).

TABLE 5.4 **Physical parameters** of the SFP. Values taken from [CHS16].

Physical parameters	Value	Description
r	$2.106 \times 10^{10} \text{ J} \cdot \text{h}^{-1}$	Residual power of the fuel
C	$4180 \text{ J} \cdot \text{kg}^{-1} \cdot \text{K}^{-1}$	Mass heat capacity
ρ	$990 \text{ kg} \cdot \text{m}^{-3}$	Density of the water
A	77 m^2	Area of the pool.
$z_S^{(1)}$	15°C	Temperature of the water sources
K	$550 \text{ m}^3 \cdot \text{h}^{-1}$	The debit water
ℓ	$2.257 \times 10^6 \text{ J} \cdot \text{kg}^{-1}$	Latent heat of vaporization
s_{\max}	3600 h	Duration of the mission
$z_0^{(2)}$	19 m	Initial level of water in the pool
$z_{\min}^{(2)}$	16 m	Critical threshold of the level of water in the pool

 TABLE 5.5 Marginal jump intensity of each component for the **standard SFP case**.

Component	Marginal jump intensity $\lambda^{(j)}$ for $j = 1, \dots, d_c$			
	c_i	when $v^{(i)} = -1$	when $v^{(i)} = 0$	when $v^{(i)} = 1$
$c_1 = \mathbf{G}_0$		$4 \cdot 10^{-2}$	4×10^{-6}	$6 \cdot 10^{-6}$
$c_{i+1} = \mathbf{G}_i, i = 1, 2, 3$		8×10^{-2}	2×10^{-6}	30×10^{-6}
$c_5 = \mathbf{S}_1$		1×10^{-2}	4×10^{-6}	20×10^{-6}
$c_6 = \mathbf{S}_2$		3×10^{-2}	1×10^{-6}	$5 \cdot 10^{-6}$
$c_{6+i} = \mathbf{L}_{i,1}, i = 1, 2, 3$		$(6 - 0.03Z_t^{(1)}) \times 10^{-2}$	$(1 + 0.05Z_t^{(1)}) \times 10^{-6}$	$(3 + 0.1Z_t^{(1)}) \times 10^{-6}$
$c_{9+i} = \mathbf{L}_{i,2}, i = 1, 2, 3$		$(6 - 0.03Z_t^{(1)}) \times 10^{-2}$	$(1 + 0.05Z_t^{(1)}) \times 10^{-6}$	$(3 + 0.1Z_t^{(1)}) \times 10^{-6}$
$c_{12+i} = \mathbf{L}_{i,3}, i = 1, 2, 3$		$(6 - 0.03Z_t^{(1)}) \times 10^{-2}$	$(1 + 0.05Z_t^{(1)}) \times 10^{-6}$	$(3 + 0.1Z_t^{(1)}) \times 10^{-6}$

It is assumed that no water can be re-injected into the pool in case of evaporation for the duration of the mission s_{\max} . Once in \mathcal{V}_F there is a first grace period before the temperature of the water reaches 100°C , but this temperature can go back down once the system is repaired. Then we have a second grace period before the water level in the pool reaches a critical threshold $z_{\min}^{(2)}$. In our model, the evaporated water is lost and the water level cannot rise again if the system is repaired.

Results for the spent fuel pool

We carry out three series of numerical simulations on the spent fuel pool system.

- (i) We first compare the performance of each version of our AIS method to a CMC method on a standard case with jump rates described in table 5.5, results described in table 5.7 and a probability of system failure about 10^{-5} . The sample size of the CMC method is

TABLE 5.6 Marginal jump intensity of each component for the **extreme SFP case**.

Component	Marginal jump intensity $\lambda_0^{(j)}$ for $j = 1, \dots, d_c$			
	c_i	when $v^{(i)} = -1$	when $v^{(i)} = 0$	when $v^{(i)} = 1$
$c_1 = G_0$		4×10^{-2}	4×10^{-6}	$6 \cdot 10^{-6}$
$c_{i+1} = G_i, i = 1, 2, 3$		10×10^{-2}	15×10^{-6}	30×10^{-6}
$c_5 = S_1$		$1 \cdot 10^{-2}$	4×10^{-6}	$20 \cdot 10^{-6}$
$c_6 = S_2$		$3 \cdot 10^{-2}$	1×10^{-6}	$5 \cdot 10^{-6}$
$c_{6+i} = L_{i,1}, i = 1, 2, 3$		$(12 - 0.04Z_t^{(1)}) \times 10^{-2}$	$(1 + 0.1Z_t^{(1)}) \times 10^{-6}$	$(3 + 0.1Z_t^{(1)}) \times 10^{-6}$
$c_{9+i} = L_{i,2}, i = 1, 2, 3$		$(12 - 0.04Z_t^{(1)}) \times 10^{-2}$	$(1 + 0.1Z_t^{(1)}) \times 10^{-6}$	$(3 + 0.1Z_t^{(1)}) \times 10^{-6}$
$c_{12+i} = L_{i,3}, i = 1, 2, 3$		$(15 - 0.05Z_t^{(1)}) \times 10^{-2}$	$(1 + 0.08Z_t^{(1)}) \times 10^{-6}$	$(3 + 0.08Z_t^{(1)}) \times 10^{-6}$

ranging from 10^5 to 10^7 . The sample size of the AIS method is ranging from 10^3 to 10^4 .

- (ii) We then check the stability of the best version of our method which seems to be based on $\xi_{\theta}^{\text{MPS}}$. We represent in [fig. 5.6](#) 50 confidence intervals at 95% level obtained with the AIS-MPS method with a sample size of 10^3 trajectories still on the standard case ([table 5.5](#)) and we compare them to the confidence interval obtained with the CMC method and a sample size of 10^7 .
- (iii) Since the method is stable, we can trust the confidence intervals produced and confront it with even rarer events for which it cannot be compared to a CMC method. Therefore we test the AIS-MPS method on an extreme case with jump rates described in [table 5.6](#), results described in [table 5.8](#) and a probability of system failure about 10^{-7} . The sample size of the AIS method is ranging from 10^3 to 10^4 .

The AIS method is initialized with the procedure proposed in [section 4.3.3](#). We choose the smallest θ (in norm) such that the probability that at least one component failure occurs before the end of the simulation is larger than $1/3$. We set $d_{\theta} = 8$. At each iteration, we generate trajectories until we have $n_{\text{CE}} = 10$ failures before updating θ for $N_L = 10^3$ and $n_{\text{CE}} = 50$ for $N_L = 10^4$. We stop when the total budget N_L is reached. The estimated probability \hat{L}_{N_L} is given by [eq. \(4.30\)](#), the estimated asymptotic standard deviation $\hat{\sigma}_{N_L}$ is the square root of the variance estimator given in [eq. \(4.50\)](#) and the confidence interval formula is given by [theorem 9](#). For the CMC method, we simply generate N_L trajectories and we count the proportion of faulty trajectories.

The first observation on [table 5.7](#) is that even the BC-IF method, which does not distinguish the role of each component, manages to drastically reduce the variance of the estimator compared to the CMC method (almost by a factor of 1000). The AIS-MPS method is extremely efficient with a variance reduction of 10^4 . Note that the coefficient of variation in [table 5.7](#), which can be used as an indicator of the performance of an estimator, does not necessarily decrease with the

TABLE 5.7 Results on the standard SFP case (with jump rates from table 5.5).

Method	N_L	\hat{I}_{N_L}	C.o.v $\frac{\hat{\sigma}_{N_L}}{\hat{I}_{N_L}}$	95% confidence interval
CMC	10^5	2×10^{-5}	223.60	$[0; 4.77 \times 10^{-5}]$
	10^6	1.3×10^{-5}	277.35	$[5.93 \times 10^{-6}; 2.01 \times 10^{-5}]$
	10^7	1.77×10^{-5}	237.68	$[1.51 \times 10^{-5}; 2.03 \times 10^{-5}]$
AIS with ξ_{θ}^{BCR}	10^3	2.16×10^{-5}	10.88	$[7.05 \times 10^{-6}; 3.63 \times 10^{-5}]$
	10^4	1.79×10^{-5}	16.82	$[1.37 \times 10^{-5}; 2.22 \times 10^{-5}]$
AIS with ξ_{θ}^{MPS}	10^3	2.19×10^{-5}	3.01	$[1.78 \times 10^{-5}; 2.60 \times 10^{-5}]$
	10^4	1.99×10^{-5}	1.01	$[1.96 \times 10^{-5}; 2.03 \times 10^{-5}]$

TABLE 5.8 Results with the AIS-MPS approximation on the extreme SFP case (with jump rates from table 5.6). Same method as in Table 5.7 except that the initialization follows the method described in section 4.3.3 with the smallest θ (in norm) such that the probability that at least one component failure occurs before the end of the simulation is larger than 0.9.

Method	N_L	\hat{I}_{N_L}	C.o.v $\frac{\hat{\sigma}_{N_L}}{\hat{I}_{N_L}}$	95% confidence interval
AIS with $\xi_{\theta}^{(MPS)}$	10^3	3.31×10^{-7}	3.35	$[2.63 \times 10^{-7}; 4.00 \times 10^{-7}]$
	10^4	3.83×10^{-7}	3.29	$[3.58 \times 10^{-7}; 4.08 \times 10^{-7}]$

sample size. This is due to the fact that in the first iterations of the method, the parameter θ is not yet well chosen and that a poor importance distribution in high dimension tends to produce too small likelihood ratios. The coefficient of variation is underestimated at this time.

Figure 5.6 confirms the performance of the AIS-MPS method. The majority of the confidence intervals produced by the AIS method with a sample size of 10^3 are shorter than the confidence interval produced by the CMC method with a sample size of 10^7 . Only 1 interval out of 50 is significantly larger than the interval produced by CMC, but remains relevant since it gives a probability of failure between 1×10^{-5} and 3.5×10^{-5} . We deduce that the AIS-MPS method is robust and that we can therefore have faith in its estimates.

Finally, we see on the table 5.8 that the AIS-MPS method still offers excellent performances for a 100 times rarer event. A reliable estimate of the probability that is of order 10^{-7} can be obtained with a sample size smaller than 10^4 .

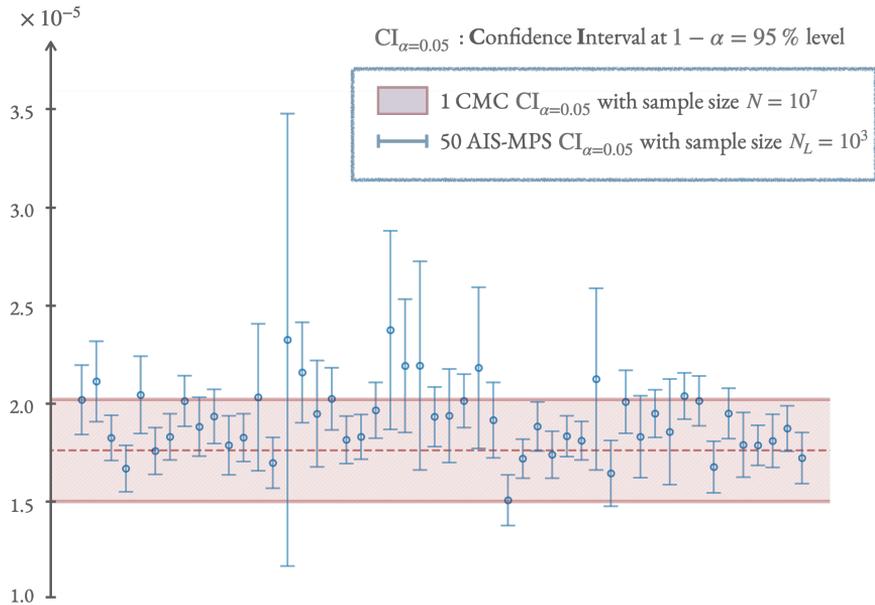


FIGURE 5.6 **Comparison of 50 confidence intervals at 95% level obtained with AIS-MPS approximation.** Each confidence interval corresponds to a run of the AIS-MPS method on the standard case of the SFP (table 5.5) with 10^3 trajectories (same conditions as for table 5.7). They are compared to the confidence intervals obtained with the CMC method on 10^7 trajectories.

CHAPTER 6

Graph-informed importance function

An adaptive importance sampling strategy has been proposed in [chapter 4](#) for rare event simulation of PDMPs. The importance distribution is built upon an approximation of the committor function of the process. This approximation, called importance function, is a parametric transformation of a proximity score β between any regime of the PDMP and the set $\mathcal{V}_{\mathcal{F}}$ of regimes allowing to access the critical region \mathcal{F} . The proximity score β^{MPS} , proposed in [chapter 5](#), has proven to be very effective when the PDMP models a coherent multicomponent industrial system. In this chapter, we aim to go further and propose a proximity score that is both more general and more accurate. The evolution of the PDMP regime can be roughly summarized by a random walk $(V_t)_t$ on a graph whose regimes are the vertices. The mean hitting time of the region $\mathcal{V}_{\mathcal{F}}$ for the random walk starting from vertex v is a natural proximity score for v . This mean hitting time is explicit when the random walk is Markovian and time-homogeneous. The random walk $(V_t)_t$ given by the PDMP being typically non-Markovian, we compute these mean hitting times for a simplified random walk $(Y_t)_t$ approximating $(V_t)_t$. Our importance sampling method is implemented with this proximity score on the final test case of the previous chapter: the spent fuel pool.

Content of this chapter

6.1	Motivations	132
6.1.1	Broadening the scope of applications	132
6.1.2	Regime as a random walk on a graph	133
6.2	Graph-based mean hitting times	136
6.2.1	Explicit computation of mean hitting times	137
6.2.2	Mean hitting times as proximity scores	142
6.3	Numerical experiments	144
6.3.1	Computation of mean hitting times	145
6.3.2	Application on the spent fuel pool	146

6.1 Motivations

The proximity score β^{MPS} , as a reminder, corresponds to the proportion of minimal path sets (MPSs) of the system having at least one broken component in a given regime. It is thus suitable for estimating the failure probability of coherent multi-component industrial systems.

6.1.1 Broadening the scope of applications

Our aim is not to limit ourselves to these systems. We wish to propose a relevant proximity score for the widest possible variety of applications.

Non-coherent systems

The reliability literature contains several proposals for extending fault tree analysis theory to non-coherent systems [IS18; BA18; Ali+17]. The thesis [Bee02], although dated, provides a useful introduction to essential fault tree analysis concepts and their extension to non-coherent systems. The previous chapter could undoubtedly be rethought in light of this work, but we have adopted a different, more general perspective.

Multi-state systems

We recall that for any regime $v \in \mathcal{V}$, we denote $\mathbf{c}(v) \in \{0, 1\}^d$ the failed or intact (0 or 1) status of each component in regime v . We have so far assumed the existence of a Boolean structure function $\varphi_{\mathcal{F}}$ such that for any $v \in \mathcal{V}$, we have: $\varphi_{\mathcal{F}}(\mathbf{c}(v)) = \mathbb{1}_{v \in \mathcal{V}_{\mathcal{F}}}$. Even by relaxing the assumption of coherence (characterized by the monotonicity of $\varphi_{\mathcal{F}}$), we have therefore assumed that the binary status of each component (failed or intact) was sufficient to determine whether a regime belongs to $\mathcal{V}_{\mathcal{F}}$. However, there exist systems, known as multi-state systems, where neither the components status nor their image by the structure function are Boolean variables. This is typically the case when multiple degradation levels are possible [SL13]. This is again a topic covered in the reliability literature [Y12; Sed+21] but it is not the path we have followed.

Beyond industrial reliability

A meticulous unification of the existing refinements of fault tree analysis theory would certainly allow for the construction of a versatile proximity score. However, this would still be limited to considering only the failure of a multi-component industrial system as the sole rare event of interest. The methodology and form of IF that we proposed in [chapter 4](#) do not constrain us this way. The PDMP can model anything else, and we could always describe the rare event of interest as reaching a region \mathcal{F} before a fixed date. Other rare/catastrophic events realized by PDMPs are considered in the literature, such as the extinction of a population [GHL22] or the ruin of an insurance company [Kri+19]. In [section 4.1.2](#), we provided arguments to justify that this

event always consisted of reaching or staying long enough in a given set of regimes $\mathcal{V}_{\mathcal{F}}$. We simply assume in the sequel that the set of regimes \mathcal{V} is finite and that we know (or can easily determine in practice) the subset $\mathcal{V}_{\mathcal{F}}$.

6.1.2 Regime as a random walk on a graph

Our ambition is therefore to propose a family of proximity scores that is independent of what the PDMP models (and more specifically, of what the regime represents in the model). Let us return to a simple explanation of the PDMP behavior. Recall that the state space of the PDMP is of the form:

$$\mathcal{X} = \bigcup_{v \in \mathcal{V}} \mathcal{X}_v = \bigcup_{v \in \mathcal{V}} (\mathcal{Z}_v \times \{v\}) = \bigcup_{v \in \mathcal{V}} (\{z \in \mathcal{Z} \mid (z, v) \in \mathcal{X}\} \times \{v\}). \quad (6.1)$$

PDMP as a random walk in a multiverse

One possible interpretation is to describe each subspace \mathcal{Z}_v for any $v \in \mathcal{V}$ as a small universe with its own physical laws (its deterministic flow ψ_v). Reaching the boundary of a universe triggers a jump to another randomly chosen universe (according to the kernel Q) with new physical laws. Jumps to other universes can also occur spontaneously. Some universes and some regions of each universe are more likely than others to trigger these random jumps (susceptibility measured by the jump intensity λ). The critical region \mathcal{F} is the union of regions located in different rarely visited universes $\mathcal{V}_{\mathcal{F}}$. A PDMP is therefore a random walk in this multiverse. We seek to quantify the proximity of each universe to the universes indexed by regimes of $\mathcal{V}_{\mathcal{F}}$.

To come back to a less metaphorical interpretation, it is easy to see that the evolution of the PDMP regime over time $(V_t)_t$, is a non-Markovian random walk on a graph. The vertices of the graph are the different regimes of \mathcal{V} and a directed edge exists from regime v to regime v' if and only if there exist two positions $z, z' \in \mathcal{Z}$ such that $q((z', v') \mid (z, v)) > 0$. This walk is indeed non-Markovian, as knowing only the current regime (i.e., the universe in which we are located) does not suffice to determine the distribution of the next jump time and the next destination regime/universe. This random walk is illustrated in [fig. 6.1](#).

Quantifying the proximity of a regime to the set $\mathcal{V}_{\mathcal{F}}$ therefore amounts to quantifying the proximity of a vertex of the graph to the subset of vertices $\mathcal{V}_{\mathcal{F}}$.

Shortest path distance on a graph

The usual distance between two vertices of a graph is the length of the shortest path (in terms of the number of edges) that separates them¹. A decreasing function of the length of this shortest path is a natural candidate for a proximity score. In reality, such a proximity score is not very

¹Since the graph is directed, we are talking here about the shortest path starting from a given vertex to any vertex of the set $\mathcal{V}_{\mathcal{F}}$

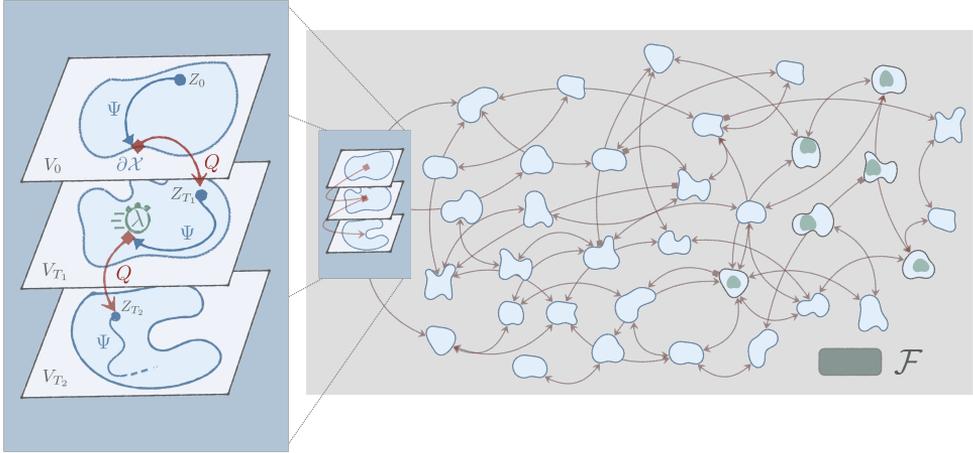


FIGURE 6.1 PDMP as a random walk in a “multiverse”.

expressive, as it only exploits a small part of the available information². This is quite apparent in the case of coherent multi-component systems from the previous chapter. The length of the shortest path then corresponds to the minimum number of possible transitions before a MCS is completely failed, or equivalently before all MPSs are damaged.

In practice, only the information carried by the most degraded MCS (or equivalently, by the shortest sequence of failures that can damage the remaining functional MPS) is used. The additional failure of a component that does not belong to the most degraded MCS does not bring us closer to the region of interest from the perspective of this particular distance. A proximity score aiming to generalize β^{MPS} must therefore at least take into account all the ways to damage the remaining MPS (and therefore the information carried by all the MCS and not only the most damaged one). A generalization of the notions of MPS and MCS on a graph is possible.

System cut sets are graph paths and system path sets are graph cuts

We need to assume the existence of an initial regime v_0 that is furthest from the region $\mathcal{V}_{\mathcal{F}}$. In the case of coherent industrial systems, the initial regime corresponds to the regime with no failed components. We will see that the cut sets of the system translate into paths on the graph, and conversely, the path sets of the system translate into cuts of the graph.

Let us start by generalizing the MCSs. In the reliability framework, a cut set is a set of components whose joint and sustained failure guarantees the critical failure of the system. This cut

²The level of connection between two people in a network is not measurable only by the minimum number of handshakes that separates them (and which is always relatively low, see https://en.wikipedia.org/wiki/Six_degrees_of_separation). Two people are close if they can be connected by multiple combinations of short sequences of handshakes.

set encodes several regimes belonging to $\mathcal{V}_{\mathcal{F}}$. Viewed dynamically, a single cut set corresponds to several possible transitions sequences leading from v_0 to $\mathcal{V}_{\mathcal{F}}$. It is therefore a set of paths³ on the graph between v_0 and $\mathcal{V}_{\mathcal{F}}$. The cut sets of the system describe the set of all the paths on the graph leading from v_0 to $\mathcal{V}_{\mathcal{F}}$, and the MCSs of the system describe the subset of these paths on the graph that do not pass through the same vertex twice (no cycle). A possible proximity score for a regime v would be an increasing function of the progress of the regime on each of these paths on the graph. However, we face the problem already mentioned in [chapter 5](#): this involves monitoring and combining many gauges (here, the paths on the graph) in a single proximity score.

Let us now generalize the MPSs. In the reliability framework, a path set is a set of components to be kept intact to ensure that the regime belongs to $\mathcal{V}_{\mathcal{F}}$. Viewed dynamically, a single path corresponds to several possible regimes through which the process must not pass. This is therefore what is called an s - t cut (or s - t separator) on the graph : a set of vertices separating the set \mathcal{V} into two disjoint subgraphs, one containing the source v_0 and the other containing the sink $\mathcal{V}_{\mathcal{F}}$. The path sets of the system correspond to the set of all s - t cuts on the graph between the source v_0 and the sink $\mathcal{V}_{\mathcal{F}}$. The MPSs of the system correspond to the s - t cuts that do not contain any others (these minimal s - t cuts are called bonds [[RRU24](#)]). A possible proximity score for a regime v would be a decreasing function of the number of minimal s - t cuts still separating the regime v from $\mathcal{V}_{\mathcal{F}}$. We would ultimately have only one gauge to monitor (the number of remaining s - t cuts).

This idea could even be generalized beyond graphs to any general set provided that it contains a source point and a target sink region. A point is close to the sink from the perspective of the source :

- (i) if it is well advanced on a portion of the paths leading from the source to the sink,
- (ii) or equivalently, if it has crossed a large proportion of the barriers separating the source from the sink.

Defining the set of all possible minimal barriers for our set means defining a score according to which being far away is simply having many barriers to cross.

In practice

As appealing as this construction may be, it remains very costly to implement in practice. On the one hand, it is necessary to be able to determine the set of all these s - t cuts between v_0 and $\mathcal{V}_{\mathcal{F}}$, to minimize them and to store them. On the other hand, for any given vertex v , we must either be able to quickly evaluate (during simulations) which s - t cuts for the source v_0 also separate v from $\mathcal{V}_{\mathcal{F}}$, or perform the calculation in advance for each vertex and store all the results. In the previous

³A path on a graph is a sequence of connected vertices from a starting vertex to an arrival vertex or set.

chapter, we exploited the particular structure adopted by the graph for coherent industrial systems to efficiently determine the MPSs/MCSs. The literature on the computation of s - t cuts is generally quite old (mainly from the 70s to the 90s) but contains some recent proposals [RRU24; KK20]. It also offers several avenues for generalizing this idea, such as [Con+20] for listing s - t cuts in temporal graphs (whose edges change over time). Finally, the existing work around the Max-flow min-cut theorem [GK13] allows us to consider the jump intensity as a weight on the edges of the graph (associated here with the notion of network capacity). The approach we propose manages to take into account all these elements without resorting to advanced notions of graph theory.

6.2 Graph-based mean hitting times

Many statistical problems on graphs have required the development of metrics capable of finely taking into account the connectivity of the graph (see [Che+23] for an overview of such metrics for graph sparsification). One approach that has proven fruitful is to rely on the properties of random walks on the graph. Of course, one can refer to algorithms of the PageRank type [Pag+99; Bero5] and their variants (personalized PageRank [Havo2; Lia+23], random walk with restart [TFP06], Katz similarity [Kat53; NB18], SimRank [JWo2; Zha+23; ZHZ24], etc.). Here is a survey on the basic theoretical concepts of random walks on graphs [Lov93] and a recent survey on applications and algorithms related to random walks on graphs [Xia+19].

A central notion associated with random walks on graphs is that of hitting time (or first passage time). The hitting time $h_{v,v'}$ between two vertices v and v' of a graph is the time it takes for a random walk to go from v to v' . We are generally interested in the mean hitting time (MHT) between two vertices:

$$H_{v,v'} = \mathbb{E}[h_{v,v'}]. \quad (6.2)$$

This is not a distance because the MHT is not symmetric. To obtain a distance, we often consider its symmetrized version called commute time:

$$C_{v,v'} = H_{v,v'} + H_{v',v}. \quad (6.3)$$

In the literature, the term “effective resistance” can also be found. The graph is interpreted as an electrical network with edges (and weights on edges) as resistances. The effective resistance is then the potential difference between two vertices when a unit current is injected at v and extracted from v' . It is a renormalized version of the commute time:

$$R_{v,v'} = \frac{C_{v,v'}}{d_{\mathcal{G}}}, \quad (6.4)$$

where $d_{\mathcal{G}}$ is the volume of the graph \mathcal{G} (the sum of degrees of all vertices). The links between

hitting times, commute times, effective resistance, Laplacian, and invariant measure are explained in [Zhu+23]. The links between hitting times and the mixing times of the random walk can also be found in [PS15]. Finally [CZ08] propose various explicit expressions of MHT for several kinds of graphs.

Measures of dissimilarity based on MHT are undoubtedly among the most popular measures for graph problems, after the shortest path distance (see also [Yen+08] for a family of dissimilarity measures generalizing both the shortest-path and the commute-time distances). They are used for clustering [Yen+05; QHo7; Ale+17], graph embedding [Gua98; QHo6], collaborative filtering [Bra05; KSo7; Fou+07], and many other fields. They have the advantage of being robust to noise (the MHT varies little after the deletion/addition of a small number of edges or after a small perturbation of weights on edges). The major asset of MHT is that they can be explicitly calculated in the case of a Markovian and time-homogeneous random walk.

6.2.1 Explicit computation of mean hitting times

The vast majority of the references cited above only examine the case of discrete-time random walks. Continuous time offers us greater flexibility and no particular disadvantages. We detail both continuous-time case and discrete-time case, the results of the second being deduced from the first (the first being more interesting for us).

Reminder on time-homogeneous Markov chains

Let us recall that a Markov chain only changes state at discrete points in time, called jump times. A Markov chain is said to be time-homogeneous if the distribution of its future states does not depend on the elapsed time (see [definition 7](#)). If a time-homogeneous Markov chain evolves on a finite state space \mathcal{V} with cardinality $\text{Card}(\mathcal{V}) = d_{\mathcal{V}}$, then its distribution can be fully characterized by a $d_{\mathcal{V}} \times d_{\mathcal{V}}$ matrix. We consider two types of Markov chains: continuous-time Markov chains (CTMCs) and discrete-time Markov chains (DTMCs).

- (i) Let $(Y_t)_{t \in \mathbb{R}_+}$ be a time-homogeneous CTMC on \mathcal{V} . Its jump times are random and follow an exponential distribution whose parameter depends on the current state of the chain. At each jump, the chain randomly changes state according to a probability vector that also depends on the current state of the chain. The distribution of the chain can be characterized by its infinitesimal generator $A = [A_{v,v'}]_{v,v' \in \mathcal{V}}$. For any $v, v' \in \mathcal{V}$ such that $v' \neq v$, the coefficient $A_{v,v'}$ is non-negative and corresponds to the jump rate from state v to v' . The rows of the matrix sum to zero, so for all $v \in \mathcal{V}$, the diagonal coefficient

satisfies $A_{v,v} = -\sum_{v' \neq v} A_{v,v'}$. For any state $v \in \mathcal{V}$ and any function $f : \mathcal{V} \rightarrow \mathbb{R}^d$, we have:

$$\lim_{t \rightarrow 0^+} \frac{\mathbb{E}[f(Y_t) \mid Y_0 = v] - f(v)}{t} = \sum_{v' \in \mathcal{V}} A_{v,v'} f(v'). \quad (6.5)$$

Starting from a state $v \in \mathcal{V}$, after a time distributed according to an exponential law with parameter $-A_{v,v}$, the CTMC jumps to a random state in $\mathcal{V} \setminus \{v\}$ according to the probability vector $\left[\frac{-A_{v,v'}}{A_{v,v}} \right]_{v' \neq v}$. If $v \in \mathcal{V}$ is an absorbing state, then $A_{v,v'} = 0$ for any $v' \in \mathcal{V}$.

(ii) Let $(\bar{Y}_k)_{k \in \mathbb{N}^*}$ be a discrete-time and time-homogeneous Markov chain on \mathcal{V} . Its jump times are deterministic and occur at each integer time step. At each jump, the chain randomly changes state according to a probability vector that also depends on the current state of the chain. The distribution of the chain can be characterized by its transition matrix $\bar{A} = \left[\bar{A}_{v,v'} \right]_{v,v' \in \mathcal{V}}$. All coefficients are non-negative and each row sums to 1. Starting from a state $v \in \mathcal{V}$, the chain randomly jumps at the next time step to a state in \mathcal{V} according to the probability vector $\left[\bar{A}_{v,v'} \right]_{v' \in \mathcal{V}}$. Unlike a CTMC, a Markov chain can therefore jump in place. If $v \in \mathcal{V}$ is an absorbing state, then $\bar{A}_{v,v} = 1$ and $\bar{A}_{v,v'} = 0$ for all $v' \neq v$. Note that any DTMC that does not jump in place can be modeled by a CTMC $(Y_t)_{t \in \mathbb{R}_+}$ whose jump times are not known, but only the order of transitions. By denoting $(S_k)_k$ the successive jump times of the CTMC, the process $(Y_{S_k})_{k \in \mathbb{N}^*}$ is a discrete-time Markov chain with transition matrix \bar{A} defined for all $v, v' \in \mathcal{V}$ (assuming that v is not an absorbing state) by:

$$\bar{A}_{v,v'} = \frac{-A_{v,v'}}{A_{v,v}} \mathbf{1}_{v \neq v'}. \quad (6.6)$$

Continuous-time version

Let $(Y_t)_t$ be a time-homogeneous continuous-time Markov chain on the finite graph \mathcal{V} with the matrix $A \in \mathbb{R}^{d_v \times d_v}$ as infinitesimal generator. We do not specify the edges of the graph since they correspond to all the pairs v, v' such that $A_{v,v'} > 0$.

⁴All such functions can be characterized by a vector of size d_v , and belong to the domain of the generator.

DEFINITION 34 · Mean hitting time

For any subset $B \subset \mathcal{V}$ and any $v \in \mathcal{V}$, let us define $h_{v,B}$ the waiting time before reaching B starting from vertex v :

$$h_{v,B} = \inf\{t \in \mathbb{R}_+ : Y_t \in B \mid Y_0 = v\}. \quad (6.7)$$

The mean hitting time (MHT) of the set B from vertex v is thus given by:

$$H_{v,B} = \mathbb{E}_{(Y_t)_t \sim \mathcal{A}}[h_{v,B}]. \quad (6.8)$$

When $B = \{v'\}$ is a single vertex, we note $h_{v,v'} = h_{v,\{v'\}}$ and $H_{v,v'} = H_{v,\{v'\}}$. We can therefore also define the matrices $h = [h_{v,v'}]_{v,v' \in \mathcal{V}}$ and $H = [H_{v,v'}]_{v,v' \in \mathcal{V}}$.

The following well-known result establishes the link between mean hitting times and the process generator.

THEOREM 11 · Mean hitting times for a time-homogeneous CTMC

For any subset $B \subset \mathcal{V}$, the vector $H_{\cdot,B} = (H_{v,B})_{v \in \mathcal{V}}$ is the unique minimal non-negative solution to the linear system:

$$\begin{cases} H_{v,B} = 0 & \text{if } v \in B, \\ \sum_{v' \in \mathcal{V}} A_{v,v'} H_{v',B} = -1 & \text{if } v \notin B. \end{cases} \quad (6.9)$$

A solution vector $[H_{v,B}]_{v \in \mathcal{V}}$ of eq. (6.9) is minimal if for any other non-negative solution $[\tilde{H}_{v,B}]_{v \in \mathcal{V}}$, we have $H_{v,B} \leq \tilde{H}_{v,B}$ for any $v \in \mathcal{V}$.

Proof. For simplicity of notation in the proof, we will directly note $h_v = h_{v,B}$ and $H_v = H_{v,B}$ for any $v \in \mathcal{V}$. The case $H_v = 0$ if $v \in B$ is obvious. Suppose $v \notin B$. Either v is absorbing and then $H_v = +\infty$, or v is not absorbing and then the random walk $(Y_t)_t$ leaves v after a random time T_1 . In the latter case, the remaining time before reaching B is $h_{Y_{T_1}}$ (with $h_{Y_{T_1}} = 0$ if $Y_{T_1} \in B$). Thus, for any non-absorbing vertex $v \in \mathcal{V} \setminus B$, we have:

$$H_v = \mathbb{E}[h_v] = \mathbb{E}[T_1 \mid Y_0 = v] + \mathbb{E}[h_{Y_{T_1}} \mid Y_0 = v]. \quad (6.10)$$

Since $T_1 \sim \text{Exp}(-A_{v,v})$ we have $\mathbb{E}[T_1 \mid Y_0 = v] = -(A_{v,v})^{-1}$. On the other hand, thanks to the strong Markov property,

$$\begin{aligned} \mathbb{E}[h_{Y_{T_1}} \mid Y_0 = v] &= \mathbb{E}[\mathbb{E}[h_{Y_{T_1}} \mid Y_{T_1}] \mid Y_0 = v] = \mathbb{E}[H_{Y_{T_1}} \mid Y_0 = v], \\ &= \sum_{v' \neq v} H_{v'} \mathbb{P}(Y_{T_1} = v' \mid Y_0 = v) = -\frac{1}{A_{v,v}} \sum_{v' \neq v} A_{v,v'} H_{v'}. \end{aligned} \quad (6.11)$$

Finally, it comes:

$$H_v = -\frac{1}{A_{v,v}} \left(1 + \sum_{v' \neq v} A_{v,v'} H_{v'} \right), \quad (6.12)$$

and thus the desired result:

$$H_v A_{v,v} + \sum_{v' \neq v} A_{v,v'} H_{v'} = -1. \quad (6.13)$$

It remains to prove the minimality of the solution vector $[H_v]_{v \in \mathcal{V}}$. Let $[\tilde{H}_v]_{v \in \mathcal{V}}$ be another non-negative solution of the linear system eq. (6.9). Let v_0 be the initial state of the random walk. If $v_0 \notin B$, then:

$$\begin{aligned} \sum_{v_1 \in \mathcal{V}} A_{v_0, v_1} \tilde{H}_{v_1} = -1 &\iff \sum_{v_1 \neq v_0} A_{v_0, v_1} \tilde{H}_{v_1} + A_{v_0, v_0} \tilde{H}_{v_0} = -1, \\ &\iff \tilde{H}_{v_0} = \frac{-1}{A_{v_0, v_0}} + \sum_{v_1 \neq v_0} \frac{-A_{v_0, v_1}}{A_{v_0, v_0}} \tilde{H}_{v_1}. \end{aligned} \quad (6.14)$$

The same result holds for \tilde{H}_{v_1} for any $v_1 \notin B$, and thus we can write:

$$\begin{aligned} \tilde{H}_{v_0} &= \frac{-1}{A_{v_0, v_0}} + \sum_{\substack{v_1 \neq v_0 \\ v_1 \notin B}} \frac{-A_{v_0, v_1}}{A_{v_0, v_0}} \left(\frac{-1}{A_{v_1, v_1}} + \sum_{v_2 \neq v_1} \frac{-A_{v_1, v_2}}{A_{v_1, v_1}} \tilde{H}_{v_2} \right), \\ &= \frac{-1}{A_{v_0, v_0}} + \sum_{\substack{v_1 \neq v_0 \\ v_1 \notin B}} \frac{A_{v_0, v_1}}{A_{v_0, v_0}} \frac{1}{A_{v_1, v_1}} + \sum_{\substack{v_1 \neq v_0 \\ v_1 \notin B}} \sum_{v_2 \neq v_1} \frac{-A_{v_0, v_1}}{-A_{v_0, v_0}} \frac{A_{v_1, v_2}}{A_{v_1, v_1}} \tilde{H}_{v_2}. \end{aligned} \quad (6.15)$$

By recurrence, for any $m \in \mathbb{N}^*$ and any $v_0 \notin B$, we have:

$$\begin{aligned} \tilde{H}_{v_0} &= \frac{-1}{A_{v_0, v_0}} + \sum_{k=1}^m \sum_{\substack{v_1 \neq v_0 \\ v_1 \notin B}} \dots \sum_{\substack{v_k \neq v_{k-1} \\ v_k \notin B}} \left(\prod_{\ell=1}^k \frac{-A_{v_{\ell-1}, v_\ell}}{A_{v_{\ell-1}, v_{\ell-1}}} \right) \left(\frac{-1}{A_{v_k, v_k}} + \sum_{\substack{v_{k+1} \neq v_k \\ v_{k+1} \notin B}} \frac{-A_{v_k, v_{k+1}}}{A_{v_k, v_k}} \tilde{H}_{v_{k+1}} \right), \\ &\geq \sum_{k=1}^m \sum_{\substack{v_1 \neq v_0 \\ v_1 \notin B}} \dots \sum_{\substack{v_k \neq v_{k-1} \\ v_k \notin B}} \left(\prod_{\ell=1}^k \frac{-A_{v_{\ell-1}, v_\ell}}{A_{v_{\ell-1}, v_{\ell-1}}} \right) \times \frac{-1}{A_{v_k, v_k}}. \end{aligned} \quad (6.16)$$

The inequality comes from the non-negativity of $\frac{-A_{v,v'}}{A_{v,v}}$ and by assumption, of \tilde{H}_v , for any $v \neq v' \in \mathcal{V}$. The right term of this inequality can be reexpressed in terms of the jump times of the random walk $(Y_t)_t$. Let T_1, T_2, \dots be the successive waiting times before each jump and S_1, S_2, \dots the times of the jump ($S_0 = 0$

and $S_k = S_{k-1} + T_k$ for $k \in \mathbb{N}^*$. For any $k \in \mathbb{N}^*$, we have:

$$\begin{aligned}
 \mathbb{E}\left[T_{k+1} \mathbb{1}_{S_{k+1} \leq h_{v_0}} \mid Y_0 = v_0\right] &= \sum_{\substack{v_1 \neq v_0 \\ v_1 \notin B}} \frac{-A_{v_0, v_1}}{A_{v_0, v_0}} \times \mathbb{E}\left[T_k \mathbb{1}_{S_k \leq h_{v_1}} \mid Y_0 = v_1\right], \\
 &= \sum_{\substack{v_1 \neq v_0 \\ v_1 \notin B}} \sum_{\substack{v_2 \neq v_1 \\ v_2 \notin B}} \frac{-A_{v_0, v_1}}{A_{v_0, v_0}} \frac{-A_{v_1, v_2}}{A_{v_1, v_1}} \times \mathbb{E}\left[T_{k-1} \mathbb{1}_{S_{k-1} \leq h_{v_2}} \mid Y_0 = v_2\right], \\
 &= \sum_{\substack{v_1 \neq v_0 \\ v_1 \notin B}} \dots \sum_{\substack{v_k \neq v_{k-1} \\ v_k \notin B}} \left(\prod_{\ell=1}^k \frac{-A_{v_{\ell-1}, v_\ell}}{A_{v_\ell, v_\ell}} \right) \times \mathbb{E}\left[T_1 \mathbb{1}_{S_1 \leq h_{v_k}} \mid Y_0 = v_k\right], \\
 &= \sum_{\substack{v_1 \neq v_0 \\ v_1 \notin B}} \dots \sum_{\substack{v_k \neq v_{k-1} \\ v_k \notin B}} \left(\prod_{\ell=1}^k \frac{-A_{v_{\ell-1}, v_\ell}}{A_{v_\ell, v_\ell}} \right) \times \frac{-1}{A_{v_k, v_k}}.
 \end{aligned} \tag{6.17}$$

Using the monotone convergence theorem, it comes:

$$\widetilde{H}_{v_0} \geq \mathbb{E}\left[\sum_{k=1}^m T_k \mathbb{1}_{S_k < h_{v_0}} \mid Y_0 = v_0\right] \xrightarrow{m \rightarrow \infty} \mathbb{E}[h_{v_0}] = H_{v_0}. \tag{6.18}$$

■

Discrete time version

This result is better known for classical Markov chains. Let $(\bar{Y}_n)_n$ be a time-homogeneous discrete-time Markov chain with transition matrix \bar{A} on \mathcal{V} . Let $B \subset \mathcal{V}$, we note:

$$\bar{h}_{v,B} = \inf\{n \in \mathbb{N} : \bar{Y}_n \in B \mid \bar{Y}_0 = v\}, \tag{6.19}$$

and

$$\bar{H}_{v,B} := \mathbb{E}[\bar{h}_{v,B}]. \tag{6.20}$$

THEOREM 12 · Mean hitting times for a time-homogeneous Markov chain

The vector $\bar{H}_{\cdot,B} := (\bar{H}_{v,B})_{v \in \mathcal{V}}$ is the unique minimal non-negative solution to the linear system:

$$\begin{cases} \bar{H}_{v,B} = 0 & \text{if } v \in B, \\ \bar{H}_{v,B} = 1 + \sum_{v' \in \mathcal{V}} \bar{A}_{v,v'} \bar{H}_{v',B} & \text{if } v \notin B. \end{cases} \tag{6.21}$$

The proof follows the same pattern as the previous one. Note that if S_k is the elapsed time before the k -th jump for the continuous-time random walk $(Y_t)_t$ and if $\bar{A}_{v,v'} = \frac{-A_{v,v'}}{A_{v,v}}$ for any vertices $v, v' \in \mathcal{V}$, we have $\bar{H}_{v,B} = \mathbb{E}[\inf\{k : Y_{S_k} \in B\} \mid Y_0 = v]$.

Dynkin's formula approach

[Theorem 11](#) actually is a special case of a much more general result.

THEOREM 13 · Dynkin's formula

Let $(Y_t)_t$ be a strong Markov process with infinitesimal generator A on a finite space \mathcal{Y} . Let also φ be a function from \mathcal{Y} to \mathbb{R} and τ a stopping time such that $\mathbb{E}[\tau] < \infty$. We have :

$$\mathbb{E}[\varphi(Y_\tau) \mid Y_0] = \varphi(Y_0) + \mathbb{E}\left[\int_0^\tau A\varphi(Y_s) ds \mid Y_0\right]. \quad (6.22)$$

A proof for general strong Markov processes is given by Dynkin in [[Dyn65](#), Theorem 1.3.C p23].

Let B be a subset of \mathcal{Y} . We are looking for a function H from \mathcal{Y} to \mathbb{R} solution of the following system:

$$\begin{cases} AH = -1 & \text{on } \mathcal{Y} \setminus B, \\ H = 0 & \text{on } B. \end{cases} \quad (6.23)$$

We define here $\tau := \inf\{t \geq 0 : Y_t \in B\}$. By definition we have $\mathbb{E}[H(Y_\tau) \mid Y_0] = 0$ and $AH(Y_s) = -1$ for any $s < \tau$. By applying Dynkin's formula ([theorem 13](#)) we get:

$$H(Y_0) = -\mathbb{E}\left[\int_0^\tau -1 ds \mid Y_0\right] = \mathbb{E}[\tau \mid Y_0]. \quad (6.24)$$

Finally, it comes:

$$\begin{cases} A(y \mapsto \mathbb{E}[\tau \mid Y_0 = y]) = -1 & \text{on } \mathcal{Y} \setminus B, \\ (y \mapsto \mathbb{E}[\tau \mid Y_0 = y]) = 0 & \text{on } B. \end{cases} \quad (6.25)$$

6.2.2 Mean hitting times as proximity scores

Let $\tau := \inf\{t \in \mathbb{R}_+ : V_t \in \mathcal{V}_F\}$. The evolution of the PDMP regime over time $(V_t)_t$ is, as mentioned earlier, a non-Markovian random walk on \mathcal{V} . Attempting to solve the linear system [eq. \(6.9\)](#) does not make sense in this case. Firstly, we do not have a well-defined generator. Secondly, the vector $(\mathbb{E}[\tau \mid V_0 = v])_{v \in \mathcal{V}}$ is ambiguous since the distribution of τ also depends on the starting position Z_0 and not only on the initial regime V_0 .

True mean hitting times for the PDMP

The reasoning used previously using Dynkin's formula ([theorem 13](#)) could be applied to the PDMP $(X_t)_t$ with infinitesimal generator \mathcal{A} . Let $x = (z, v) \in \mathcal{X}$, we have :

$$\begin{cases} \mathcal{A}(\mathbb{E}[\tau | X_0 = x]) &= -1 & \text{if } v \notin \mathcal{V}_{\mathcal{F}}, \\ \mathbb{E}[\tau | X_0 = x] &= 0 & \text{if } v \in \mathcal{V}_{\mathcal{F}}. \end{cases} \quad (6.26)$$

We would then like to determine the function $H : \mathcal{X} \mapsto \mathbb{R}_+$ equal to 0 on \mathcal{F} and solution on $\mathcal{X} \setminus \mathcal{F}$ of :

$$\left\langle \frac{d}{dt} \psi_v(z, t); \nabla_z H((z, v)) \right\rangle + \lambda(x) \left[\int_{\mathcal{X}} H(x') q(x' | x) \nu_x(dx') - H(x) \right] = -1. \quad (6.27)$$

Solving this problem explicitly is of course out of reach. Rather than obtaining an approximate solution for the exact process, we prefer to determine the exact solution for an approximated process.

Simplified random walk

We consider a Markovian and time-homogeneous random walk $(Y_t)_t$ on \mathcal{V} . The random walk can be chosen to be discrete-time, but we prefer the flexibility offered by continuous time⁵. We denote by $\mathcal{A} \in \mathbb{R}^{d_v \times d_v}$ its infinitesimal generator. It is then a PDMP whose flow is zero (thus no jumps at the boundaries) and whose jump intensity and jump kernel depend only on the regime. The intensity in regime v is given by $-\mathcal{A}_{v,v}$ and the jump kernel between a vertex v and v' is $-\mathcal{A}_{v,v'} / \mathcal{A}_{v,v}$. We call $(Y_t)_t$ the Simplified Random Walk (SRW).

Ideally, we would like to choose a SRW whose behavior is close to that of $(V_t)_t$. In practice, a simple (uniform) random walk with the right support already offers excellent performance on our test case. By "right support", we mean a support that respects that of $(V_t)_t$ in the following sense : for all $v, v' \in \mathcal{V}$, $\mathcal{A}_{v,v'}$ is non-zero if and only if there exist $z, z' \in \mathcal{Z}$ such that $\lambda((z, v))$ and $q((z', v') | (z, v))$ are non-zero. It is possible to do better, but the quality of a time-homogeneous random walk depends heavily on the problem at hand.

Proximity scores

For any $v \in \mathcal{V}$, let $H_{v, \mathcal{V}_{\mathcal{F}}}$ be the mean hitting time of the critical region $\mathcal{V}_{\mathcal{F}}$ starting from the regime v for the SRW $(Y_t)_t$. We can apply [theorem 11](#) to explicitly determine the vector $(H_{v, \mathcal{V}_{\mathcal{F}}})_{v \in \mathcal{V}}$ by solving a linear system. We define a new proximity score based on mean hitting times.

⁵The sojourn time in a given regime can be more impactful than the number of jumps required to reach the target region.

DEFINITION 35 · MHT-based proximity score

For any time-homogeneous Markovian random walk $(Y_t)_t$ on \mathcal{V} , let $(H_{v, \mathcal{V}_F})_{v \in \mathcal{V}}$ be the vector of mean hitting times of \mathcal{V}_F for $(Y_t)_t$:

$$H_{v, \mathcal{V}_F} = \mathbb{E}_{(Y_t), \sim \mathcal{A}} [\inf\{t \in \mathbb{R}_+ : Y_t \in \mathcal{V}_F\} \mid Y_0 = v]. \quad (6.28)$$

The proximity score associated with the simplified random walk $(Y_t)_t$ on \mathcal{V} is given by:

$$\beta^{\text{MHT}} : v \in \mathcal{V} \mapsto 1 - \frac{H_{v, \mathcal{V}_F}}{\max_{v' \in \mathcal{V}} H_{v', \mathcal{V}_F}}. \quad (6.29)$$

We indeed have $\beta^{\text{MHT}}(v) \in [0, 1]$ for any $v \in \mathcal{V}$ and more specifically:

- (i) $\beta^{\text{MHT}}(v) = 0$ if and only if v belongs to the set of vertices from which the SRW $(Y_t)_t$ takes the longest time to reach \mathcal{V}_F (i.e., the regimes furthest from \mathcal{V}_F in the sense we have defined).
- (ii) $\beta^{\text{MHT}}(v) = 1$ if and only if $v \in \mathcal{V}_F$.

The degree of expressiveness of this score is directly given by that of the mean hitting times $(H_{v, \mathcal{V}_F})_{v \in \mathcal{V}}$. It can potentially assign a distinct value to each vertex, even on large graphs. In this respect, it is much more refined than the score β^{MPS} for industrial systems⁶. Suppose that the failure of an additional component brings us from a vertex v to a vertex v' , but this component did not belong to any intact MPS. We would then have $\beta^{\text{MPS}}(v) = \beta^{\text{MPS}}(v')$. However, the neighbors of v' correspond to regimes with a number of damaged MPS greater than or equal to that of the neighbors of v . This is because this additional failure likely increases the number of broken components in at least one MPS, thereby lengthening its total repair time. We would then like to consider v' as slightly closer to \mathcal{V}_F than v . It is an other strength of the proximity score β^{MHT} . It does not require us to think about how much importance to give to this type of phenomenon. If the difference in situation is perceptible for the SRW, then it will be taken into account.

6.3 Numerical experiments

We now turn to the implementation of this method. First, we will focus on the computation of these mean hitting times in practice. Then, we will present the application of our adaptive importance sampling method with the proximity score β^{MHT} .

⁶It is also more refined than what would be obtained by a generalization of β^{MPS} on arbitrary graphs, as suggested at the end of [section 6.1.2](#).

6.3.1 Computation of mean hitting times

The explicit derivation of the mean hitting times has a numerical cost depending on the size of the graph. The development of efficient algorithms for computing hitting/commute times in large graphs has been a very active field for about thirty years (in fact, since the advent of the internet).

Hitting and commute times computation for large graphs

The current literature is mainly interested in the case of very large graphs (with millions or even billions of vertices). It is then rarely possible to estimate the entire matrix H (i.e., the MHT between all pairs of vertices). The goal is to quickly estimate, upon request, the MHT (and more frequently the commute time or effective resistance) between two given vertices v and v' [Pen+21] (or between a vertex v and a given group $B \subset \mathcal{V}$ [Guo+24]). An exploratory Monte Carlo approach is then favored, exploiting the centrality of particular vertices [Lia+23] or combining exploration with a clever rewriting of the linear system [YT23; LL20]. Monte Carlo approaches are relevant for individual queries but become very expensive when the MHT must be estimated for several possible starting vertices. The rewriting of the linear system, on the other hand, often relies on the assumption of a uniform random walk which comes with exploitable properties (for example, its invariant measure is known explicitly).

In our case, we want to be able to efficiently estimate $H_{v,v'}$ for any starting vertex v (or rather H_{v,\mathcal{V}_F} for any v). This remains much less expensive than estimating $H_{v,v'}$ for any destination vertex v' . The second problem involves solving a system eq. (6.9) (or eq. (6.21)) for each destination vertex v' . Several options exist in this case, such as truncating the MHTs [SM12], using distributed algorithms [Can+19], or cleverly determining an approximation of the pseudo-inverse of the graph Laplacian [Bol21; AG17].

The relevance of MHTs has recently been questioned by the article [VRH10]. It notes that the MHTs of a uniform random walk become very uninformative when the size of the graph explodes because they then depend only on the degree of the destination vertex. This is not a problem in our case. On the one hand, the size of the graphs we consider is probably still far from this asymptotic situation, on the other hand, we can often propose better than a uniform random walk. Moreover, the result is only valid for certain families of random graphs.

Finally, let us mention the article [KTD19] which is interested in the calculation of MHTs not for a given random walk but for a family of possible walks with the formalism of imprecise probabilities.

Sparse methods for linear systems

The vector of mean hitting times of \mathcal{V}_F is solution of a known linear system given in eq. (6.9) if the simplified random walk is time-continuous, or in eq. (6.21) if the SRW is time-discrete.

Since $H_{v,\mathcal{V}_F} = 0$ (resp. $\bar{H}_{v,\mathcal{V}_F} = 0$) in the case $v \in \mathcal{V}_F$, the size of the system to be solved is actually given by $d_{\mathcal{V} \setminus \mathcal{V}_F}$, the cardinal of the set $\{\mathcal{V} \setminus \mathcal{V}_F\}$. We note $\mathbf{1}$ the vector of length $d_{\mathcal{V} \setminus \mathcal{V}_F}$ whose coordinates are all 1 and Id the identity matrix of size $d_{\mathcal{V} \setminus \mathcal{V}_F} \times d_{\mathcal{V} \setminus \mathcal{V}_F}$. One just needs to solve:

$$A' H'_{,\mathcal{V}_F} = -\mathbf{1}, \quad (6.30)$$

in the time-continuous case with $A' = (A_{v,v'})_{v,v' \in \mathcal{V} \setminus \mathcal{V}_F}$ and $H'_{,\mathcal{V}_F} = (H_{v,\mathcal{V}_F})_{v \in \mathcal{V} \setminus \mathcal{V}_F}$, and:

$$(\bar{A}' - \text{Id}) \bar{H}'_{,\mathcal{V}_F} = -\mathbf{1}, \quad (6.31)$$

in the time-discrete case with $\bar{A}' = (\bar{A}_{v,v'})_{v,v' \in \mathcal{V} \setminus \mathcal{V}_F}$ and $\bar{H}'_{,\mathcal{V}_F} = (\bar{h}_v)_{v \in \mathcal{V} \setminus \mathcal{V}_F}$. A very naive approach to solving this system numerically involves inverting A' (resp. $\bar{A}' - \text{Id}$). The cost of this resolution is in $O(d_{\mathcal{V} \setminus \mathcal{V}_F}^3)$, which remains a problem for large graphs.

The graph is large if there is a large number of possible regimes for the modeled process. However, the number of possible transitions from a given regime then generally represents only a small fraction of the existing regimes. In other words, it is often relevant to choose an SRW whose generator is sparse. If most of the coefficients of A' (or \bar{A}') are zero, then there are many sparse methods for storing the matrix, performing matrix/vector products, and especially solving linear systems involving A' that are memory-efficient and computationally efficient. These methods were primarily developed for solving PDEs but are now common. Interested readers are referred to this reference book [Saa03] and a history of these methods by the same author [Saa20].

We recommend the use of a GMRES method [SS86; Zou23]. It does not require strong assumptions about A' (such as symmetry), and has given good results in practice. It is now a well established method and therefore efficiently implemented in many numerical toolboxes. Moreover, any subtleties in its implementation (restart, preconditioning, etc.) are extensively covered in the literature [GLJ19].

6.3.2 Application on the spent fuel pool

The adaptive importance sampling (AIS) method presented in [chapter 4](#) is now tested with the proximity score β^{MHT} . It is once again applied to the test case of the spent fuel pool (SFP [example 10](#)), under the same conditions as in [section 5.3.2](#). We compare its performance to a classical Monte Carlo method (CMC) and to our AIS method with the proximity score β^{MPS} .

Size of the linear system

The set of regimes \mathcal{V} for the SFP system is given by the status of its 15 components (active, inactive, failed). Therefore, there are $3^{15} = 14\,348\,907$ possible regimes. A naively-encoded

CTMC on this set would have a generator matrix with approximately 2×10^{14} coefficients. Since the membership of a regime v to $\mathcal{V}_{\mathcal{F}}$ only depends on which components are broken, we can restrict ourselves to a subset of size $2^{15} = 32\,768$. Within this subset, 27 854 regimes belong to $\mathcal{V}_{\mathcal{F}}$. The linear system to be solved to determine the mean hitting times is therefore of size 4914. Such a reduction can be explained by combinatorial reasons. Among the 32 768 possible configurations, most correspond to regimes with a large number of broken components and therefore often a cut set. For example, 85% of the 32 768 configurations have at least 6 broken components out of 15 (i.e. more than 40% of the SFP system).

Then comes the sparsity of the system. The possible transitions for the SRW correspond to the additional failure or repair of a component. Therefore, there are $d_c = 15$ possible transitions at each moment and thus at most 15 non-zero values per row in the final matrix. All of this combined allows us to store a matrix in sparse format with only 53 030 non-zero coefficients.

Quasi-uniform random walk

We propose a simplified random walk $(Y_t)_t$, simple enough to illustrate the effectiveness of the method without complex hyperparameter tuning. The SRW is time-continuous. Two regimes are neighbors on the graph if and only if one of them has precisely one more broken component than the other. This preserves the support of the true random walk $(V_t)_t$. We assign all components a constant failure rate $\bar{\lambda}$ and a constant repair rate $\bar{\mu}$. The SRW $(Y_t)_t$ therefore has an infinitesimal generator given by the matrix \mathcal{A} defined for any $v, v' \in \mathcal{V}$ by:

$$A_{v,v'} = \bar{\lambda} \mathbb{1}_{b(v')=b(v)+1} + \bar{\mu} \mathbb{1}_{b(v')=b(v)-1} - (\bar{\lambda}(d_c - b(v)) + \bar{\mu}b(v)) \mathbb{1}_{v=v'}, \quad (6.32)$$

with $b(v)$ the number of broken components in the regime v . We have chosen the failure rate $\bar{\lambda} = 1$ and the repair rate $\bar{\mu} = 20$. This walk remains very simple because it is uniform among failures and uniform among repairs. The essential aspect is that failures occur less and less frequently as the number of broken components increases.

The input information is deliberately limited. On the one hand, this allows us, as mentioned earlier, to observe the method's performance in a suboptimal situation. On the other hand, since we are not looking at mean hitting times as output, it is not certain that the SRW $(Y_t)_t$ that is "the closest" to $(V_t)_t$ is the most effective for estimating \mathcal{I} . Moreover, a generator \mathcal{A} with very heterogeneous values is more likely to be ill-conditioned and to complicate the numerical resolution of the system eq. (6.9).

Numerical results

We carry out two series of numerical simulations on the spent fuel pool system.

- (i) We first compare the performance of the AIS method based on the proximity score β^{MHT}

TABLE 6.1 **Results on the standard SFP case** (with jump rates from table 5.5).

Method	N_L	\hat{I}_{N_L}	C.o.v $\frac{\hat{\sigma}_{N_L}}{\hat{I}_{N_L}}$	95% confidence interval
CMC	10^5	2×10^{-5}	223.60	$[0; 4.77 \times 10^{-5}]$
	10^6	1.3×10^{-5}	277.35	$[5.93 \times 10^{-6}; 2.01 \times 10^{-5}]$
	10^7	1.77×10^{-5}	237.68	$[1.51 \times 10^{-5}; 2.03 \times 10^{-5}]$
AIS with ξ_θ^{MPS}	10^3	2.19×10^{-5}	3.01	$[1.78 \times 10^{-5}; 2.60 \times 10^{-5}]$
	10^4	1.99×10^{-5}	1.01	$[1.96 \times 10^{-5}; 2.03 \times 10^{-5}]$
AIS with ξ_θ^{MHT}	10^3	1.86×10^{-5}	1.62	$[1.67 \times 10^{-5}; 2.04 \times 10^{-5}]$
	10^4	2.01×10^{-5}	0.88	$[1.98 \times 10^{-5}; 2.05 \times 10^{-5}]$

to the results obtained with the CMC method and the AIS method based on β^{MPS} in chapter 5 for the standard case with jump rates described in table 5.5. The results are described in table 6.1 and a probability of system failure about 10^{-5} . The sample size of the AIS method is ranging from 10^3 to 10^4 . The sample size of the CMC method is ranging from 10^5 to 10^7 .

- (ii) We then check the stability of the AIS method based on ξ_θ^{MHT} . We represent in fig. 6.2 15 confidence intervals at 95% level obtained with the AIS-MHT method with a sample size of 10^3 trajectories still on the standard case (table 5.5) and we compare them to the confidence interval obtained with the CMC method and a sample size of 10^7 .

The AIS method is initialized with the procedure proposed in section 4.3.3. We choose the smallest θ (in norm) such that the probability that at least one component failure occurs before the end of the simulation is larger than $1/3$. We set $d_\theta = 8$. At each iteration, we generate trajectories until we have $n_{\text{CE}} = 10$ failures before updating θ for $N_L = 10^3$ and $n_{\text{CE}} = 50$ for $N_L = 10^4$. We stop when the total budget N_L is reached. The estimated probability \hat{I}_{N_L} is given by eq. (4.30). The estimated coefficient of variation (C.o.v) is the ratio between the estimated asymptotic standard deviation $\hat{\sigma}_{N_L}$ (square root of the variance estimator given in eq. (4.50)) and the estimated probability \hat{I}_{N_L} . The confidence interval formula is given by theorem 9. For the CMC method, we simply generate N_L trajectories and we count the proportion of faulty trajectories.

We observe in table 6.1 that the method's performance with the IF ξ_θ^{MHT} is very similar, if not better, than that obtained with the IF ξ_θ^{MPS} . Indeed, for the same sample size, the coefficient of variation (C.o.v) and the length of the confidence intervals are slightly smaller for ξ_θ^{MHT} . The analysis of the results is therefore similar to that proposed in section 5.3.2. The variance reduction compared to the CMC method is of the order of 10^4 . We also observe in fig. 6.2 that

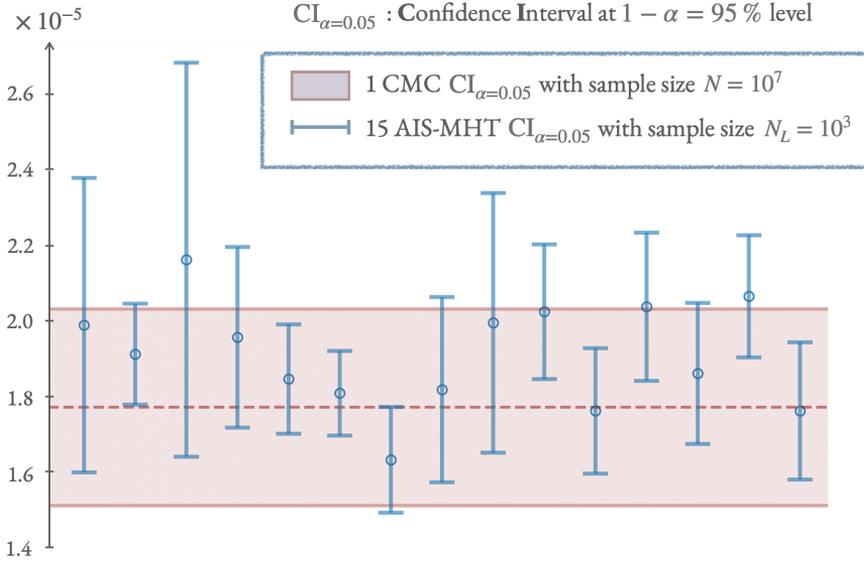


FIGURE 6.2 **Comparison of 15 confidence intervals at 95% level obtained with AIS-MHT method approximation.** Each confidence interval corresponds to a run of the AIS-MHT method on the standard case of the SFP (table 5.5) with 10^3 trajectories (same conditions as for table 5.7). They are compared to the confidence interval obtained with the CMC method on 10^7 trajectories.

the method is just as empirically robust with $\xi_{\theta}^{\text{MHT}}$ as with $\xi_{\theta}^{\text{MPS}}$ (see fig. 5.6). The confidence intervals are indeed all of similar relative length for a sample size of 10^3 and comparable to that of the CMC method for a sample size of 10^7 .

We conclude that the IF $\xi_{\theta}^{\text{MHT}}$, which is much more general than $\xi_{\theta}^{\text{MPS}}$, performs at least as well in the specific case of coherent industrial systems for which $\xi_{\theta}^{\text{MPS}}$ is designed. However, the calculation of the mean hitting times before executing the AIS method seems to us more expensive for very large industrial systems than the determination of the MPSs (and rarely benefits from directly dedicated methods mastered by reliability engineers).

CONCLUSION

CHAPTER 7

Conclusion and perspectives

Disclaimer

In accordance with the regulations of the doctoral school, the introduction was written in French. This conclusion is written in English but a French translation is proposed in appendix (right after the English translation of the introduction).

Content of this chapter

7.1	Conclusion	154
7.1.1	Summary of contributions	154
7.1.2	Discussion	157
7.2	Perspectives	158
7.2.1	Refinements to the AIS Method	158
7.2.2	Bandits: Best Arm Identification	161

7.1 Conclusion

The topic of this thesis was motivated by industrial reliability issues faced by the company EDF (Électricité de France). We specifically focused on estimating the failure probability of industrial systems involved in the operation of nuclear power plants and dams.

Problem reminder

The industrial systems we are concerned with can be represented as piecewise deterministic Markov processes (PDMPs). Mathematically, this estimation problem is expressed as:

$$\mathcal{I} = \mathbb{E}_{\mathbf{X} \sim \mathbf{p}}[\mathbf{I}(\mathbf{X})], \quad (7.1)$$

with \mathbf{X} a PDMP trajectory, \mathbf{p} its reference distribution, and \mathbf{I} is the indicator function for the event of interest: the trajectory leads to failure. Since PDMP trajectories are costly to simulate and the failure probability we are looking for is typically very small, we cannot rely on a Classical Monte Carlo method.

It is known that the optimal distribution for importance sampling of PDMPs is fully characterized by the committor function of the process. This committor function associates to each moment t of a PDMP trajectory the probability that the trajectory will lead to the event of interest, given its state at time t . We hypothesized that it is possible to approximate the committor function at a reasonable cost, and that an importance distribution constructed from a good approximation of the committor function would produce a low-variance estimator.

7.1.1 Summary of contributions

All the contributions of this thesis are related to the problem we just described. They can be grouped into a single global estimation methodology based on the approximation of the committor function. This methodology consists of two phases. In the first “offline” phase, i.e., before any simulation, a family of approximations of the committor function is constructed. Each approximation corresponds to a possible importance distribution. In the second “online” phase, the best importance distribution is sequentially determined within the family defined in the previous phase. The simulations generated during this second phase allow both refining the approximation of the committor function and contributing to the final estimation of the quantity of interest.

Approximation a priori de la fonction committor

We first introduced in [chapter 4](#) the general form of a parametric family of approximations for the committor function:

$$\xi_{\theta}(x) = \exp \left[\sum_{j=1}^{d_{\Theta}} \theta_j \times \phi_j(\beta(v)) \right]. \quad (7.2)$$

with,

- (i) a parameter vector $\theta \in \Theta \subset \mathbb{R}^{d_{\Theta}}$ with dimension $d_{\Theta} > 0$,
- (ii) a proximity score $\beta : \mathcal{V} \mapsto [0, 1]$, where \mathcal{V} is the set of regimes of the PDMP,
- (iii) a family of basis functions $(\phi_j)_{j=1}^{d_{\Theta}}$ with $\phi_j : [0, 1] \mapsto \mathbb{R}$ for any $j \in \llbracket 1, d_{\Theta} \rrbracket$.

An implicit contribution of this thesis is to clearly distinguish:

- ♦ the prior information provided to the model, here represented by the proximity score β ,
- ♦ from the information learned during the simulations, represented by the successive values of the parameter θ , used to guide the importance distribution.

We proposed two families of proximity scores that can be determined a priori, without simulations, and at a reasonable cost using suitable algorithms. In [chapter 5](#), we constructed a proximity score valid for coherent multi-component industrial systems, based on the proportion of minimal paths damaged in a given regime. These minimal paths exploit the system's structure function and correspond to the barriers the process must necessarily cross to reach failure. In [chapter 6](#), we introduced a more general score, not making any assumptions about what the PDMP models, and based on the mean hitting times of a random walk on a graph represented by the PDMP regimes. This score can be explicitly computed for a homogeneous Markov random walk on this graph. It is therefore conditioned on the choice of the homogeneous walk approximating the actual walk performed by the PDMP, but a very simple walk, such as a uniform random walk, proved sufficient in practice.

Adaptive strategy

We proposed an adaptive preferential sampling methodology (AIS) based on a cross-entropy (CE) procedure. The drawback of classical sequential methods is that they use only part of the simulations generated for optimization steps and the final estimation step. The advantage of importance sampling lies in its ability to recycle all past samples at any time by adjusting the likelihood ratio. This significantly increases the precision of the method for the same computational budget.

Convergence results for simple recycling schemes exist in the literature, but their assumptions are often difficult to interpret and verify in practice. For PDMPs, we proposed interpretable and minimally restrictive assumptions for practitioners. Our theorem guarantees convergence and asymptotic normality, enabling the construction of asymptotic confidence intervals for the quantity of interest.

Dissemination of the work

The contributions of this thesis have resulted in two research articles, with the three contribution chapters reorganized accordingly.

- ◊ The first article, [Che+24a], corresponds to an earlier version of the methodology described in chapter 4, and introduces proximity scores tailored to the modeling of industrial systems, detailed in chapter 5. This work was published in 2024 in the *SIAM Journal of Uncertainty Quantification*.

Guillaume Chennetier, Hassane Chraïbi, Anne Dutfoy, and Josselin Garnier. “Adaptive importance sampling based on fault tree analysis for piecewise deterministic Markov process”. In: *SIAM/ASA Journal on Uncertainty Quantification* 12.1 (2024), pp. 128–156

- ◊ The second article corresponds to enhancements proposed in the methodology of chapter 4 relative to the first article [Che+24a], and introduces new proximity scores, described in chapter 6, based on hitting times for a random walk on a graph. This article is planned for submission in 2024.

Guillaume Chennetier, Hassane Chraïbi, Anne Dutfoy, and Josselin Garnier. “Graph-informed importance sampling for rare event estimation with piecewise deterministic Markov process”. In: (Submission in 2024)

The results obtained throughout the thesis were also presented at various national and international conferences. Selected examples:

- ◊ UNCECOMP (Conference on Uncertainty Quantification in Computational Science and Engineering), 2023 edition,
- ◊ MCM (Monte Carlo Methods and applications), 2023 edition,
- ◊ SIAM UQ (Conference on Uncertainty Quantification), 2024 edition,
- ◊ JDS (Journées de Statistique), 2022, 2023, 2024 editions.

Two awards were also granted for my work at the MASCOT-NUM meetings, 2022 and 2024 editions, organized by the thematic group RT-UQ:

- ◇ MASCOT-NUM 2022 edition: prize of best PhD Student poster.
- ◇ MASCOT-NUM 2024 edition: prize of best PhD Student talk.

7.1.2 Discussion

It is very clear that the nature of the integral we aim to estimate lends itself exclusively to probabilistic simulation. The highly specific geometry of the PDMP trajectory space makes it challenging to explore through other means. However, importance sampling is a risky approach for multimodal distributions in high dimensions. This thesis has, at least partially, justified the use of this method for this problem.

Relevance of the approach

One criticism of importance sampling is its intrusive nature. This is not an issue here since the practitioner fully controls all the simulation parameters of the model. We chose to build upon the methodology introduced in Thomas Galtier's thesis [Gal19]. It was not immediately obvious:

- ◇ that expressing the family of importance distributions solely in terms of an approximation of the committor function would offer the necessary flexibility,
- ◇ or that the minimizer of the Kullback-Leibler divergence, within a fairly constrained family, would yield a low-variance estimator.

This gamble proved effective in practice. The current literature on importance sampling heavily relies on modern learning methods to infer the optimal importance distribution during simulations in high-dimensional and multimodal contexts. These methods are well-suited for learning very complex functions with a high simulation budget. In our case, the target distribution is not particularly complex, as it is simply the product of a known distribution and an indicator function. However, the budget is limited to fewer than 10^4 simulations, which precludes optimizing the importance distribution iteratively in a non-parametric framework (or one with a very large number of parameters).

The key idea is as follows: a lightly parameterized family of distributions is sufficiently flexible, provided it is given the right input information. We demonstrated that, in practice, very few assumptions about the PDMP are necessary to provide this prior information.

Limitations

Not all questions are resolved at the end of this thesis. Firstly, the weights in an importance sampling method can always explode, and it is very challenging to assess the quality of an importance sampling estimate without comparing it to a brute-force Monte Carlo method.

Evaluation metrics based on variance or effective sample size are themselves poorly estimated when the method fails.

Our method is designed for cases where the cost of simulating a PDMP trajectory represents the sole significant expense of the procedure (due to the resolution of differential equations involved in the flow). It remains to be determined how high this cost needs to be (relative to the cost of constructing the family of committor function approximations and optimizing across iterations) for the method to remain competitive.

Conversely, if the physics play a major role in the system’s dynamics, simulations will be expensive, but approximating the committor function using only the regime and not the position becomes less defensible. However, incorporating the position would make density evaluation and all optimization phases much more costly, as the flow would need to be recalculated multiple times for the same trajectory.

7.2 Perspectives

Numerous opportunities naturally arise for extending this thesis, either by refining the proposed methods or broadening the range of applications. We can list a few examples:

- ◊ Applying our method to new test cases outside of industrial reliability. Since the proximity score β^{MHT} makes no assumptions about the nature of what the PDMP models, new applications—such as population dynamics or transportation network modeling—are entirely possible.
- ◊ Comparing the performance of our importance sampling method to splitting methods. Additionally, it would be worth investigating whether the importance functions ξ_θ we proposed are also effective for these methods.
- ◊ Developing the theoretical framework for importance sampling to “General stochastic hybrid systems” (GSHS) [BL06]. These stochastic processes generalize PDMPs: the process’s evolution between jumps is governed by stochastic differential equations instead of ordinary ones. A similar problem to ours for GSHS was addressed using an interacting particle system in [MB23].

Two additional key directions are detailed below.

7.2.1 Refinements to the AIS Method

The adaptive importance sampling (AIS) method proposed in [chapter 4](#) is relatively simple. It combines a cross-entropy minimization procedure with a sample-recycling scheme, both in their most basic forms. This simplicity allowed us to guarantee the convergence and asymptotic

normality of our algorithm under practical and verifiable conditions. Many refinements to these methods exist in the literature but complicate the convergence proofs.

One major difficulty with the cross-entropy procedure is determining a good initial importance distribution:

- ◊ If the initial distribution is too close to the reference distribution \mathbf{p} , a large number of trajectories must be generated for some to realize the event of interest. The first iteration of the algorithm is thus costly.
- ◊ If the initial distribution is too far from \mathbf{p} , it becomes difficult to control the likelihood ratio explosion risk.
- ◊ If the initial distribution covers only part of the target distribution's support (due to its strong multimodality), some regions of the trajectory space may never be observed after several optimization steps.

To mitigate the impact of the initial distribution \mathbf{p} , the target distribution \mathbf{g}^* can be varied. The idea is to define a family $(\mathbf{g}_\gamma^*)_{\gamma \in \mathbb{R}_+}$ of intermediate target distributions, parameterized by γ , interpolating between the reference distribution \mathbf{p} and the target distribution \mathbf{g}^* . By convention, $\lim_{\gamma \rightarrow 0^+} \mathbf{g}_\gamma^* = \mathbf{g}^*$.

Multilevel Cross entropy

Many rare event problems can be expressed as threshold exceedances:

$$\mathbf{I} : \mathbf{X} \mapsto \mathbb{1}_{\beta(\mathbf{X}) > s^*}, \quad (7.3)$$

where β is a performance score associated with the realization \mathbf{X} . In reliability analysis, events are often of the form $\{\beta \leq 0\}$, and β is referred to as the “limit state function”. We discussed in [section 1.3.2](#) two variants of the Cross entropy method: “multilevel Cross entropy” (mCE) and “improved Cross entropy” (iCE). The intermediate target distributions take the form $\mathbf{g}_\gamma^* \propto |\mathbf{I}_\gamma| \times \mathbf{p}$ where \mathbf{I}_γ replaces \mathbf{I} .

- ◊ The mCE method uses lower threshold indicators:

$$\mathbf{I}_\gamma^{\text{mCE}} : \mathbf{x} \mapsto \mathbb{1}_{\beta(\mathbf{x}) > s^* - \gamma}, \quad (7.4)$$

- ◊ While the iCE method uses the cumulative distribution function Φ of a $\mathcal{N}(0, 1)$ distribution as an approximation of the indicator function:

$$\mathbf{I}_\gamma^{\text{iCE}} : \mathbf{x} \mapsto \Phi\left(\frac{\beta(\mathbf{x}) - s^*}{\gamma}\right). \quad (7.5)$$

CONCLUSION

The update strategy for γ differs between methods. At iteration ℓ (without recycling scheme), a sample $\mathbf{X}_1, \dots, \mathbf{X}_{n_\ell} \sim \mathbf{g}_{\theta^{(\ell)}}$ is produced. In the mCE method, $\gamma_{\ell+1}^{\text{mCE}}$ is chosen to ensure a proportion ρ of trajectories exceed the corresponding threshold.

$$\gamma_{\ell+1}^{\text{mCE}} \in \inf \left\{ \gamma \in \mathbb{R}_+ : \sum_{k=1}^n \mathbf{I}_\gamma^{\text{mCE}}(\mathbf{X}_k) > \rho n \right\}. \quad (7.6)$$

In the iCE method, $\gamma_{\ell+1}^{\text{iCE}}$ is selected such that the coefficient of variation of $\left[\mathbf{I}_\gamma(\mathbf{X}_k) \frac{p(\mathbf{X}_k)}{\mathbf{g}_{\theta^{(\ell)}}(\mathbf{X}_k)} \right]_{k=1}^{n_\ell}$ is the closest possible to a fixed target value $\text{CV}_{\text{target}}$.

$$\gamma_{\ell+1}^{\text{iCE}} \in \arg \min_{\gamma \in [0, \gamma_\ell^{\text{iCE}}]} \left(\widehat{\text{CV}}_\gamma - \text{CV}_{\text{target}} \right)^2. \quad (7.7)$$

Trajectory Score

Let us return to our case: $\mathbf{X} = (X_t)_{t \in [0, \tau]}$ is a PDMP trajectory of duration τ . The process takes values in \mathcal{X} and the trajectory lies in \mathbf{T}_τ . We are interested in the probability that the trajectory \mathbf{X} reaches a region $\mathcal{F} \subset \mathcal{X}$ before a given time s_{max} . We define:

$$\tau = \inf \{ t \geq 0 : X_t \in \mathcal{F} \} \wedge s_{\text{max}}. \quad (7.8)$$

We denote $\mathbf{F} = \{ \mathbf{x} \in \mathbf{T}_\tau : \tau < s_{\text{max}} \}$. Our goal is to estimate $\mathcal{I} = \mathbb{E}_{\mathbf{X} \sim \mathbf{p}}[\mathbf{I}(\mathbf{X})]$ with:

$$\mathbf{I} : \mathbf{x} \in \mathbf{T}_\tau \mapsto \mathbf{1}_{\mathbf{x} \in \mathbf{F}}. \quad (7.9)$$

To apply the mCE or iCE procedure, a relevant performance score β for a trajectory must be chosen. One can draw inspiration from a classic choice used in splitting methods applied to stochastic processes ([Lou+17; CGR19]). If a scoring function β is already defined on the state space \mathcal{X} , the total score of a trajectory in \mathbf{T} can be selected as the maximum value of β along the trajectory:

$$\beta(\mathbf{X}) = \sup_{t \in [0, \tau]} \beta(X_t). \quad (7.10)$$

Two natural choices arise: the proximity score β already defined in this manuscript and the importance function ξ_θ at the current iteration.

Recycling and Confidence Intervals

Applying the mCE/iCE method to our case is not straightforward. First, the event $\{\mathbf{X} \in \mathbf{F}\}$ cannot be fully expressed as a threshold exceedance for the score function β . However, if $\mathbf{X} \in \mathbf{F}$, then $\sup_{t \in [0, \tau]} \beta(X_t) = 1$. Thus, it is necessary to transition from \mathbf{I}_γ to \mathbf{I}_0 , and then from \mathbf{I}_0 to \mathbf{I} .

Moreover, varying the target distribution at each iteration (when updating γ) takes us outside the conditions of [theorem 8](#). Consequently, asymptotic confidence intervals constructed using all past simulations are no longer accessible.

One possible solution is to divide the AIS algorithm into two phases. During a first "burn-in" phase, the target distribution evolves according to one of the described schemes. Once the target distribution corresponds to \mathbf{g}^* , all simulations generated during this phase are discarded, but the importance distribution from the last iteration is retained. In the second phase, the AIS method continues with recycling but without varying the target distribution or using simulations from the first phase.

Other refinements to the AIS method, such as adaptive multiple importance sampling (AMIS [[MPS19](#)]), can be implemented during phase one without disrupting the convergence and asymptotic normality of the estimator produced in phase two.

7.2.2 Bandits: Best Arm Identification

The initial problem was estimating the failure probability of an industrial system for a given distribution, represented by the failure and repair rates of its components. One expected use of our method is determining, among several configurations (i.e., among multiple reference distributions $\mathbf{p}_1, \dots, \mathbf{p}_K$), which is the most reliable:

$$(\text{QoI}) \quad \arg \min_{i \in \llbracket 1, K \rrbracket} \mathbb{E}_{\mathbf{X} \sim \mathbf{p}_i} [\mathbf{I}(\mathbf{X})]. \quad (7.11)$$

A naive approach would independently apply the AIS method K times (once for each expectation) and compare the K resulting estimators. If the distributions $(\mathbf{p}_1, \dots, \mathbf{p}_K)$ share a common support, the same sample can be used to estimate each expectation by simply adjusting the likelihood ratio. For any i.i.d. sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ from the distribution \mathbf{g} , and for any $i \in \llbracket 1, K \rrbracket$, we have (under appropriate support conditions):

$$\hat{\mathcal{I}}_i = \frac{1}{n} \sum_{k=1}^n \mathbf{I}(\mathbf{X}_k) \frac{\mathbf{p}_i(\mathbf{X}_k)}{\mathbf{g}(\mathbf{X}_k)}. \quad (7.12)$$

Best Arm Identification

This problem is a variant of the best arm identification (BAI) problem for multi-armed bandits [[AB10](#); [GGL12](#); [GK16](#)]. In its classical formulation:

- ♦ K distributions, referred to as "arms," are available,
- ♦ at each iteration, one arm can be chosen, generating an observation under this distribution,

- ◊ the goal is to identify the distribution with the highest mean.

The sequential arm selection strategy should minimize the total number of arm pulls at a fixed confidence level ("fixed-confidence setting") or maximize the confidence level of the final decision under a fixed simulation budget ("fixed-budget setting").

In our case, the objective is to find the arm with the lowest mean. Observations can be drawn under an importance distribution of our choice, providing joint information on all K expectations per draw. If the importance distribution is fixed (non-adaptive setting), the only decision involves stopping the method once the result is sufficiently confident. An interesting problem lies in determining the confidence level of the decision for a given sample size n .

With Adaptive Importance Sampling

To achieve optimal performance, it is expected that the importance distribution will need to be updated as simulations proceed. While the choice of the target distribution \mathbf{g}^* is clear for estimating the mean of a single arm, it becomes less straightforward when estimating the expectation of a vector:

$$\mathbb{E}_{\mathbf{X} \sim \mathbf{g}} \left[\frac{\mathbf{I}(\mathbf{X})}{\mathbf{g}(\mathbf{X})} \left(\mathbf{p}_j(\mathbf{X}) \right)_{i=1}^K \right]. \quad (7.13)$$

For the specific case of two arms, the density proportional to $|\mathbf{I}| \times |\mathbf{p}_1 - \mathbf{p}_2|$ allows the exact estimation of the difference between the two means in two draws.

Developing an adaptive importance sampling strategy capable of identifying the minimizer of a family of expectations while providing the confidence level associated with this decision remains, to our knowledge, an open problem.

APPENDIX

References

- [ASS18] Candy Abboud, Rachid Senoussi, and Samuel Soubeyrand. “Piecewise-deterministic Markov processes for spatio-temporal population dynamics”. In: *Statistical Inference for Piecewise-deterministic Markov Processes* (2018), pp. 209–255 (cit. on p. 47).
- [ABV23] Michael S Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. “Stochastic interpolants: A unifying framework for flows and diffusions”. In: *arXiv preprint arXiv:2303.08797* (2023) (cit. on p. 30).
- [Ale+17] Vedat Levi Alev, Nima Anari, Lap Chi Lau, and Shayan Oveis Gharan. “Graph clustering using effective resistance”. In: *arXiv preprint arXiv:1711.06530* (2017) (cit. on p. 137).
- [Ali+17] Hananeh Aliee, Emanuele Borgonovo, Michael Glaß, and Jürgen Teich. “On the Boolean extension of the Birnbaum importance to non-coherent systems”. In: *Reliability Engineering & System Safety* 160 (2017), pp. 191–200 (cit. on p. 132).
- [And22] Charly Andral. “An attempt to trace the birth of importance sampling”. In: *arXiv preprint arXiv:2206.12286* (2022) (cit. on p. 16).
- [AAT92] George L Ang, Alfredo H-S Ang, and Wilson H Tang. “Optimal importance-sampling density estimator”. In: *Journal of engineering mechanics* 118.6 (1992), pp. 1146–1163 (cit. on p. 28).
- [AB01] Siu-Kui Au and James L Beck. “Estimation of small failure probabilities in high dimensions by subset simulation”. In: *Probabilistic engineering mechanics* 16.4 (2001), pp. 263–277 (cit. on p. 13).
- [AB10] Jean-Yves Audibert and Sébastien Bubeck. “Best arm identification in multi-armed bandits”. In: *COLT-23th Conference on learning theory-2010*. 2010, 13–p (cit. on pp. 161, 205).
- [APB22] Grégoire Aufort, Pierre Pudlo, and Denis Burgarella. “Tempered, Anti-truncated, Multiple Importance Sampling”. In: *arXiv preprint arXiv:2205.01501* (2022) (cit. on p. 30).

- [AG17] Necdet Serhat Aybat and Mert Gürbüzbalaban. “Decentralized computation of effective resistances and acceleration of consensus algorithms”. In: *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2017, pp. 538–542 (cit. on p. 145).
- [BMM13] Mathieu Balesdent, Jerome Morio, and Julien Marzat. “Kriging-based adaptive importance sampling algorithms for rare event estimation”. In: *Structural Safety* 44 (2013), pp. 1–10 (cit. on p. 28).
- [BK16] Polina V Banushkina and Sergei V Krivov. “Optimal reaction coordinates”. In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 6.6 (2016), pp. 748–763 (cit. on p. 78).
- [Bee02] Sally C Beeson. “Non-coherent fault tree analysis”. PhD thesis. Loughborough University, 2002 (cit. on p. 132).
- [Ben+16] Anis Ben Abdesslem, Romain Azaïs, Marie Touzet-Cortina, Anne Gégout-Petit, and Monique Puiggali. “Stochastic modelling and prediction of fatigue crack propagation using piecewise-deterministic Markov processes”. In: *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 230.4 (2016), pp. 405–416 (cit. on p. 46).
- [BH22] Michel Benaïm and Tobias Hurth. *Markov Chains on Metric Spaces: A Short Course*. Springer Nature, 2022 (cit. on p. 8).
- [Ben+15] Michel Benaïm, Stéphane Le Borgne, Florent Malrieu, and Pierre-André Zitt. “Qualitative properties of certain piecewise deterministic Markov processes”. In: *Annales de l’IHP Probabilités et statistiques*. Vol. 51. 3. 2015, pp. 1040–1075 (cit. on p. 47).
- [BBL08] Thomas Bengtsson, Peter Bickel, and Bo Li. “Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems”. In: *Probability and statistics: Essays in honor of David A. Freedman*. Vol. 2. Institute of Mathematical Statistics, 2008, pp. 316–335 (cit. on p. 30).
- [Bero5] Pavel Berkhin. “A survey on PageRank computing”. In: *Internet mathematics* 2.1 (2005), pp. 73–120 (cit. on p. 136).
- [Ber23] A Bertazzi. “Approximations and transformations of piecewise deterministic Monte Carlo algorithms”. PhD thesis. Delft University of Technology, 2023 (cit. on p. 47).
- [BBD22] Andrea Bertazzi, Joris Bierkens, and Paul Dobson. “Approximations of piecewise deterministic Markov processes and their convergence properties”. In: *Stochastic Processes and their Applications* 154 (2022), pp. 91–153 (cit. on p. 56).

- [BH05] Robert B Best and Gerhard Hummer. “Reaction coordinates and rates from transition paths”. In: *Proceedings of the National Academy of Sciences* 102.19 (2005), pp. 6732–6737 (cit. on p. 89).
- [Bet17] Michael Betancourt. “A conceptual introduction to Hamiltonian Monte Carlo”. In: *arXiv preprint arXiv:1701.02434* (2017) (cit. on p. 9).
- [Bha22] Soumendranath Bhakat. “Collective variable discovery in the age of machine learning: reality, hype and everything in between”. In: *RSC advances* 12.38 (2022), pp. 25010–25024 (cit. on p. 89).
- [BFR19] Joris Bierkens, Paul Fearnhead, and Gareth Roberts. “The zig-zag process and super-efficient sampling for Bayesian analysis of big data”. In: (2019) (cit. on p. 47).
- [Bit+23] Andreas Bittracher, Mattes Mollenhauer, Péter Koltai, and Christof Schütte. “Optimal reaction coordinates: Variational characterization and sparse computation”. In: *Multiscale Modeling & Simulation* 21.2 (2023), pp. 449–488 (cit. on p. 89).
- [Bje88] Peter Bjerager. “Probability integration by directional simulation”. In: *Journal of Engineering Mechanics* 114.8 (1988), pp. 1285–1302 (cit. on p. 12).
- [Bol21] Daniel Boley. “On fast computation of directed graph Laplacian pseudo-inverse”. In: *Linear Algebra and its Applications* 623 (2021), pp. 128–148 (cit. on p. 145).
- [BK11] Zdravko I Botev and Dirk P Kroese. “The generalized cross entropy method, with applications to probability density estimation”. In: *Methodology and Computing in Applied Probability* 13 (2011), pp. 1–27 (cit. on p. 28).
- [Bot+13] Zdravko I Botev, Dirk P Kroese, Reuven Y Rubinstein, and Pierre L’Ecuyer. “The cross-entropy method for optimization”. In: *Handbook of statistics*. Vol. 31. Elsevier, 2013, pp. 35–59 (cit. on p. 28).
- [BVD18] Alexandre Bouchard-Côté, Sebastian J Vollmer, and Arnaud Doucet. “The bouncy particle sampler: A nonreversible rejection-free Markov chain Monte Carlo method”. In: *Journal of the American Statistical Association* 113.522 (2018), pp. 855–867 (cit. on p. 47).
- [Bra05] Matthew Brand. “A random walks perspective on maximizing satisfaction and profit”. In: *Proceedings of the 2005 SLAM international conference on data mining*. SIAM, 2005, pp. 12–19 (cit. on p. 137).
- [Bré+16] Charles-Edouard Bréhier, Maxime Gazeau, Ludovic Goudenège, Tony Lelièvre, and Mathias Rousset. “Unbiasedness of some generalized adaptive multilevel splitting algorithms”. In: (2016) (cit. on p. 15).

- [Bri+21] Nicolae Brinzei, Jean-François Aubry, Miroslav Kvassay, and Elena Zaitseva. “System Reliability Analysis and Assessment by Means of Logic Differential Calculus and Hasse Diagram”. In: *Reliability Engineering and Computational Intelligence*. Springer, 2021, pp. 3–21 (cit. on p. 123).
- [BA18] Nicolae Brinzei and Jean-François Aubry. “Graphs models and algorithms for reliability assessment of coherent and non-coherent systems”. In: *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 232.2 (2018), pp. 201–215 (cit. on p. 132).
- [Buc04] James Antonio Bucklew. *Introduction to rare event simulation*. Vol. 5. Springer, 2004 (cit. on pp. xix, 16, 24, 191).
- [BM24] Evelyn Buckwar and Amira Meddah. “Numerical Approximations and Convergence Analysis of Piecewise Diffusion Markov Processes, with Application to Glioma Cell Migration”. In: *arXiv preprint arXiv:2401.13428* (2024) (cit. on p. 56).
- [BS20] Carlos E Budde and Mariëlle Stoelinga. “Automated rare event simulation for fault tree analysis via minimal cut sets”. In: *Measurement, Modelling and Evaluation of Computing Systems: 20th International GI/ITG Conference, MMB 2020, Saarbrücken, Germany, March 16–18, 2020, Proceedings 20*. Springer, 2020, pp. 259–277 (cit. on pp. 89, 123).
- [Bug+17] Monica F Bugallo, Victor Elvira, Luca Martino, David Luengo, Joaquin Miguez, and Petar M Djuric. “Adaptive importance sampling: The past, the present, and the future”. In: *IEEE Signal Processing Magazine* 34.4 (2017), pp. 60–79 (cit. on pp. 16, 23).
- [BL06] Manuela L Bujorianu and John Lygeros. “Toward a general theory of stochastic hybrid systems”. In: *Stochastic hybrid systems: theory and safety critical applications* (2006), pp. 3–30 (cit. on pp. 158, 202).
- [BL19] Jonathon Byrd and Zachary Lipton. “What is the effect of importance weighting in deep learning?” In: *International conference on machine learning*. PMLR, 2019, pp. 872–881 (cit. on p. 18).
- [Can+19] B Can, S Soori, NS Aybat, MM Dehvani, and M Gürbüzbalaban. “Decentralized computation of effective resistances and acceleration of distributed optimization algorithms. arXiv e-print”. In: *arXiv preprint arXiv:1907.13110* (2019) (cit. on p. 145).
- [Car+22] Gabriel Cardoso, Sergey Samsonov, Achille Thin, Eric Moulines, and Jimmy Olsson. “Br-snis: bias reduced self-normalized importance sampling”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 716–729 (cit. on p. 21).

- [Čep11] Marko Čepin. *Assessment of power system reliability: methods and applications*. Springer Science & Business Media, 2011 (cit. on p. 123).
- [Cér+12] Frédéric Cérou, Pierre Del Moral, Teddy Furon, and Arnaud Guyader. “Sequential Monte Carlo for rare event estimation”. In: *Statistics and computing* 22.3 (2012), pp. 795–808 (cit. on pp. 14, 15).
- [Cér+19] Frédéric Cérou, Bernard Delyon, Arnaud Guyader, and Mathias Rousset. “On the asymptotic normality of adaptive multilevel splitting”. In: *SIAM/ASA Journal on Uncertainty Quantification* 7.1 (2019), pp. 1–30 (cit. on p. 15).
- [CG07] Frédéric Cérou and Arnaud Guyader. “Adaptive multilevel splitting for rare event analysis”. In: *Stochastic Analysis and Applications* 25.2 (2007), pp. 417–443 (cit. on p. 15).
- [CGR19] Frédéric Cérou, Arnaud Guyader, and Mathias Rousset. “Adaptive multilevel splitting: Historical perspective and recent results”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29.4 (2019) (cit. on pp. 13, 78, 160, 203).
- [Cha+18] Vincent Chabridon, Mathieu Balesdent, Jean-Marc Bourinet, Jérôme Morio, and Nicolas Gayton. “Reliability-based sensitivity estimators of rare event probability in the presence of distribution parameter uncertainty”. In: *Reliability Engineering & System Safety* 178 (2018), pp. 164–178 (cit. on p. 18).
- [CD18] Sourav Chatterjee and Persi Diaconis. “The sample size required in importance sampling”. In: *The Annals of Applied Probability* 28.2 (2018), pp. 1099–1135 (cit. on p. 19).
- [CZ08] Haiyan Chen and Fuji Zhang. “The expected hitting times for finite Markov chains”. In: *Linear Algebra and its Applications* 428.11–12 (2008), pp. 2730–2749 (cit. on p. 137).
- [Che21] Ming Chen. “Collective variable-based enhanced sampling and machine learning”. In: *The European Physical Journal B* 94 (2021), pp. 1–17 (cit. on p. 89).
- [Che+23] Yuhan Chen, Haojie Ye, Sanketh Vedula, Alex Bronstein, Ronald Dreslinski, Trevor Mudge, and Nishil Talati. “Demystifying graph sparsification algorithms in graph properties preservation”. In: *Proceedings of the VLDB Endowment* 17.3 (2023), pp. 427–440 (cit. on p. 136).
- [Che+24a] Guillaume Chennetier, Hassane Chraïbi, Anne Dutfoy, and Josselin Garnier. “Adaptive importance sampling based on fault tree analysis for piecewise deterministic Markov process”. In: *SIAM/ASA Journal on Uncertainty Quantification* 12.1 (2024), pp. 128–156 (cit. on pp. xxii, 108, 156, 194, 199).

- [Che+24b] Guillaume Chennetier, Hassane Chraïbi, Anne Dutfoy, and Josselin Garnier. “Graph-informed importance sampling for rare event estimation with piecewise deterministic Markov process”. In: (Submission in 2024) (cit. on pp. [xxii](#), [156](#), [194](#), [199](#)).
- [CP20] Nicolas Chopin and Omiros Papaspiliopoulos. *An introduction to sequential Monte Carlo*. Vol. 4. Springer, 2020 (cit. on p. [12](#)).
- [Chr+19] H Chraïbi, A Dutfoy, T Galtier, and J Garnier. “On the optimal importance process for piecewise deterministic markov process”. In: *ESAIM: Probability and Statistics* 23 (2019), pp. 893–921 (cit. on pp. [xix](#), [xx](#), [66](#), [71](#), [76](#), [79–83](#), [113](#), [114](#), [116](#), [191](#)).
- [Chr+21] Hassane Chraïbi, Anne Dutfoy, Thomas Galtier, and Josselin Garnier. “Optimal potential functions for the interacting particle system method”. In: *Monte Carlo Methods and Applications* 27.2 (2021), pp. 137–152 (cit. on pp. [xix](#), [191](#)).
- [CHS16] Hassane Chraïbi, JC Houbedine, and Alain Sibler. “PyCATSHOO: Toward a new platform dedicated to dynamic reliability assessments of hybrid systems”. In: *PSAM congress*. 2016 (cit. on pp. [xvii](#), [116](#), [126](#), [127](#), [189](#)).
- [Clo+17] Bertrand Cloez, Renaud Dessalles, Alexandre Genadot, Florent Malrieu, Aline Marguet, and Romain Yvinec. “Probabilistic and piecewise deterministic models in biology”. In: *ESAIM: Proceedings and Surveys* 60 (2017), pp. 225–245 (cit. on p. [47](#)).
- [Con+20] Alessio Conte, Pierluigi Crescenzi, Andrea Marino, and Giulia Punzi. “Enumeration of sd separators in DAGs with application to reliability analysis in temporal graphs”. In: *45th International Symposium on Mathematical Foundations of Computer Science (MFCS 2020)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik. 2020 (cit. on p. [136](#)).
- [CDo8] Oswaldo LV Costa and François Dufour. “Stability and ergodicity of piecewise deterministic Markov processes”. In: *SIAM Journal on Control and Optimization* 47.2 (2008), pp. 1053–1077 (cit. on p. [47](#)).
- [CHI1] Yves Crama and Peter L Hammer. *Boolean functions: Theory, algorithms, and applications*. Cambridge University Press, 2011 (cit. on p. [119](#)).
- [Dai02] Yu-Hong Dai. “Convergence properties of the BFGS algorithm”. In: *SIAM Journal on Optimization* 13.3 (2002), pp. 693–701 (cit. on p. [105](#)).
- [DJ24] Agnimitra Dasgupta and Erik A Johnson. “REIN: Reliability Estimation via Importance sampling with Normalizing flows”. In: *Reliability Engineering & System Safety* 242 (2024), p. 109729 (cit. on p. [29](#)).

- [Dav84] Mark HA Davis. “Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 46.3 (1984), pp. 353–376 (cit. on pp. xvii, 46, 102, 189).
- [Dav18] Mark HA Davis. *Markov models & optimization*. Routledge, 2018 (cit. on pp. 46, 53–55).
- [De +05] Pieter-Tjerk De Boer, Dirk P Kroese, Shie Mannor, and Reuven Y Rubinstein. “A tutorial on the cross-entropy method”. In: *Annals of operations research* 134 (2005), pp. 19–67 (cit. on p. 27).
- [DDZ15] 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 (cit. on pp. xvii, 46, 189).
- [DCM19] Daniel Delahaye, Supatcha Chaimatanan, and Marcel Mongeau. “Simulated annealing: From basics to applications”. In: *Handbook of metaheuristics* (2019), pp. 1–35 (cit. on pp. 4, 30).
- [DP21] Bernard Delyon and F Portier. “Safe adaptive importance sampling: A mixture approach”. In: (2021) (cit. on p. 32).
- [DP18] Bernard Delyon and François Portier. “Asymptotic optimality of adaptive importance sampling”. In: *Advances in Neural Information Processing Systems* 31 (2018) (cit. on pp. 23, 97–99, 101).
- [Dem+23] Julien Demange-Chryst, François Bachoc, Jérôme Morio, and Timothé Krauth. “Variational autoencoder with weighted samples for high-dimensional non-parametric adaptive importance sampling”. In: *arXiv preprint arXiv:2310.09194* (2023) (cit. on p. 29).
- [Des+21] Loïc Desgeorges, Pierre-Yves Piriou, Thibault Lemattre, and Hassane Chraïbi. “Formalism and semantics of PyCATSHOO: A simulator of distributed stochastic hybrid automata”. In: *Reliability Engineering & System Safety* 208 (2021), p. 107384 (cit. on pp. xvii, 46, 189).
- [Dio+17] CM Diop, E Dumonteil, H Louvin TL, and M Rousset. “Adaptive Multilevel Splitting for Monte Carlo particle transport”. In: *Eur. Phys. J. N* (2017) (cit. on p. 15).
- [DG21] Qiming Du and Arnaud Guyader. “Variance estimation in adaptive sequential Monte Carlo”. In: *The Annals of Applied Probability* 31.3 (2021), pp. 1021–1060 (cit. on p. 15).

- [DGM21] Alain Durmus, Arnaud Guillin, and Pierre Monmarché. “Piecewise deterministic Markov processes and their invariant measures”. In: *Annales de l’Institut Henri Poincaré (B) Probabilités et statistiques*. Vol. 57. 3. Institut Henri Poincaré. 2021, pp. 1442–1475 (cit. on p. 47).
- [Dyn65] Eugene Borisovich Dynkin. *Markov processes*. Springer, 1965 (cit. on p. 142).
- [Eck87] Roger Eckhardt. “Stan Ulam, John von Neumann and the Monte Carlo Method”. In: *Los Alamos Science* 100.15 (1987), p. 131 (cit. on p. 10).
- [Ehr+23] Max Ehre, Rafael Flock, Martin Fußeder, Jason Papaioannou, and Daniel Straub. “Certified dimension reduction for Bayesian updating with the cross-entropy method”. In: *SIAM/ASA Journal on Uncertainty Quantification* 11.1 (2023), pp. 358–388 (cit. on p. 32).
- [EMS21a] Maxime El Masri, Jérôme Morio, and Florian Simatos. “Improvement of the cross-entropy method in high dimension for failure probability estimation through a one-dimensional projection without gradient estimation”. In: *Reliability Engineering & System Safety* 216 (2021), p. 107991 (cit. on p. 32).
- [EMS21b] Maxime El Masri, Jérôme Morio, and Florian Simatos. “Optimal projection to improve parametric importance sampling in high dimension”. In: *arXiv preprint arXiv:2107.06091* (2021) (cit. on p. 32).
- [EM21] Víctor Elvira and Luca Martino. “Advances in importance sampling”. In: *arXiv preprint arXiv:2102.05407* (2021) (cit. on p. 16).
- [Elv+19] Víctor Elvira, Luca Martino, David Luengo, and Mónica F Bugallo. “Generalized Multiple Importance Sampling”. In: *Statistical Science* 34.1 (2019), pp. 129–155 (cit. on pp. 22, 23).
- [ES94] Paul Embrechts and Hanspeter Schmidli. “Ruin estimation for a general insurance risk model”. In: *Advances in applied probability* 26.2 (1994), pp. 404–422 (cit. on p. 46).
- [FKH13] Giacomo Fiorin, Michael L Klein, and Jérôme Hénin. “Using collective variables to drive molecular dynamics simulations”. In: *Molecular Physics* 111.22–23 (2013), pp. 3345–3362 (cit. on p. 89).
- [Fou+07] Francois Fouss, Alain Pirotte, Jean-Michel Renders, and Marco Saerens. “Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation”. In: *IEEE Transactions on knowledge and data engineering* 19.3 (2007), pp. 355–369 (cit. on p. 137).
- [FG13] Bert E Fristedt and Lawrence F Gray. *A modern approach to probability theory*. Springer Science & Business Media, 2013 (cit. on p. 37).

- [GGL12] Victor Gabillon, Mohammad Ghavamzadeh, and Alessandro Lazaric. “Best arm identification: A unified approach to fixed budget and fixed confidence”. In: *Advances in Neural Information Processing Systems* 25 (2012) (cit. on pp. 161, 205).
- [Gal19] Thomas Galtier. “Accelerated Monte-Carlo methods for piecewise deterministic Markov processes for a faster reliability assessment of power generation systems within the PyCATSHOO toolbox”. PhD thesis. Université de Paris, 2019 (cit. on pp. xix, 66, 71, 72, 74–76, 80, 157, 191, 200).
- [GK16] Aurélien Garivier and Emilie Kaufmann. “Optimal best arm identification with fixed confidence”. In: *Conference on Learning Theory*. PMLR. 2016, pp. 998–1027 (cit. on pp. 161, 205).
- [Garoo] Marnix Joseph Johann Garvels. “The splitting method in rare event simulation”. In: (2000) (cit. on p. 15).
- [GLJ19] Aditi Ghai, Cao Lu, and Xiangmin Jiao. “A comparison of preconditioned Krylov subspace methods for large-scale nonsymmetric linear systems”. In: *Numerical Linear Algebra with Applications* 26.1 (2019), e2215 (cit. on p. 146).
- [GK13] Robert Ghrist and Sanjeevi Krishnan. “A topological max-flow-min-cut theorem”. In: *2013 IEEE Global Conference on Signal and Information Processing*. IEEE. 2013, pp. 815–818 (cit. on p. 136).
- [Gil15] Michael B Giles. “Multilevel monte carlo methods”. In: *Acta numerica* 24 (2015), pp. 259–328 (cit. on p. 15).
- [GC11] Mark Girolami and Ben Calderhead. “Riemann manifold langevin and hamiltonian monte carlo methods”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 73.2 (2011), pp. 123–214 (cit. on p. 9).
- [GR96] Geof H Givens and Adrian E Raftery. “Local adaptive importance sampling for multivariate densities with strong nonlinear relationships”. In: *Journal of the American Statistical Association* 91.433 (1996), pp. 132–141 (cit. on p. 28).
- [GS02] Peter W Glynn and Roberto Szechtman. “Some new perspectives on the method of control variates”. In: *Monte Carlo and Quasi-Monte Carlo Methods 2000: Proceedings of a Conference held at Hong Kong Baptist University, Hong Kong SAR, China, November 27–December 1, 2000*. Springer. 2002, pp. 27–49 (cit. on p. 12).
- [GHL22] Branda Goncalves, Thierry Huillet, and Eva Löcherbach. “On population growth with catastrophes”. In: *Stochastic Models* 38.2 (2022), pp. 214–249 (cit. on pp. 47, 132).

- [Gra23] Sebastiano Grazi. “Design and applications of Monte Carlo methods based on piecewise deterministic Markov processes”. PhD thesis. Delft University of Technology, 2023 (cit. on p. 47).
- [Gri+19] Pascal Grittmann, Iliyan Georgiev, Philipp Slusallek, and Jaroslav Křivánek. “Variance-aware multiple importance sampling”. In: *ACM Transactions on Graphics (TOG)* 38.6 (2019), pp. 1–9 (cit. on p. 23).
- [Gua98] Stephen Guattery. *Graph embeddings, symmetric real matrices, and generalized inverses*. 98. Citeseer, 1998 (cit. on p. 137).
- [GCE23] Thomas Guilmeau, Emilie Chouzenoux, and Víctor Elvira. “On variational inference and maximum likelihood estimation with the λ -exponential family”. In: *arXiv preprint arXiv:2310.05781* (2023) (cit. on p. 27).
- [Guo+24] Qintian Guo, Dandan Lin, Sibao Wang, Raymond Chi-Wing Wong, and Wenqing Lin. “Efficient Algorithms for Group Hitting Probability Queries on Large Graphs”. In: *IEEE Transactions on Knowledge and Data Engineering* (2024) (cit. on p. 145).
- [HM56] John Michael Hammersley and Keith William Morton. “A new Monte Carlo technique: antithetic variates”. In: *Mathematical proceedings of the Cambridge philosophical society*. Vol. 52. 3. Cambridge University Press, 1956, pp. 449–475 (cit. on p. 12).
- [Has70] W Keith Hastings. “Monte Carlo sampling methods using Markov chains and their applications”. In: (1970) (cit. on p. 8).
- [Havo2] Taher H Haveliwala. “Topic-sensitive pagerank”. In: *Proceedings of the 11th international conference on World Wide Web*. 2002, pp. 517–526 (cit. on p. 136).
- [HO14] Hera Y He and Art B Owen. “Optimal mixture weights in multiple importance sampling”. In: *arXiv preprint arXiv:1411.3954* (2014) (cit. on p. 23).
- [Her+17] Ulysse Herbach, Arnaud Bonnaïfoux, Thibault Espinasse, and Olivier Gandrillon. “Inferring gene regulatory networks from single-cell data: a mechanistic approach”. In: *BMC systems biology* 11.1 (2017), pp. 1–15 (cit. on p. 47).
- [Hes95] Tim Hesterberg. “Weighted average importance sampling and defensive mixture distributions”. In: *Technometrics* 37.2 (1995), pp. 185–194 (cit. on p. 32).
- [HG+14] Matthew D Hoffman, Andrew Gelman, et al. “The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo.” In: *J. Mach. Learn. Res.* 15.1 (2014), pp. 1593–1623 (cit. on p. 9).

- [Hol+24] Lars Holdijk, Yuanqi Du, Ferry Hooft, Priyank Jaini, Berend Ensing, and Max Welling. “Stochastic Optimal Control for Collective Variable Free Sampling of Molecular Transition Paths”. In: *Advances in Neural Information Processing Systems* 36 (2024) (cit. on p. 89).
- [HZ02] Feifang Hu and James V Zidek. “The weighted likelihood”. In: *Canadian Journal of Statistics* 30.3 (2002), pp. 347–371 (cit. on p. 30).
- [IS18] Ayyoub Imakhlaf and Mohamed Sallak. “Birnbbaum’s measure extension for non-coherent systems based on BDD”. In: *2018 Annual Reliability and Maintainability Symposium (RAMS)*. IEEE, 2018, pp. 1–7 (cit. on p. 132).
- [Iono8] Edward L Ionides. “Truncated importance sampling”. In: *Journal of Computational and Graphical Statistics* 17.2 (2008), pp. 295–311 (cit. on p. 30).
- [JWo2] Glen Jeh and Jennifer Widom. “Simrank: a measure of structural-context similarity”. In: *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. 2002, pp. 538–543 (cit. on p. 136).
- [KH51] Herman Kahn and Theodore E Harris. “Estimation of particle transmission by random sampling”. In: *National Bureau of Standards applied mathematics series* 12 (1951), pp. 27–30 (cit. on pp. 13, 16).
- [Kar+19] O Karlik, M Sik, P Vévoda, T Skrivan, and J Krivanek. “MIS compensation: optimizing sampling techniques in multiple importance sampling”. In: *ACM Transactions on Graphics (TOG)* 38.6 (2019), pp. 1–12 (cit. on p. 23).
- [Kat53] Leo Katz. “A new status index derived from sociometric analysis”. In: *Psychometrika* 18.1 (1953), pp. 39–43 (cit. on p. 136).
- [KB+18] Youngjoo Kim, Hyochoong Bang, et al. “Introduction to Kalman filter and its applications”. In: *Introduction and Implementations of the Kalman Filter* 1 (2018), pp. 1–16 (cit. on p. 12).
- [Kin92] John Frank Charles Kingman. *Poisson processes*. Vol. 3. Clarendon Press, 1992 (cit. on p. 42).
- [KTD19] Thomas Krak, Natan T’Joens, and Jasper De Bock. “Hitting times and probabilities for imprecise Markov chains”. In: *International Symposium on Imprecise Probabilities: Theories and Applications*. PMLR, 2019, pp. 265–275 (cit. on p. 145).
- [Kre23] Nathalie Krell. “Can Piecewise deterministic Markov process help biologist?” PhD thesis. Université de Rennes, 2023 (cit. on p. 46).

- [Kri+19] Peter Kritzer, Gunther Leobacher, Michaela Szölgényi, and Stefan Thonhauser. “Approximation methods for piecewise deterministic Markov processes and their costs”. In: *Scandinavian actuarial journal* 2019.4 (2019), pp. 308–335 (cit. on pp. 46, 56, 132).
- [Kri13] Sergei V Krivov. “On reaction coordinate optimality”. In: *Journal of chemical theory and computation* 9.1 (2013), pp. 135–146 (cit. on p. 89).
- [KR04] Dirk P Kroese and Reuven Y Rubinstein. “The transform likelihood ratio method for rare event simulation with heavy tails”. In: *Queueing Systems* 46 (2004), pp. 317–351 (cit. on pp. 26, 27).
- [KS07] Jérôme Kunegis and Stephan Schmidt. “Collaborative filtering using electrical resistance network models”. In: *Advances in Data Mining. Theoretical Aspects and Applications: 7th Industrial Conference, ICDM 2007, Leipzig, Germany, July 14-18, 2007. Proceedings* 7. Springer, 2007, pp. 269–282 (cit. on p. 137).
- [KK20] Kazuhiro Kurita and Yasuaki Kobayashi. “Efficient enumerations for minimal multicuts and multiway cuts”. In: *arXiv preprint arXiv:2006.16222* (2020) (cit. on p. 136).
- [La 20] Arnaud de La Fortelle. “Generalized inverses of increasing functions and Lebesgue decomposition”. In: *Markov Processes And Related Fields* (2020) (cit. on p. 5).
- [Lee+85] Wen-Shing Lee, Doris L Grosh, Frank A Tillman, and Chang H Lie. “Fault tree analysis, methods, and applications - a review”. In: *IEEE transactions on reliability* 34.3 (1985), pp. 194–203 (cit. on p. 123).
- [LTT18] Vincent Lemaire, Michèle Thieullen, and Nicolas Thomas. “Exact simulation of the jump times of a class of piecewise deterministic Markov processes”. In: *Journal of Scientific Computing* 75 (2018), pp. 1776–1807 (cit. on pp. 47, 62).
- [LS79] PA W Lewis and Gerald S Shedler. “Simulation of nonhomogeneous Poisson processes by thinning”. In: *Naval research logistics quarterly* 26.3 (1979), pp. 403–413 (cit. on p. 61).
- [LL20] Enzhi Li and Zhengyi Le. “Frustrated random walks: A faster algorithm to evaluate node distances on connected and undirected graphs”. In: *Physical review E* 102.5 (2020), p. 052135 (cit. on p. 145).
- [Lia+23] Meihao Liao, Rong-Hua Li, Qiangqiang Dai, Hongyang Chen, Hongchao Qin, and Guoren Wang. “Efficient Resistance Distance Computation: the Power of Landmark-based Approaches”. In: *Proceedings of the ACM on Management of Data* 1.1 (2023), pp. 1–27 (cit. on pp. 136, 145).

- [LB18] Yen Ting Lin and Nicolas E Buchler. “Efficient analysis of stochastic gene dynamics in the non-adiabatic regime using piecewise deterministic Markov processes”. In: *Journal of The Royal Society Interface* 15.138 (2018), p. 20170804 (cit. on p. 47).
- [LL21] Chunyan Ling and Zhenzhou Lu. “Support vector machine-based importance sampling for rare event estimation”. In: *Structural and Multidisciplinary Optimization* 63 (2021), pp. 1609–1631 (cit. on p. 29).
- [Lou+17] Henri Louvin, Eric Dumonteil, Tony Lelièvre, Mathias Rousset, and Cheikh M Diop. “Adaptive multilevel splitting for Monte Carlo particle transport”. In: *EPJ Web of Conferences*. Vol. 153. EDP Sciences. 2017, p. 06006 (cit. on pp. 89, 160, 203).
- [Lov93] László Lovász. “Random walks on graphs”. In: *Combinatorics, Paul erdos is eighty* 2.1-46 (1993), p. 4 (cit. on p. 136).
- [Lu+08] Zhenzhou Lu, Shufang Song, Zhufeng Yue, and Jian Wang. “Reliability sensitivity method by line sampling”. In: *Structural Safety* 30.6 (2008), pp. 517–532 (cit. on p. 12).
- [MB23] Hao Ma and Henk AP Blom. “Interacting particle system based estimation of reach probability of general stochastic hybrid systems”. In: *Nonlinear Analysis: Hybrid Systems* 47 (2023), p. 101303 (cit. on pp. 158, 202).
- [MPS19] Jean-Michel Marin, Pierre Pudlo, and Mohammed Sedki. “Consistency of adaptive importance sampling and recycling schemes”. In: *Bernoulli* 25.3 (2019), pp. 1977–1998 (cit. on pp. 23, 161, 204).
- [Mar+18] Luca Martino, Victor Elvira, J Míguez, A Artés-Rodríguez, and PM Djurić. “A comparison of clipping strategies for importance sampling”. In: *2018 IEEE Statistical Signal Processing Workshop (SSP)*. IEEE. 2018, pp. 558–562 (cit. on p. 30).
- [Met+87] Nicholas Metropolis et al. “The beginning of the Monte Carlo method”. In: *Los Alamos Science* 15.584 (1987), pp. 125–130 (cit. on p. 10).
- [Met+53] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. “Equation of state calculations by fast computing machines”. In: *The journal of chemical physics* 21.6 (1953), pp. 1087–1092 (cit. on p. 8).
- [MU49] Nicholas Metropolis and Stanislaw Ulam. “The monte carlo method”. In: *Journal of the American statistical association* 44.247 (1949), pp. 335–341 (cit. on p. 9).

- [Míg17] Joaquín Míguez. “On the performance of nonlinear importance samplers and population Monte Carlo schemes”. In: *2017 22nd International Conference on Digital Signal Processing (DSP)*. IEEE, 2017, pp. 1–5 (cit. on p. 30).
- [MMV18] Joaquín Míguez, Inés P Mariño, and Manuel A Vázquez. “Analysis of a nonlinear importance sampling scheme for Bayesian parameter estimation in state-space models”. In: *Signal Processing* 142 (2018), pp. 281–291 (cit. on p. 30).
- [Mit98] Melanie Mitchell. *An introduction to genetic algorithms*. MIT press, 1998 (cit. on p. 4).
- [Moro4] Pierre Moral. *Feynman-Kac formulae: genealogical and interacting particle systems with applications*. Springer, 2004 (cit. on p. 12).
- [MLo9] Carl N Morris and Kari F Lock. “Unifying the named natural exponential families and their relatives”. In: *The American Statistician* 63.3 (2009), pp. 247–253 (cit. on p. 26).
- [MPP23] Line Mouaffac, Karen Palacio-Rodriguez, and Fabio Pietrucci. “Optimal reaction coordinates and kinetic rates from the projected dynamics of transition paths”. In: *Journal of Chemical Theory and Computation* 19.17 (2023), pp. 5701–5711 (cit. on pp. 78, 89).
- [Mül+19] Thomas Müller, Brian McWilliams, Fabrice Rousselle, Markus Gross, and Jan Novák. “Neural importance sampling”. In: *ACM Transactions on Graphics (ToG)* 38.5 (2019), pp. 1–19 (cit. on p. 29).
- [Mun+16] Rémi Munos, Tom Stepleton, Anna Harutyunyan, and Marc Bellemare. “Safe and efficient off-policy reinforcement learning”. In: *Advances in neural information processing systems* 29 (2016) (cit. on p. 18).
- [NB18] Eisha Nathan and David A Bader. “Incrementally updating Katz centrality in dynamic graphs”. In: *Social Network Analysis and Mining* 8.1 (2018), p. 26 (cit. on p. 136).
- [Ned09] Jan C Neddermeyer. “Computationally efficient nonparametric importance sampling”. In: *Journal of the American Statistical Association* 104.486 (2009), pp. 788–802 (cit. on p. 28).
- [NJ10] Alexander G Nikolaev and Sheldon H Jacobson. “Simulated annealing”. In: *Handbook of metaheuristics* (2010), pp. 1–39 (cit. on p. 4).
- [OZ00] Art Owen and Yi Zhou. “Safe and effective importance sampling”. In: *Journal of the American Statistical Association* 95.449 (2000), pp. 135–143 (cit. on p. 32).
- [Owe13] Art B. Owen. *Monte Carlo theory, methods and examples*. <https://artowen.su.domains/mc/>, 2013 (cit. on pp. 16, 18, 21).

- [Pag+99] Lawrence Page, Sergey Brin, Rajeev Motwani, Terry Winograd, et al. “The pagerank citation ranking: Bringing order to the web”. In: (1999) (cit. on p. 136).
- [Pap+15] Iason Papaioannou, Wolfgang Betz, Kilian Zwirgmaier, and Daniel Straub. “MCMC algorithms for subset simulation”. In: *Probabilistic Engineering Mechanics* 41 (2015), pp. 89–103 (cit. on p. 14).
- [PGS19] Iason Papaioannou, Sebastian Geyer, and Daniel Straub. “Improved cross entropy-based importance sampling with a flexible mixture model”. In: *Reliability Engineering & System Safety* 191 (2019), p. 106564 (cit. on pp. 28, 32).
- [Pap+19] Matteo Papini, Alberto Maria Metelli, Lorenzo Lupo, and Marcello Restelli. “Optimistic policy optimization via multiple importance sampling”. In: *International Conference on Machine Learning*. PMLR. 2019, pp. 4989–4999 (cit. on p. 23).
- [Pen+21] Pan Peng, Daniel Lopatta, Yuichi Yoshida, and Gramoz Goranci. “Local algorithms for estimating effective resistance”. In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2021, pp. 1329–1338 (cit. on p. 145).
- [PS15] Yuval Peres and Perla Sousi. “Mixing times are hitting times of large sets”. In: *Journal of Theoretical Probability* 28.2 (2015), pp. 488–519 (cit. on p. 137).
- [QH06] Huaijun Qiu and Edwin R Hancock. “Graph embedding using commute time”. In: *Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshops, SSPR 2006 and SPR 2006, Hong Kong, China, August 17-19, 2006. Proceedings*. Springer. 2006, pp. 441–449 (cit. on p. 137).
- [QH07] Huaijun Qiu and Edwin R Hancock. “Clustering and embedding using commute times”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.11 (2007), pp. 1873–1890 (cit. on p. 137).
- [RRU24] Alice Raffaele, Romeo Rizzi, and Takeaki Uno. “Listing the bonds of a graph in $O(n)$ -delay”. In: *Discrete Applied Mathematics* 348 (2024), pp. 105–121 (cit. on pp. 135, 136).
- [Rau93] Antoine Rauzy. “New algorithms for fault trees analysis”. In: *Reliability Engineering & System Safety* 40.3 (1993), pp. 203–211 (cit. on p. 123).
- [Rén61] Alfréd Rényi. “On measures of entropy and information”. In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. Vol. 4. University of California Press. 1961, pp. 547–562 (cit. on p. 27).
- [RC99] Christian P Robert and George Casella. *Monte Carlo statistical methods*. Vol. 2. Springer, 1999 (cit. on p. 6).

- [RR98] Christian P Robert and Sylvia Richardson. “Markov chain Monte Carlo methods”. In: *Discretization and MCMC Convergence Assessment*. Springer, 1998, pp. 1–25 (cit. on p. 9).
- [RT+09] Gerardo Rubino, Bruno Tuffin, et al. *Rare event simulation using Monte Carlo methods*. Vol. 73. Wiley Online Library, 2009 (cit. on pp. xix, 16, 191).
- [RK04] Reuven Y Rubinstein and Dirk P Kroese. *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning*. Vol. 133. Springer, 2004 (cit. on p. 24).
- [RK16] Reuven Y Rubinstein and Dirk P Kroese. *Simulation and the Monte Carlo method*. John Wiley & Sons, 2016 (cit. on pp. 24, 27, 29).
- [RT17] Ryszard Rudnicki and Marta Tyran-Kamińska. *Piecewise deterministic processes in biological models*. Vol. 1. Springer, 2017 (cit. on p. 46).
- [RS15] Enno Ruijters and Mariëlle Stoelinga. “Fault tree analysis: A survey of the state-of-the-art in modeling, analysis and tools”. In: *Computer science review* 15 (2015), pp. 29–62 (cit. on p. 123).
- [RA21] Ali Muhammad Ali Rushdi and Motaz Hussain Amashah. “Derivation of minimal cutsets from minimal pathsets for a multi-state system and utilization of both sets in checking reliability expressions”. In: *Journal of Engineering Research and Reports* 20.8 (2021), pp. 22–33 (cit. on p. 123).
- [Rus+22] Patrik Rusnak, Elena Zaitseva, Frank PA Coolen, Miroslav Kvassay, and Vitaly Levashenko. “Logic differential calculus for reliability analysis based on survival signature”. In: *IEEE Transactions on Dependable and Secure Computing* 20.2 (2022), pp. 1529–1540 (cit. on p. 123).
- [SS86] Youcef Saad and Martin H Schultz. “GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems”. In: *SIAM Journal on scientific and statistical computing* 7.3 (1986), pp. 856–869 (cit. on p. 146).
- [Saa03] Yousef Saad. *Iterative methods for sparse linear systems*. SIAM, 2003 (cit. on p. 146).
- [Saa20] Yousef Saad. “Iterative methods for linear systems of equations: A brief historical journey”. In: *Contemporary Mathematics, edited by: Brenner, S., Shparlinski, I., Shu, C.-W., and Szylid, D., American Mathematical Society, Providence, Rhode Island* 754 (2020), pp. 197–215 (cit. on p. 146).
- [SM12] Purnamrita Sarkar and Andrew Moore. “A tractable approach to finding closest truncated-commute-time neighbors in large graphs”. In: *arXiv preprint arXiv:1206.5259* (2012) (cit. on p. 145).

- [SE22] Mateu Sbert and Víctor Elvira. “Generalizing the balance heuristic estimator in multiple importance sampling”. In: *Entropy* 24.2 (2022), p. 191 (cit. on p. 23).
- [SHS16] Mateu Sbert, Vlastimil Havran, and László Szirmay-Kalos. “Variance Analysis of Multi-sample and One-sample Multiple Importance Sampling”. In: *Computer Graphics Forum*. Vol. 35. 7. Wiley Online Library. 2016, pp. 451–460 (cit. on p. 23).
- [SHS18] Mateu Sbert, Vlastimil Havran, and Laszlo Szirmay-Kalos. “Multiple importance sampling revisited: breaking the bounds”. In: *EURASIP Journal on Advances in Signal Processing* 2018.1 (2018), pp. 1–15 (cit. on p. 23).
- [SS23] Mateu Sbert and László Szirmay-Kalos. “A linear heuristic for multiple importance sampling”. In: *EURASIP Journal on Advances in Signal Processing* 1 (2023), p. 31 (cit. on p. 23).
- [Sch98] Manfred Schäl. “On piecewise deterministic Markov control processes: control of jumps and of risk processes in insurance”. In: *Insurance: Mathematics and Economics* 22.1 (1998), pp. 75–91 (cit. on p. 46).
- [Sed+21] Peter Sedlacek, Elena Zaitseva, Vitaly Levashenko, and Miroslav Kvassay. “Critical state of non-coherent multi-state system”. In: *Reliability Engineering & System Safety* 215 (2021), p. 107824 (cit. on p. 132).
- [SL13] Maria Carmen Segovia and Pierre-Etienne Labeau. “Reliability of a multi-state system subject to shocks using phase-type distributions”. In: *Applied mathematical modelling* 37.7 (2013), pp. 4883–4904 (cit. on p. 132).
- [SD24] Gabriel San Martín Silva and Enrique López Droguett. “Quantum Fault Trees and Minimal Cut Sets Identification”. In: *arXiv preprint arXiv:2404.05853* (2024) (cit. on p. 123).
- [Sob90] IM Sobol’. “Quasi-monte carlo methods”. In: *Progress in Nuclear Energy* 24.1-3 (1990), pp. 55–61 (cit. on p. 12).
- [SK23] Chenxiao Song and Reiichiro Kawai. “Monte Carlo and variance reduction methods for structural reliability analysis: A comprehensive review”. In: *Probabilistic Engineering Mechanics* (2023), p. 103479 (cit. on p. 12).
- [TJG22] Armin Tabandeh, Gaofeng Jia, and Paolo Gardoni. “A review and assessment of importance sampling methods for reliability analysis”. In: *Structural Safety* 97 (2022), p. 102216 (cit. on p. 20).
- [Tho19] Nicolas Thomas. “Stochastic numerical methods for Piecewise Deterministic Markov Processes: applications in Neuroscience”. PhD thesis. Sorbonne université, 2019 (cit. on p. 47).

- [Tij03] Henk C Tijms. *A first course in stochastic models*. John Wiley and sons, 2003 (cit. on p. 46).
- [TFP06] Hanghang Tong, Christos Faloutsos, and Jia-Yu Pan. “Fast random walk with restart and its applications”. In: *Sixth international conference on data mining (ICDM’06)*. IEEE. 2006, pp. 613–622 (cit. on p. 136).
- [TPZ16] Pietro Turati, Nicola Pedroni, and Enrico Zio. “Advanced RESTART method for the estimation of the probability of failure of highly reliable hybrid dynamic systems”. In: *Reliability Engineering & System Safety* 154 (2016), pp. 117–126 (cit. on p. 123).
- [Uri+21] Felipe Uribe, Iason Papaioannou, Youssef M Marzouk, and Daniel Straub. “Cross-entropy-based importance sampling with failure-informed dimension reduction for rare event simulation”. In: *SIAM/ASA Journal on Uncertainty Quantification* 9.2 (2021), pp. 818–847 (cit. on p. 32).
- [Van+17] Paul Vanetti, Alexandre Bouchard-Côté, George Deligiannidis, and Arnaud Doucet. “Piecewise-deterministic markov chain monte carlo”. In: *arXiv preprint arXiv:1707.05296* (2017) (cit. on pp. 47, 55).
- [VG95] Eric Veach and Leonidas J Guibas. “Optimally combining sampling techniques for Monte Carlo rendering”. In: *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. 1995, pp. 419–428 (cit. on pp. 22, 23).
- [Veh+15] Aki Vehtari, Daniel Simpson, Andrew Gelman, Yuling Yao, and Jonah Gabry. “Pareto smoothed importance sampling”. In: *arXiv preprint arXiv:1507.02646* (2015) (cit. on p. 30).
- [VC95] Johan Verbeke and Ronald Cools. “The newton-raphson method”. In: *International Journal of Mathematical Education in Science and Technology* 26.2 (1995), pp. 177–193 (cit. on p. 63).
- [Vil07] José Villén-Altamirano. “Importance functions for RESTART simulation of highly-dependable systems”. In: *Simulation* 83.12 (2007), pp. 821–828 (cit. on pp. 15, 89).
- [Vir+20] Pauli Virtanen et al. “SciPy 1.0: fundamental algorithms for scientific computing in Python”. In: *Nature methods* 17.3 (2020), pp. 261–272 (cit. on p. 105).
- [VRH10] Ulrike Von Luxburg, Agnes Radl, and Matthias Hein. “Hitting and commute times in large graphs are often misleading”. In: *arXiv preprint arXiv:1003.1266* (2010) (cit. on p. 145).
- [WJ94] Matt P Wand and M Chris Jones. *Kernel smoothing*. CRC press, 1994 (cit. on p. 28).

- [WS16] Ziqi Wang and Junho Song. “Cross-entropy-based adaptive importance sampling using von Mises-Fisher mixture for high dimensional reliability analysis”. In: *Structural Safety* 59 (2016), pp. 42–52 (cit. on p. 32).
- [Wes+20] Rex West, Iliyan Georgiev, Adrien Gruson, and Toshiya Hachisuka. “Continuous multiple importance sampling”. In: *ACM Transactions on Graphics (TOG)* 39.4 (2020), pp. 136–1 (cit. on p. 23).
- [Xia+19] Feng Xia, Jiaying Liu, Hansong Nie, Yonghao Fu, Liangtian Wan, and Xiangjie Kong. “Random walks: A review of algorithms and applications”. In: *IEEE Transactions on Emerging Topics in Computational Intelligence* 4.2 (2019), pp. 95–107 (cit. on p. 136).
- [YT23] Renchi Yang and Jing Tang. “Efficient Estimation of Pairwise Effective Resistance”. In: *Proceedings of the ACM on Management of Data* 1.1 (2023), pp. 1–27 (cit. on p. 145).
- [Yen+08] Luh Yen, Marco Saerens, Amin Mantrach, and Masashi Shimbo. “A family of dissimilarity measures between nodes generalizing both the shortest-path and the commute-time distances”. In: *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2008, pp. 785–793 (cit. on p. 137).
- [Yen+05] Luh Yen, Denis Vanvyve, Fabien Wouters, François Fouss, Michel Verleysen, Marco Saerens, et al. “clustering using a random walk based distance measure.” In: *ESANN*. 2005, pp. 317–324 (cit. on p. 137).
- [YJ12] Gu Yingkui and Li Jing. “Multi-state system reliability: A new and systematic review”. In: *Procedia engineering* 29 (2012), pp. 531–536 (cit. on p. 132).
- [ZHZ24] Chuanyan Zhang, Xiaoguang Hong, and Yongqing Zheng. “Fast computation of General SimRank on heterogeneous information network”. In: *Discover Computing* 27.1 (2024), p. 8 (cit. on p. 136).
- [Zha+23] Liangfu Zhang, Cuiping Li, Chuanwen Luo, and Hong Chen. “All-Pairs SimRank Updates on Dynamic Graphs”. In: *2023 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCLOUD/SocialCom/SustainCom)*. IEEE. 2023, pp. 131–138 (cit. on p. 136).
- [Zha96] Ping Zhang. “Nonparametric importance sampling”. In: *Journal of the American Statistical Association* 91.435 (1996), pp. 1245–1253 (cit. on p. 28).

- [ZWS20] Xufang Zhang, Lei Wang, and John Dalsgaard Sørensen. “AKOIS: an adaptive Kriging oriented importance sampling method for structural system reliability analysis”. In: *Structural Safety* 82 (2020), p. 101876 (cit. on p. 28).
- [ZO99] Yan-Gang Zhao and Tetsuro Ono. “A general procedure for first/second-order reliability method (FORM/SORM)”. In: *Structural safety* 21.2 (1999), pp. 95–112 (cit. on p. 12).
- [Zhu+23] Mingzhe Zhu, Liwang Zhu, Huan Li, Wei Li, and Zhongzhi Zhang. “Resistance Distances in Directed Graphs: Definitions, Properties, and Applications”. In: *arXiv preprint arXiv:2302.04114* (2023) (cit. on p. 137).
- [Zou23] Qinmeng Zou. “GMRES algorithms over 35 years”. In: *Applied Mathematics and Computation* 445 (2023), p. 127869 (cit. on p. 146).

English summary

Disclaimer

This is the English translation of the [French summary of the foreword part](#).

Abstract

The purpose of this thesis is to provide new methods for estimating rare event probabilities based on the simulation of *Piecewise Deterministic Markov Processes* (PDMPs). This very general class of stochastic processes offers the flexibility needed to accurately represent complex dynamic industrial systems. In particular, it allows for the joint modeling of the deterministic and continuous dynamics of the physical variables of the system (temperature, pressure, liquid levels, etc.), and the random jump dynamics that govern the change in status of its components (failures, repairs, control mechanisms, etc.).

The industrial challenge is to enable the tool PyCATSHOO, used by the company Électricité de France for its probabilistic safety assessment studies, to efficiently estimate the failure probability of such systems with guaranteed accuracy. A classical Monte Carlo approach requires, for a fixed level of accuracy, a number of simulations inversely proportional to the probability sought. It is therefore not suitable for highly reliable systems with high simulation costs. *Importance sampling* is a popular variance reduction method in the rare event context. It consists of generating simulations under a biased distribution that favors the occurrence of the event, and correcting the bias a posteriori. Recent work has proposed a theoretical framework for implementing importance sampling of PDMPs, and has highlighted the connection between the optimal biased distribution and the so-called *committor function* of the process.

Using tools from reliability analysis and the theory of random walks on graphs, new families of approximations of the committor function are introduced in this thesis. The proposed methodology is adaptive: an approximation of the committor function is constructed a priori and then refined during the simulations of a cross-entropy procedure. The simulations are then recycled to produce an importance sampling estimator of the target probability. Convergence results have been obtained, making it possible to overcome the dependence between simulations and construct asymptotic confidence intervals. This method produces excellent results in practice on the tested industrial systems.

About the author

My name is Guillaume Chennetier, and I was born in 1995 in Nice, France. I stayed in the Nice area for my schooling until high school graduation and then pursued higher education in Toulouse with a bachelor's degree in economics and mathematics¹. In the third year of my bachelor's program, I shifted to applied mathematics, still in Toulouse². I then moved to Paris to enroll in the Master 1 program *Mathematics and Applications* at Sorbonne University. I specialized in stochastic mathematics and joined the Master 2 program in the "Statistics" track for the academic year 2019–2020. I completed my final-year internship within the R&D division of the company EDF (Électricité de France), working on the same subject and under the same supervision that would guide my subsequent doctoral thesis.

Thesis context

This thesis was conducted as part of an industrial collaboration under the CIFRE³ program. Although my work time was fully dedicated to the thesis, it was shared between an academic laboratory and a company.

Industrial environment

EDF (Électricité de France) is the leading electricity producer and supplier in France and Europe. It operates three R&D centers in France: Les Renardières, Chatou, and Paris-Saclay. My internship and doctoral work took place at the EDF Lab Paris-Saclay center within the PERICLES department (PERformance and prevention of Industrial risks of the parC through simuLation and studIEs).



I was part of the i2i group, specializing in probabilistic safety and availability studies of systems. This team comprises around fifteen EDF staff members, complemented each year by a few doctoral students and interns. My supervisors were Hassane CHRAIBI⁴, a research engineer and project manager, and Anne DUTFOY⁵, a senior researcher.

¹Former bachelor *MASS*, now *MIASHS*, at "Université Toulouse I Capitole".

²Bachelor *MAPI*³ at "Université Toulouse III Paul Sabatier".

³Conventions Industrielles de Formation par la REcherche.

⁴hassane.chraibi@edf.fr

⁵anne.dutfoy@edf.fr

Academic environment

My thesis was supervised by [Josselin GARNIER](#), professor at l'École polytechnique (member institution of *Institut polytechnique de Paris*).

I was a doctoral student at the [CMAP](#) laboratory (Centre de mathématiques appliquées) at École polytechnique, a member of the [CMAP-SIMPAS](#) statistics team, and of the [INRIA-ASCII](#) probabilities team. My doctoral school was I also was [EDMH](#) (École Doctorale de Mathématiques Hadamard).



Beyond my academic affiliations, my thesis work aligned with the research themes of the Uncertainty Quantification Thematic Network [RT-UQ](#) (formerly [MASCOT-NUM](#)). I greatly benefited from opportunities to engage with and share my work within this community. Finally, I have been serving as secretary of the [Young Statisticians group](#) of the [Société Française de Statistique](#) (SFdS, French Statistical Society) since June 2023.

Motivations

This thesis is primarily driven by an industrial problem situated at the intersection of two fields: reliability analysis and rare event simulation.

Reliability of industrial systems

In 2023, 77.7% of the electricity produced by the EDF Group worldwide came from nuclear sources; and 9.1% from hydropower. The critical failure of a power plant or a dam would have major consequences, so EDF is obligated to ensure the safety of its nuclear and hydropower facilities.

Probabilistic Safety Assessment

The [i21](#) group specializes in probabilistic safety assessments (PSA). According to [Fundamental Safety Rule](#) (RFS n° 2002-01) drafted by EDF, the [IRSN](#)⁶ and the [ASN](#)⁷:

⁶Institut de radioprotection et de sûreté nucléaire (Institute for Radiological Protection and Nuclear Safety)

⁷Autorité de sûreté nucléaire (French Nuclear Safety Authority)

“Les EPS sont une méthode d'évaluation des risques fondée sur une investigation systématique des scénarios accidentels. Elles se composent d'un ensemble d'analyses techniques permettant d'apprécier les fréquences d'événements redoutés et leurs conséquences. Elles permettent d'obtenir une appréciation globale du niveau de sûreté, intégrant aussi bien la fiabilité des équipements que le comportement des opérateurs.”

That can be translated by:

“PSAs are a risk evaluation method based on a systematic investigation of accidental scenarios. They consist of a series of technical analyses to evaluate the frequencies of adverse events and their consequences. They provide an overall assessment of the safety level, integrating both equipment reliability and operator behavior.”

System safety is managed according to four aspects, commonly summarized by the acronym RAMS (Reliability, Availability, Maintainability, Safety):

- (i) **Reliability**: the ability of a device to perform a required function under specified conditions and for a defined period.
- (ii) **Maintainability**: under given usage conditions, the ability of a device to be maintained or restored to a state in which it can perform its required function, using prescribed maintenance methods.
- (iii) **Availability**: the ability of a system to be operational and perform its function under given conditions at a specific time.
- (iv) **Safety**: the ability of a system to avoid causing critical or catastrophic events under given conditions and for a defined duration.

This thesis focuses on the reliability of multi-component industrial systems involved in the operation of hydropower and nuclear plants. In reliability analysis, a multi-component system refers to a complex system composed of multiple interconnected elements or subsystems, called components. Each component can be an equipment piece, a machine, an electronic device, etc. It contributes in some way—often interactively with other components—to the system's overall function. These interactions may involve flows of matter, energy, information, or control between components. For instance, one component may receive inputs from another and produce outputs that are used by other components.

A component can operate in different modes, such as active, inactive, degraded, or failed. These modes are referred to as the component's status. Status changes can occur automatically via control mechanisms or as a result of random events like failures or repairs.

Modeling PDMPs with the PyCATSHOO tool

The reliability of an industrial system refers to its ability to perform its overall function over a given period under specified conditions. A critical failure is declared when this function can no longer be fulfilled. This typically occurs when the system’s continuous physical variables (e.g., temperature, pressure, or liquid level in a tank) exceed critical thresholds. If the system is well-designed, such scenarios can only arise if the system remains in a severely degraded state for an extended period—this means after the failure of essential groups of components and before their repair. In the nuclear and hydropower industries, such systems generally exhibit very high reliability. This is explained by two factors:

- (i) *High redundancy*: the system can be reconfigured using multiple identical components to maintain functionality while awaiting the repair or replacement of defective components.
- (ii) *Dyssymmetric frequencies*: the mean time to failure or incident is usually significantly longer than the mean time to repair.

As a result, critical failures are extremely rare. Their probabilities cannot be practically estimated using real-world data, but these systems lend themselves well to computer simulations. A dynamic representation of the system must model both the trajectories of the continuous physical variables being monitored and the status of each component over time. This is the role of a computational tool developed by EDF called PyCATSHOO [CHS16; Des+21].



FIGURE 1 **Logo of the PyCATSHOO tool.** For more information, visit: <http://www.pycatshoo.org/>.

A key feature of PyCATSHOO is its ability to simulate dynamic industrial systems using Python and C++ APIs, representing them as “Piecewise Deterministic Markov Processes” (PDMPs) [Dav84; DDZ15]. PDMPs form a broad class of stochastic processes, encompassing all Markov processes without a diffusive component. They are referred to as hybrid processes because their state variable consists of a continuous part (called “position”) and a discrete part (called “regime”). The position evolves deterministically, depending on the regime. At random times

(with distributions that continuously depend on the process's state) or at deterministic times (when the position reaches the boundary of the state space), the process undergoes jumps to new random positions and regimes, before resuming a new deterministic trajectory.

Failures and repairs of components are modeled as discrete random events, while the evolution of continuous variables is governed by deterministic differential equations derived from physical laws. Each component has reliability characteristics that may depend continuously on physical variables (e.g., a submerged component might weaken due to excessive pressure or temperature). Conversely, the differential equations governing the evolution of physical variables are parametrized by the components' statuses (e.g., water may flow at different rates depending on whether a valve is stuck open or closed). This behavior is well-suited for PDMP modeling and simulation. However, these simulations can be costly, primarily due to the need to solve differential equations.

Rare event simulation

Our objective is to estimate the probability that a given industrial system experiences a critical failure before a specified time. Mathematically, this corresponds to the probability that a PDMP trajectory, following a given distribution, reaches a target region in its state space before a fixed time.

The issue with the Classical Monte Carlo approach

Given the availability of a PDMP simulator (PyCATSH00), the most natural idea is to rely on a classical Monte Carlo estimation. The principle is straightforward: generate a large number of potential PDMP trajectories and calculate the proportion of these trajectories that lead to the event of interest (system failure). This estimator has, in theory, all the desirable properties: it is unbiased, converges to the target quantity, and its convergence rate—determined by the central limit theorem—is independent of the dimension of the space. However, it is not suitable for estimating the probabilities of rare events. This is a well-known problem. When the target probability is much smaller than the inverse of the number of generated simulations n , two scenarios are likely:

- (i) with high probability, no realization of the event of interest is observed, leading to an incorrect estimate of 0,
- (ii) “by chance”, $k \geq 1$ realizations of the event of interest are observed, resulting in an estimate of k/n , which is much larger than the target probability.

In short, the estimator is always significantly off from the true value in relative terms. This is a variance problem. It is generally accepted that the sample size needs to exceed 100 times the

inverse of the target probability to obtain a reliable estimate (with a coefficient of variation slightly below 0.1).

As stated earlier, the critical failure of the industrial systems of interest is a rare event, with typical failure probabilities ranging between 10^{-5} et 10^{-7} . Using PyCATSHOO to estimate these probabilities with Monte Carlo simulations would require hundreds of millions, if not billions, of runs. Given the complex physics of the systems involved, the simulation cost of a single PDMP trajectory is already high. This is because each jump in the trajectory requires solving new differential equations, often with dedicated industrial codes, to determine deterministic trajectory segments. Estimating the system's failure probability is, therefore, extremely costly.

Work of Thomas Galtier

Alternative methods to Classical Monte Carlo, specifically tailored for estimating rare events probabilities, exist and constitute a rich and dynamic field in applied mathematics ([Buc04; RT+09]). In his PhD thesis, defended in 2019 ([Gal19]), Thomas Galtier studied the application of two major families of methods to PDMPs: importance sampling methods [Chr+19] and then splitting methods based on interacting particle systems [Chr+21]. This thesis builds upon Thomas Galtier's work within the same research team. Remarkable results have been achieved using importance sampling methods on small industrial systems. The principle involves simulating a PDMP trajectory, not under its true distribution, but under a biased distribution that increases the likelihood of the event of interest (system failure), and then correcting the bias using the appropriate likelihood ratio. Thomas Galtier and his co-authors defined:

- ♦ the dominant measure, under which the probability density of a fixed-duration PDMP trajectory can be explicitly formulated—essential for evaluating the likelihood ratios involved in importance sampling,
- ♦ the optimal importance distribution for PDMPs, enabling the construction of a zero-variance estimator for the target quantity.

The optimal importance distribution process is a piecewise deterministic process (PDP) that can be rendered Markovian. It evolves within the same state space and follows the same deterministic trajectories as the original PDMP. The distribution of its jump times and destinations is fully characterized by the so-called “committor function” of the process. The committor function returns the probability that a PDMP trajectory will realize the event of interest before a fixed time, given its current state.

Since this function is obviously unknown in practice, it must be approximated. As previously mentioned, Thomas Galtier and his co-authors proposed an approximation method that delivered promising results for small industrial systems ([Chr+19]) but was much less effective for complex and/or large-scale systems.

The goal of this thesis is to propose a more comprehensive and systematic methodology for approximating the committor function of PDMPs that model complex phenomena. This involves understanding, either a priori or through simulations, the mechanisms and pathways that lead a PDMP trajectory to realize the event of interest.

Structure of the manuscript

The body of this manuscript, apart from this introduction and the preceding sections, is divided into three parts, each consisting of several chapters. Their content is briefly described below.

State of the art

The purpose of this section is to introduce the concepts necessary to formulate the problem addressed in this thesis and to describe the existing methods on which our work is based. For coherence and conciseness, this state-of-the-art does not aim to cover all the notions used to solve the problem. These will be introduced as needed in the chapters of [section 3.3.2](#) dedicated to the contributions.

Prerequisites

This state-of-the-art aims to be accessible to beginner mathematicians. However, the reader is expected to be familiar with the following concepts:

- ◇ Measure theory basics: sigma-algebras, measures, sigma-finite measures, Borel sets, integration with respect to a measure, Radon-Nikodym derivatives, and densities relative to a measure. These concepts are mainly used in [chapter 3](#).
- ◇ Elementary probability theory related to measure theory: expectation, variance, conditioning, classical probability distributions, and their properties (Gaussian, uniform, exponential).
- ◇ Basic mathematical statistics: empirical mean and variance, asymptotic behavior (“law of large numbers” and “central limit theorem”), and confidence intervals.
- ◇ Discrete-time stochastic processes: martingales, Markov chains, and stopping times.

Chapter 1: Monte Carlo simulation and rare events

In probabilistic numerical simulation, a common task is the evaluation of a quantity of interest defined as the expectation of a function under a reference distribution. Depending on the nature and dimension of the space, it is not always possible to approximate this integral using a deterministic method. One then resorts to a Monte Carlo method: the expectation is approximated

by an empirical average of i.i.d. draws by computer. When the reference distribution puts little mass where the function to be integrated is large (in absolute value), the relative variance of the Monte Carlo estimator becomes large. It is then preferable to use so-called variance reduction methods. We will focus in particular on the case of rare event simulation. The objective of this chapter is firstly to recall the basics of random variable simulation and Monte Carlo methods. We then describe two families of variance reduction methods adapted to rare events: importance splitting and importance sampling methods.

Chapter 2: Piecewise deterministic Markov processes

The central mathematical object of this thesis is the PDMP (piecewise deterministic Markov process). The aim of this chapter is to provide the reader with all the notions about PDMPs needed to fully understand the contributions of this thesis. A PDMP is a non-diffusive continuous-time jump stochastic process. Its specificity (which actually makes it a very general process) is its hybrid behavior. On the one hand, it takes values on a space composed of a continuous part and a discrete part. On the other hand, it evolves according to a deterministic dynamics between jumps, and jumps both at random and deterministic times. This behavior is actually quite intuitive for readers familiar with inhomogeneous Poisson processes. We first recall the basics of these processes (which are special cases of PDMPs), before addressing the essential characteristics of PDMPs. Finally, we propose classical approaches for simulating PDMP trajectories.

Chapter 3: Importance sampling of piecewise deterministic Markov processes

This chapter aims to present the fundamentals of importance sampling of PDMPs. This method requires the evaluation of the likelihood of a PDMP trajectory. We first give the definition of a dominant measure on the space of PDMP trajectories with respect to which they admit a probability density function. We then describe the process with optimal importance distribution. This optimal importance process is a piecewise deterministic process (not necessarily Markovian). It preserves the state space and flow of the original PDMP. The jump dynamics of the optimal importance process can be explicitly described. In the rare event estimation case, there is a Markovian reformulation of this optimal process. The expression of its jump intensity and jump kernel is then directly connected to the so-called “committor function” of the process. This function returns the probability that a trajectory realizes the event of interest given its current state. This committor function is not known but can be approximated. At the end of the chapter, we present an implementation of importance sampling based on a simple approximation of the committor function.

Contributions

The three chapters presenting the contributions are a reorganization of two articles written during the thesis:

- ◇ The first article, [Che+24a], corresponds to an earlier version of the methodology described in [chapter 4](#), and introduces proximity scores tailored to the modeling of industrial systems, detailed in [chapter 5](#). This work was published in 2024 in the *SIAM Journal of Uncertainty Quantification*.

Guillaume Chennetier, Hassane Chraïbi, Anne Dutfoy, and Josselin Garnier. “Adaptive importance sampling based on fault tree analysis for piecewise deterministic Markov process”. In: *SIAM/ASA Journal on Uncertainty Quantification* 12.1 (2024), pp. 128–156

- ◇ The second article corresponds to enhancements proposed in the methodology of [chapter 4](#) relative to the first article [Che+24a], and introduces new proximity scores, described in [chapter 6](#), based on hitting times for a random walk on a graph. This article is planned for submission in 2024.

Guillaume Chennetier, Hassane Chraïbi, Anne Dutfoy, and Josselin Garnier. “Graph-informed importance sampling for rare event estimation with piecewise deterministic Markov process”. In: (Submission in 2024)

Chapter 4: Adaptive importance sampling strategy

The optimality conditions of an importance distribution for importance sampling of PDMPs were recalled in [chapter 3](#). We are interested in the case of rare event simulation realized by PDMPs. The optimal importance distribution is then characterized by an optimal jump intensity and optimal jump kernel, which are given by the committor function of the process. This committor function associates to each possible state of a PDMP trajectory the probability that it realizes the event of interest. We first introduce the general form of a family of approximations of the committor function. We then detail an adaptive importance sampling procedure with recycling to jointly determine a good candidate within the family of approximations, and to estimate the probability of interest. We prove the asymptotic optimality and normality of the method under easily verifiable and interpretable conditions. Finally, we discuss implementation choices of the method.

Chapter 5: Importance function for multicomponent systems

An adaptive importance sampling strategy has been proposed in [chapter 4](#) for rare event simulation of PDMPs. The importance distribution is built upon an approximation of the committor

function of the process. This approximation, called importance function, is a parametric transformation of a proximity score β between any regime of the PDMP and the set $\mathcal{V}_{\mathcal{F}}$ of regimes allowing to access the critical region \mathcal{F} . In this chapter, we consider PDMPs modeling multi-component industrial systems. The rare event of interest is the critical failure of the system. This critical failure can only occur when certain key combinations of components remain in a failed state for a sufficiently long time. Fault tree analysis provides tools to rigorously define and determine these combinations of components. We propose a proximity score based on the notion of minimal path sets, which corresponds to the combinations of components that prevent the system failure. Our adaptive importance sampling method is tested with this proximity score on toy series/parallel systems then on a more complex industrial test case: the spent fuel pool.

Chapter 6: Graph-informed importance function

An adaptive importance sampling strategy has been proposed in [chapter 4](#) for rare event simulation of PDMPs. The importance distribution is built upon an approximation of the committor function of the process. This approximation, called importance function, is a parametric transformation of a proximity score β between any regime of the PDMP and the set $\mathcal{V}_{\mathcal{F}}$ of regimes allowing to access the critical region \mathcal{F} . The proximity score β^{MPS} , proposed in [chapter 5](#), has proven to be very effective when the PDMP models a coherent multicomponent industrial system. In this chapter, we aim to go further and propose a proximity score that is both more general and more accurate. The evolution of the PDMP regime can be roughly summarized by a random walk $(V_t)_t$ on a graph whose regimes are the vertices. The mean hitting time of the region $\mathcal{V}_{\mathcal{F}}$ for the random walk starting from vertex v is a natural proximity score for v . This mean hitting time is explicit when the random walk is Markovian and time-homogeneous. The random walk $(V_t)_t$ given by the PDMP being typically non-Markovian, we compute these mean hitting times for a simplified random walk $(Y_t)_t$ approximating $(V_t)_t$. Our importance sampling method is implemented with this proximity score on the final test case of the previous chapter: the spent fuel pool.

Conclusion

Chapter 7: Conclusion and perspectives

Finally, the main results of the thesis are summarized and discussed in [chapter 7](#). There are various ways in which the proposed method can be enriched, or its application extended to new test cases. We propose several avenues.

French conclusion

Disclaimer

This is the French translation of the [English conclusion](#).

Conclusion

Le sujet de cette thèse était motivé par des problématiques de fiabilité industrielle rencontrées par l'entreprise EDF (Électricité de France). Nous nous sommes plus précisément intéressés à l'estimation de la probabilité de défaillance de systèmes industriels intervenant dans le fonctionnement de centrales nucléaires et de barrages.

Rappel du problème

Les systèmes industriels qui nous intéressent peuvent être représentés par des processus de Markov déterministes par morceau (PDMPs). Mathématiquement, ce problème d'estimation s'exprime sous la forme :

$$\mathcal{I} = \mathbb{E}_{\mathbf{X} \sim \mathbf{p}}[\mathbf{I}(\mathbf{X})], \quad (14)$$

avec \mathbf{X} une trajectoire de PDMP, \mathbf{p} sa distribution de référence et \mathbf{I} l'indicatrice de l'événement d'intérêt : la trajectoire réalise la défaillance. Les trajectoires de PDMP étant coûteuses à simuler, et la probabilité de défaillance recherchée étant typiquement très faible, on ne peut pas recourir à une méthode de Monte-Carlo classique.

On sait que la distribution optimale pour l'échantillonnage d'importance de PDMPs est entièrement caractérisée par la fonction committor du processus. Cette fonction committor associe à chaque instant t d'une trajectoire de PDMP la probabilité que cette trajectoire réalise l'événement d'intérêt sachant son état à cet instant t . Nous avons fait le pari qu'il était possible d'approcher la fonction committor à un coût raisonnable et qu'une distribution d'importance construite sur une bonne approximation de la fonction committor produirait un estimateur de faible variance.

Synthèse des contributions

L'ensemble des contributions de cette thèse se rapportent au problème que l'on vient de décrire. Elles peuvent être rassemblées dans une seule méthodologie globale d'estimation basée sur

l'approximation de la fonction committor. Cette méthodologie se décompose en deux phases. Dans une première phase “hors-ligne” c'est-à-dire avant toute simulation, on construit une famille d'approximations de la fonction committor. À chaque approximation correspond une distribution d'importance possible. Dans une seconde phase “en-ligne”, on détermine séquentiellement la meilleure distribution d'importance au sein de la famille définie à la phase précédente. Les simulations générées pendant cette seconde phase permettent à la fois de raffiner l'approximation de la fonction committor et de contribuer à l'estimation finale de la quantité d'intérêt.

Approximation a priori de la fonction committor

Nous avons introduit d'abord dans le [chapter 4](#) la forme générale d'une famille paramétrique d'approximations de la fonction committor:

$$\xi_{\theta}(x) = \exp \left[\sum_{j=1}^{d_{\Theta}} \theta_j \times \phi_j(\beta(v)) \right]. \quad (15)$$

avec,

- (i) un vecteur de paramètres $\theta \in \Theta \subset \mathbb{R}^{d_{\Theta}}$ de dimension $d_{\Theta} > 0$,
- (ii) un score de proximité $\beta : \mathcal{V} \mapsto [0, 1]$, où \mathcal{V} est l'ensemble des régimes du PDMP,
- (iii) une famille de fonctions de base $(\phi_j)_{j=1}^{d_{\Theta}}$ avec $\phi_j : [0, 1] \mapsto \mathbb{R}$ pour tout $j \in \llbracket 1, d_{\Theta} \rrbracket$.

Une contribution implicite de cette thèse, est de distinguer clairement:

- ◊ l'information a priori que l'on fournit au modèle, ici représentée par le score de proximité β ,
- ◊ de l'information apprise au cours des simulations, représentée par les valeurs successives du paramètre θ , pour guider la distribution d'importance.

Nous avons proposé deux familles de scores de proximité pouvant être déterminées a priori, sans simulation, et à un coût raisonnable en employant les algorithmes adaptés. Dans le [chapter 5](#), nous avons construit un score de proximité valide pour les systèmes industriels multicomposants cohérents, et basé sur la proportion de chemins minimaux endommagés dans un régime donné. Ces chemins minimaux exploitent la fonction de structure du système et correspondent aux barrières que le processus doit nécessairement traverser pour atteindre la défaillance. Dans le [chapter 6](#), nous avons proposé un score plus général, ne faisant pas d'hypothèse sur ce que modélise le PDMP, et basé sur les temps moyens d'atteinte d'une marche aléatoire sur un graphe représenté par les régimes du PDMP. Ce score peut être calculé explicitement pour une marche

aléatoire markovienne homogène sur ce graphe. Il est donc conditionné au choix de la marche homogène approchant la marche réellement effectuée par le PDMP, mais une marche très simple comme la marche aléatoire uniforme s’est révélée suffisante en pratique.

Stratégie adaptative

Nous avons proposé une méthodologie d’échantillonnage préférentiel adaptative (AIS) basée sur une procédure d’entropie croisée (CE). L’inconvénient des méthodes séquentielles classiques est de n’utiliser qu’une partie des simulations générées pour les étapes d’optimisation ainsi que pour l’étape finale d’estimation. L’avantage de l’échantillonnage d’importance est de pouvoir, en ajustant le rapport de vraisemblance, recycler tous les échantillons passés à tout moment. On peut donc grandement augmenter la précision de la méthode à budget computationnel équivalent.

Des résultats de convergence pour des schémas de recyclage simples existent dans la littérature, mais leurs hypothèses sont difficiles à interpréter et vérifier dans la pratique. Nous avons proposé des hypothèses interprétables et peu contraignantes pour le praticien dans le cas des PDMPs. Notre théorème garantit convergence et normalité asymptotique, ce qui permet de construire des intervalles de confiance asymptotiques pour la quantité d’intérêt.

Diffusion des travaux

Les contributions de cette thèse ont fait l’objet de deux articles, dont les trois chapitres de contributions sont une réorganisation.

- ◇ Le premier article [Che+24a] correspond à une version antérieure de la méthodologie décrite dans le [chapter 4](#), et introduit les scores de proximité adaptés à la modélisation de systèmes industriels décrits dans le [chapter 5](#). Il a été publié en 2024 par le [SIAM Journal of Uncertainty Quantification](#).

Guillaume Chennetier, Hassane Chraïbi, Anne Dutfoy, and Josselin Garnier. “Adaptive importance sampling based on fault tree analysis for piecewise deterministic Markov process”. In: *SIAM/ASA Journal on Uncertainty Quantification* 12.1 (2024), pp. 128–156

- ◇ Le second article correspond aux ajouts proposés dans la méthodologie du [chapter 4](#) par rapport à celle de l’article [Che+24a], et introduit de nouveaux scores de proximité, décrits dans le [chapter 6](#), basés sur les temps d’atteinte d’une marche aléatoire sur un graphe. Il sera prochainement soumis auprès d’un éditeur.

Guillaume Chennetier, Hassane Chraïbi, Anne Dutfoy, and Josselin Garnier. “Graph-informed importance sampling for rare event estimation with piecewise deterministic Markov process”. In: (Submission in 2024)

Les résultats obtenus au fur et à mesure de la thèse ont également été présentés dans différentes conférences nationales et internationales. Exemples choisis:

- ◇ UNCECOMP (Conference on Uncertainty Quantification in Computational Science and Engineering), édition 2023,
- ◇ MCM (Monte Carlo Methods and applications), édition 2023,
- ◇ SIAM UQ (Conference on Uncertainty Quantification), édition 2024,
- ◇ JDS (Journées de Statistique), éditions 2022, 2023, 2024.

Deux prix ont également été attribués à mes travaux lors des rencontres MASCOT-NUM, éditions 2022 et 2024, organisés par le groupe thématique RT-UQ:

- ◇ MASCOT-NUM édition 2022 : prix du meilleur poster,
- ◇ MASCOT-NUM édition 2024 : prix du meilleur oral.

Discussion

Il est très clair que la nature de l'intégrale que l'on souhaite estimer ne se prête à rien d'autre qu'à de la simulation probabiliste. La géométrie très particulière de l'espace des trajectoires de PDMP le rend difficile à parcourir autrement. Pour autant, l'échantillonnage d'importance est une approche risquée pour des distributions multimodales en grande dimension. Cette thèse a justifié au moins en partie l'usage de cette méthode sur ce problème.

Pertinence de l'approche

Un reproche fait à l'échantillonnage d'importance est son caractère intrusif. Ce n'est pas un problème ici puisque le praticien a de toutes façons la main sur l'intégralité des paramètres de simulation du modèle. Nous avons choisi de poursuivre la méthodologie initiée dans la thèse de Thomas Galtier [Gal19]. Il n'était pas évident :

- ◇ qu'exprimer la famille de distributions d'importance uniquement en fonction d'une approximation de la fonction committor offre la flexibilité nécessaire,
- ◇ que le minimiseur de la divergence de Kullback-Leibler, au sein d'une famille assez contrainte, produise un estimateur de faible variance.

Ce pari s'est révélé payant dans la pratique. La littérature actuelle sur l'échantillonnage d'importance s'appuie fortement sur des méthodes d'apprentissage modernes pour apprendre la distribution d'importance optimale au cours des simulations dans un contexte de grande dimension et multimodalité. Ces méthodes sont adaptées pour l'apprentissage de fonctions très

complexes avec un budget de simulation élevé. Dans notre cas, la distribution cible n'est pas si complexe puisque simplement le produit d'une distribution connue et d'une indicatrice. Le budget est en revanche restreint à moins de 10^4 simulations, ce qui ne permet pas d'optimiser la distribution d'importance au cours des itérations dans un cadre non paramétrique (ou avec un très grand nombre de paramètres).

L'idée clé est la suivante : une famille de distributions faiblement paramétrée est bien assez flexible, pourvu qu'on lui fournisse la bonne information en entrée. Nous avons montré que très peu d'hypothèses sur le PDMP étaient nécessaires en pratique pour fournir cette information a priori.

Limites

Toutes les questions ne sont pas tranchées à l'issue de cette thèse. Premièrement, les poids d'une méthode d'échantillonnage d'importance peuvent toujours exploser et il est très difficile d'évaluer la qualité d'une estimation par échantillonnage d'importance sans la confronter à une méthode de Monte-Carlo massif. Les métriques d'évaluation basées sur la variance ou l'*effective sample size* sont elles-mêmes mal estimées lorsque la méthode se trompe.

Notre méthode est pensée pour le cas où la simulation d'une trajectoire de PDMP représente le seul coût non négligeable de la procédure (en raison de la résolution des équations différentielles intervenant dans le flot). Il faudrait déterminer à quel point ce coût doit être élevé (par rapport au coût de la construction de la famille d'approximations de la fonction committor, et à celui de l'optimisation entre les itérations) pour que la méthode reste compétitive.

Inversement, si la physique joue un rôle majeur dans la dynamique du système, les simulations coûteront cher, mais l'approximation de la fonction committor par une fonction du régime uniquement et pas de la position est moins justifiable. La prise en compte de la position rendrait cependant l'évaluation de densité et toutes les phases d'optimisation bien plus coûteuses car il faudrait réévaluer le flot plusieurs fois pour une même trajectoire.

Perspectives

De nombreuses pistes s'ouvrent naturellement pour prolonger cette thèse, en raffinant les méthodes proposées ou en étendant les cas d'application. Nous pouvons citer quelques exemples.

- ◇ Appliquer notre méthode à de nouveaux cas tests hors fiabilité industrielle. Le score de proximité β^{HT} ne faisant pas d'hypothèse sur la nature de ce que modélise le PDMP, rien n'interdit de nouvelles applications, par exemple en dynamique de population ou en modélisation de réseaux de transport.
- ◇ Comparer les performances de notre méthode d'échantillonnage d'importance à des méthodes de type splitting. Il faudrait de plus vérifier si les fonctions d'importance ξ_{φ} que

nous avons proposées sont aussi pertinentes pour ces méthodes.

- ◊ Développer le cadre théorique de l'échantillonnage d'importance aux "General stochastic hybrid systems" (GSHS) [BL06]. Il s'agit de processus stochastiques généralisant les PDMPs: l'évolution du processus entre les sauts est décrite par des équations différentielles stochastiques et non plus ordinaires. Un problème similaire au nôtre dans le cas des GSHS a été traité à l'aide d'un système de particules en interaction dans [MB23].

Deux autres pistes importantes sont développées ci-dessous.

Raffinements de la méthode AIS

La méthode d'échantillonnage d'importance adaptatif que nous avons proposée dans le [chapter 4](#) est assez simple. Elle combine une procédure de minimisation d'entropie croisée avec un schéma de recyclage des échantillons passés, tous les deux dans leur version la plus classique. Cette simplicité nous a permis de garantir la convergence et la normalité asymptotique de notre algorithme sous des conditions vérifiables et peu contraignantes en pratique. De nombreux raffinements de ces méthodes existent dans la littérature mais compliquent précisément les preuves de convergence. L'une des principales difficultés posées par la procédure d'entropie croisée est de déterminer une bonne distribution d'importance initiale.

- ◊ Si celle-ci est trop proche de la distribution de référence \mathbf{p} , alors il faut générer un grand nombre de trajectoires pour que certaines réalisent l'événement d'intérêt. La première itération de l'algorithme coûte donc cher.
- ◊ Si celle-ci est trop éloignée de \mathbf{p} , alors on contrôle difficilement le risque d'explosion des rapports de vraisemblance.
- ◊ Enfin, si celle-ci ne recouvre qu'une partie du support de la distribution cible (en raison de la forte multi-modalité de cette dernière), on manque des régions de l'espace des trajectoires qui ne seront sans doute jamais observées après plusieurs étapes d'optimisation.

On peut minimiser l'impact de cette distribution initiale \mathbf{p} en faisant varier la distribution cible \mathbf{g}^* . L'idée est de se donner une famille $(\mathbf{g}_\gamma^*)_{\gamma \in \mathbb{R}_+}$ de distributions cibles intermédiaires, paramétrées par un réel γ , interpolant entre la distribution de référence \mathbf{p} et la distribution cible \mathbf{g}^* . On suppose par convention que $\lim_{\gamma \rightarrow 0^+} \mathbf{g}_\gamma^* = \mathbf{g}^*$.

Entropie croisée multiniveaux

De nombreux problèmes d'événements rares s'écrivent sous la forme d'un dépassement de seuil :

$$\mathbf{I} : \mathbf{X} \mapsto \mathbb{1}_{\beta(\mathbf{X}) > s^*}, \quad (16)$$

avec β un score de performance associé à la réalisation \mathbf{X} . Dans le domaine de l'analyse de fiabilité, il est courant de plutôt considérer un événement de la forme $\{\beta \leq 0\}$. Le score de performance β est alors appelé "limit state function". Nous avons déjà évoqué en [section 1.3.2](#) deux variantes de la méthode d'entropie croisée : les procédures de "multilevel Cross entropy" (mCE) et "improved Cross entropy" (iCE). Les distributions cibles intermédiaires sont de la forme $\mathbf{g}_\gamma^* \propto |\mathbf{I}_\gamma| \times \mathbf{p}$ où \mathbf{I}_γ remplace l'indicatrice \mathbf{I} . La méthode mCE utilise des indicatrices de seuils faibles, tandis que la méthode iCE utilise la fonction de répartition Φ d'une loi $\mathcal{N}(0, 1)$ comme approximation de l'indicatrice :

$$\mathbf{I}_\gamma^{\text{mCE}} : \mathbf{x} \mapsto \mathbb{1}_{\beta(\mathbf{x}) > s^* - \gamma}, \quad (17)$$

$$\mathbf{I}_\gamma^{\text{iCE}} : \mathbf{x} \mapsto \Phi\left(\frac{\beta(\mathbf{x}) - s^*}{\gamma}\right). \quad (18)$$

La stratégie de mise à jour du paramètre γ diffère également selon la méthode. À l'itération ℓ (sans recyclage des échantillons passés), on dispose d'un échantillon $\mathbf{X}_1, \dots, \mathbf{X}_{n_\ell} \sim \mathbf{g}_{\beta(\ell)}$. Dans le cas mCE, on sélectionne $\gamma_{\ell+1}^{\text{mCE}}$ tel qu'une proportion $\rho \in (0, 1)$ dépasse le seuil correspondant. Dans le cas iCE, on sélectionne $\gamma_{\ell+1}^{\text{iCE}}$ tel que le coefficient de variation de $\left[\mathbf{I}_\gamma(\mathbf{X}_k) \frac{\mathbf{p}(\mathbf{X}_k)}{\mathbf{g}_{\beta(\ell)}(\mathbf{X}_k)}\right]_{k=1}^{n_\ell}$ soit le plus proche possible d'une valeur cible $\text{CV}_{\text{target}}$:

$$\gamma_{\ell+1}^{\text{mCE}} \in \inf\left\{\gamma \in \mathbb{R}_+ : \sum_{k=1}^n \mathbf{I}_\gamma^{\text{mCE}}(\mathbf{X}_k) > \rho n\right\}. \quad (19)$$

$$\gamma_{\ell+1}^{\text{iCE}} \in \arg \min_{\gamma \in [0, \gamma_{\ell+1}^{\text{iCE}}]} \left(\widehat{\text{CV}}_\gamma - \text{CV}_{\text{target}}\right)^2. \quad (20)$$

Score d'une trajectoire

Revenons à notre cas : $\mathbf{X} = (X_t)_{t \in [0, \tau]}$ est une trajectoire de PDMP de durée τ . Le processus est à valeur dans \mathcal{X} et la trajectoire vit dans \mathbf{T}_τ . On s'intéresse à la probabilité que la trajectoire \mathbf{X} atteigne une région $\mathcal{F} \subset \mathcal{X}$ avant une date s_{\max} . On a :

$$\tau = \inf\{t \geq 0 : X_t \in \mathcal{F}\} \wedge s_{\max}. \quad (21)$$

On note $\mathbf{F} = \{\mathbf{x} \in \mathbf{T}_\tau : \tau < s_{\max}\}$. On souhaite donc estimer $\mathcal{I} = \mathbb{E}_{\mathbf{X} \sim \mathbf{p}}[\mathbf{I}(\mathbf{X})]$ avec :

$$\mathbf{I} : \mathbf{x} \in \mathbf{T}_\tau \mapsto \mathbb{1}_{\mathbf{x} \in \mathbf{F}}. \quad (22)$$

Pour pouvoir appliquer la procédure mCE ou iCE, il faut choisir un score de performance β pertinent pour une trajectoire. On peut s'inspirer d'un choix classique pour les méthodes de splitting appliquées à des processus stochastiques ([[Lou+17](#); [CGR19](#)]). Si on dispose déjà d'une fonction de score β sur l'espace d'états \mathcal{X} , le score total d'une trajectoire sur \mathbf{T} peut être choisi

comme la valeur maximale obtenue par le score β le long de la trajectoire:

$$\beta(\mathbf{X}) = \sup_{t \in [0, \tau)} \beta(X_t). \quad (23)$$

Deux choix naturels s'offrent à nous : le score de proximité déjà noté β dans ce manuscrit, et la fonction d'importance $\xi_{\mathcal{G}}$ à l'itération courante.

Recyclage et intervalles de confiance

L'application de la méthode mCE/iCE dans notre cas n'est pas immédiate. Tout d'abord, l'événement $\{\mathbf{X} \in \mathbf{F}\}$ ne s'exprime pas totalement comme un dépassement de seuil pour la fonction de score β , mais on a $\mathbf{X} \in \mathbf{F} \implies \sup_{t \in [0, \tau)} \beta(X_t) = 1$. Il faut donc passer d'abord de \mathbf{I}_γ à \mathbf{I}_0 , puis de \mathbf{I}_0 à \mathbf{I} . Ensuite, la variation de la distribution cible à chaque itération (lors de la mise à jour de γ) nous fait sortir des conditions du [theorem 8](#). On n'a alors plus accès aux intervalles de confiance asymptotiques construits à partir de l'intégralité des simulations passées.

Une possibilité est de décomposer l'algorithme AIS en deux phases. Pendant une première phase dite de "burn in", la distribution cible varie selon l'un des schémas décrits. Lorsque la distribution cible utilisée correspond à \mathbf{g}^* , on efface toutes les simulations générées mais on conserve la distribution d'importance obtenue à la dernière itération. Dans une seconde phase, on poursuit la méthode AIS avec recyclage mais sans faire varier la distribution cible et sans utiliser les simulations générées en première phase.

D'autres raffinements de la méthode AIS comme l'adaptive multiple importance sampling (AMIS [[MPS19](#)]) sont possibles en phase 1, et ne perturbent pas la convergence et la normalité asymptotique de l'estimateur produit en phase 2.

Bandits : identification du meilleur bras

Le problème de départ était l'estimation de la probabilité de défaillance d'un système industriel pour une distribution donnée, représentée par les taux de panne et de réparations des différents composants. Un usage attendu de notre méthode est de déterminer parmi plusieurs configurations (entendre parmi plusieurs distributions de référence $\mathbf{p}_1, \dots, \mathbf{p}_K$), laquelle est la plus fiable.

$$(QoI) \quad \arg \min_{i \in \llbracket 1, K \rrbracket} \mathbb{E}_{\mathbf{X} \sim \mathbf{p}_i} [\mathbf{I}(\mathbf{X})]. \quad (24)$$

Une approche naïve consisterait à appliquer K fois indépendamment la méthode AIS (une fois pour chaque espérance) et comparer les K estimateurs obtenus. Si les distributions $(\mathbf{p}_1, \dots, \mathbf{p}_K)$ ont support commun, on peut en fait utiliser le même échantillon pour estimer chacune des espérances en changeant simplement de rapport de vraisemblance. Pour tout échantillon i.i.d.

$\mathbf{X}_1, \dots, \mathbf{X}_n$ de distribution \mathbf{g} , et tout $i \in \llbracket 1, K \rrbracket$, on a (sous les bonnes conditions de support) :

$$\hat{\mathcal{I}}_i = \frac{1}{n} \sum_{k=1}^n \mathbf{I}(\mathbf{X}_k) \frac{\mathbf{p}_i(\mathbf{X}_k)}{\mathbf{g}(\mathbf{X}_k)}. \quad (25)$$

Identification du meilleur bras

On a affaire à une variante du problème d'identification du meilleurs bras (BAI pour best arm identification) pour des bandits multibras [AB10; GGL12; GK16]. Dans sa formulation classique :

- ◊ K distributions sont disponibles et sont appelées “bras”,
- ◊ à chaque itération, on peut choisir l'un des bras et générer une observation sous cette distribution,
- ◊ l'objectif est de déterminer la distribution avec la moyenne la plus haute.

La stratégie séquentielle de sélection des bras doit permettre de minimiser le nombre total de bras actionnés à niveau de confiance fixé (“fixed-confidence setting”), ou de maximiser le niveau de confiance de la décision finale à budget de simulation fixé (“fixed-budget setting”).

Dans notre cas, on cherche plutôt le bras avec la moyenne la plus basse. On peut tirer nos observations sous la distribution d'importance de notre choix et obtenir de l'information conjointement sur chacune des K espérances à chaque tirage. Si la distribution d'importance est fixée (cadre non adaptatif), la seule décision à prendre est l'arrêt de la méthode lorsqu'on est assez confiant dans le résultat. Il y a déjà un travail intéressant à faire pour déterminer le niveau de confiance de la décision pour un échantillon de taille n donné.

Avec échantillonnage d'importance adaptatif

Pour obtenir les meilleures performances possibles, on s'attend à devoir mettre à jour la distribution d'importance au fur et à mesure des simulations. Le choix de la distribution cible \mathbf{g}^* est évident dans le cas de l'estimation de la moyenne d'un seul bras. Il ne l'est plus ici, car il n'y a pas de définition évidente de la distribution optimale permettant d'estimer l'espérance d'un vecteur :

$$\mathbb{E}_{\mathbf{X} \sim \mathbf{g}} \left[\frac{\mathbf{I}(\mathbf{X})}{\mathbf{g}(\mathbf{X})} \left(\mathbf{p}_j(\mathbf{X}) \right)_{i=1}^K \right]. \quad (26)$$

On a tout de même le cas particulier du BAI à deux bras : la densité proportionnelle à $|\mathbf{I}| \times |\mathbf{p}_1 - \mathbf{p}_2|$ permet d'estimer exactement la différence entre les deux moyennes en deux tirages. Construire une stratégie d'échantillonnage d'importance adaptatif capable de déterminer le minimiseur d'une famille d'espérances, et de donner le niveau de confiance associé à cette décision, est à notre connaissance un problème encore ouvert.

Titre : Simulation d'événements rares avec des processus de Markov déterministes par morceaux, application en fiabilité industrielle pour l'outil PyCATSHOO.

Mots clés : Processus stochastiques, Échantillonnage d'importance, Fonction committor, Analyse fiabiliste

Résumé : Cette thèse porte sur l'estimation de probabilités d'événements rares pour des processus de Markov déterministes par morceaux (PDMPs). Ces processus offrent la flexibilité nécessaire pour représenter des systèmes industriels dynamiques complexes. Ils peuvent modéliser conjointement la dynamique déterministe et continue des variables physiques du système (température, pression, niveaux de liquide, etc.), et la dynamique de saut aléatoire qui régit le changement de statut de ses composants (pannes, réparations, mécanismes de contrôle, etc.). L'enjeu industriel est de permettre à l'outil PyCATSHOO, utilisé par l'entreprise Électricité de France pour ses études probabilistes de sûreté, d'estimer la probabilité de défaillance de tels systèmes efficacement, et avec une précision garantie. Une approche Monte-Carlo classique réclame, à niveau de précision fixé, une quantité de simulations inversement proportionnelle à la probabilité recherchée. Elle n'est donc pas adaptée au cas de systèmes hautement fiables et dont le coût de simulation est élevé. L'échantillonnage d'importance est une méthode de réduction de variance populaire en situation d'événements rares. Elle

consiste à générer les simulations sous une distribution biaisée favorisant la réalisation de l'événement, et à rectifier le biais a posteriori. Des travaux récents ont proposé un cadre théorique d'implémentation de l'échantillonnage d'importance pour des PDMPs, et mis en évidence le lien existant entre la distribution biaisée optimale et la fonction dite "committor" du processus. À l'aide d'outils issus de l'analyse fiabiliste et de la théorie des marches aléatoires sur graphes, de nouvelles familles d'approximations de la fonction committor sont introduites dans cette thèse. La méthodologie proposée est adaptative : une approximation de la fonction committor est construite a priori puis raffinée au cours des simulations d'une procédure d'entropie croisée. Les simulations sont ensuite recyclées pour produire un estimateur par échantillonnage d'importance de la probabilité cible. Les résultats de convergence obtenus permettent la construction d'intervalles de confiance asymptotiques malgré la dépendance entre les simulations. Cette méthode produit d'excellents résultats en pratique sur les systèmes industriels testés.

Title : Rare event simulation with piecewise deterministic Markov processes, a reliability application for PyCATSHOO tool

Keywords : Stochastic processes, Importance sampling, Committor function, Reliability analysis

Abstract : The purpose of this thesis is to provide new methods for estimating rare event probabilities for Piecewise Deterministic Markov Processes (PDMPs). These processes offer the flexibility needed to represent complex dynamic industrial systems. They can jointly model the deterministic and continuous dynamics of the physical variables of the system (temperature, pressure, liquid levels, etc.), and the random jump dynamics that govern the change in status of its components (failures, repairs, control mechanisms, etc.). The industrial challenge is to enable the tool PyCATSHOO, used by the company Électricité de France for its probabilistic safety assessment studies, to efficiently estimate the failure probability of such system with guaranteed accuracy. A classical Monte Carlo approach requires, for a fixed level of accuracy, a number of simulations inversely proportional to the probability sought. It is therefore not suitable for highly reliable systems with high simulation costs. Importance sampling is a popular variance reduction method in the rare event context. It consists of ge-

nerating simulations under a biased distribution that favors the occurrence of the event, and correcting the bias a posteriori. Recent work has proposed a theoretical framework for implementing importance sampling of PDMPs, and has highlighted the connection between the optimal biased distribution and the so-called "committor function" of the process. Using tools from reliability analysis and the theory of random walks on graphs, new families of approximations of the committor function are introduced in this thesis. The proposed methodology is adaptive: an approximation of the committor function is constructed a priori and then refined during the simulations of a cross-entropy procedure. The simulations are then recycled to produce an importance sampling estimator of the target probability. Convergence results have been obtained, making it possible to overcome the dependence between simulations and construct asymptotic confidence intervals. This method produces excellent results in practice on the tested industrial systems.