

UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

MASTER MATHÉMATIQUES ET APPLICATIONS

TRAVAIL ENCADRÉ DE RECHERCHE

Simulation D'événements Rares : Application de l'algorithme Branchement Multi-niveaux Adaptative

Travail réalisé par : Mohammed Hajji

Proposé par : Jacques Printems

05 Septembre 2019.

Table des matières

1	Introduction	4
2	Cadre Général	6
2.1	Paramètres de base	6
2.2	Simulation de Monte Carlo (MC)	6
2.3	La méthode Échantillonnage Préférentiel	7
2.4	La méthode <i>Branchement Multi-niveaux</i>	7
3	La méthode de Branchement Multi-niveaux Classique	8
3.1	Fonction d'importance	8
3.2	Étapes et seuils	9
3.3	Les Branchements	10
3.4	Départ et arrêt	10
3.5	Analyses	11
3.5.1	Distribution de nombre de succès	11
3.5.2	Estimateur	12
3.5.3	Branchement Fixé Versus Effort Fixé	13
3.5.4	Analyse de la variance pour un paramètre simplifié	14
3.5.5	Efficacité	14
3.5.6	Implémentation	17
4	La version Branchement Multi-niveaux Adaptative	18
4.1	Description de l'algorithme	18
4.2	Convergence de l'estimateur	20
4.3	Complexité et Implémentation	25
4.4	Sur la fonction d'importance	26
4.5	Conclusion	26
5	Application de l'algorithme BMA dans la dynamique moléculaire	27
5.1	Cadre de l'application	27
5.2	Adaptation de l'algorithme à la dynamique moléculaire	28
5.3	Illustration Numérique : MC versus BMA	30
5.4	Cas unidimensionnel	31
5.4.1	Durée du trajectoire réactive	31
5.4.2	Temps de calcul	33
5.4.3	Temps de transition	33
5.5	Cas bidimensionnel	34
5.5.1	Insensibilité au choix de la coordonnée de réaction	35
5.5.2	Temps de transition	36
6	Conclusion	36

1 Introduction

L'estimation de probabilité d'événements rares est une question cruciale dans de nombreux domaines tels que la fiabilité, la télécommunication, la gestion du trafic aérien, ou encore la dynamique moléculaire. Dans les systèmes complexes, l'étude analytique et hors de question et la méthodes de Monte Carlo est très souvent utilisée. Un événement rare est par définition un événement de probabilité très faible (moins de 10^{-9}), mais d'importance pratique cruciale, ce qui justifie une étude précise. La méthode Monte Carlo classique (MC) s'avérant prohibitivement coûteuse dans ce cas, il importe d'appliquer des techniques spécifiques pour leur estimation. Celles-ci se divisent en deux grandes catégories : Échantillonnage Préférentiel d'un côté, Branchement Multi-niveaux de l'autre.

L'échantillonnage préférentiel, EP dans la suite, est apparue dans la littérature en tant qu'outil général puissant pour réduire la variance d'un estimateur, ce qui, dans le cas d'une estimation d'événement rare, signifie également une augmentation de l'occurrence de l'événement rare. Par conséquent, le principe de la méthode EP est de choisir une distribution de sorte que les trajectoires du processus ont plus de chance d'atteindre d'événement rare. L'utilisation d'une distribution biaisée conduira à un estimateur biaisé si nous l'appliquons directement aux simulations. Cependant, les différentes simulations sont pondérées afin de corriger ce biais ; l'estimateur EP est alors sans biais. Le poids qui est donné à chaque simulation est le ratio de vraisemblance, qui est la densité de Radon-Nikodym de la vraie distribution par rapport à la distribution biaisée. Le point fondamental dans l'implémentation d'une simulation utilisant l'EP est le choix de la distribution biaisée. Choisir ou créer une bonne distribution biaisée est l'art des EP. Le choix d'une distribution auxiliaire efficace dans le cas d'un espace d'états continu nécessite que le praticien ait une expérience significative du système étudié et peut faire face à un problème de calcul délicat. L'avantage peut alors être une énorme économie de temps de calculs alors que l'inconvénient pour une mauvaise distribution peut être des calculs plus longs qu'une simple simulation de Monte-Carlo.

La méthode de Branchement Multi-niveaux cependant plus répandue, consiste à séparer le domaine en plusieurs niveaux et nécessite une connaissance suffisante du système pour décider où placer les niveaux. Dans les versions non adaptatives de cette méthode, les niveaux d'importance (ou seuils) sont fixés a priori par le praticien et il est facile de construire des estimateurs non biaisés. Malheureusement, cela n'est pas toujours possible si le système est très complexe. Récemment un algorithme de sélection appelé algorithme de Branchement Multi-niveaux Adaptatif (BMA) a été proposé pour faire face à ce problème : L'estimateur est asymptotiquement consistante, coûte un peu plus que la méthode BM classique et a la même efficacité en termes de variance asymptotique.

Le but de ce mémoire est d'étudier l'estimateur de probabilité d'un événement rare par le méthode Branchement Multi-niveaux dans ses deux versions, classique et adaptative (récemment proposé, *Cérou et Guyader, 2007*) et notamment comment la version adaptative a entrainer un rebond dans le domaine applicative pour estimer les événements rares. Nous allons ensuite exploiter la méthode BM dans sa version adaptative pour estimer un événement rare dans

la dynamique moléculaire, plus précisément, pour estimer les trajectoires réactives, à savoir, des trajectoires à l'équilibre laissant un état métastable et se terminant dans un autre. Nous démontrons l'intérêt de la méthode sur deux cas simples : un potentiel unidimensionnel à deux puits et un potentiel bidimensionnel présentant deux canaux pour passer d'un état métastable à un autre. Nous utilisons aussi une extension de l'algorithme BMA permettant de calculer les temps de transition d'une particule d'un état métastable à un autre.

2 Cadre Général

2.1 Paramètres de base

Soit $X := (X_t)_{t \geq 0}$ un processus stochastique de Markov, continue à droite, à valeurs dans l'espace \mathbb{R}^d , et $(\mathcal{F}_t)_{t \geq 0}$ sa filtration naturelle. Soient A et B deux sous-ensembles disjoints de \mathbb{R}^d et $x_0 \in \mathbb{R}^d \setminus B$, l'état initial associé à X . Généralement, x_0 est un voisinage de A . le processus donc exige, $X_0 = x_0$, quitte l'ensemble A si $x_0 \in A$, atteint éventuellement l'ensemble B ou retourne à l'ensemble A . Posons

$$T_A = \inf \{t > 0 : X_t \in A\}$$

le premier temps où X atteint A (ou retourne à A si $x_0 \in A$) et,

$$T_B = \inf \{t > 0 : X_t \in B\}$$

le premier temps où X atteint B . Par les hypothèses précédentes sur X et $(\mathcal{F}_t)_{t \geq 0}$, T_A et T_B sont des temps d'arrêts. Le but est d'estimer $\gamma = \mathbb{P}[T_B < T_A]$, la probabilité que le processus X atteigne B avant A dans le cas particulière où cet événements est très rare, c'est-à-dire γ est très petit (moins de 10^{-9}). Cet événement rare peut alors être exprimé par $R := \{T_B < T_A\}$. Cela peut se produire dans de nombreuses situations pratiques

2.2 Simulation de Monte Carlo (MC)

Pour estimer γ , la méthode de MC standard utilise n copies du processus X indépendantes avec x_0 comme point initial, jusqu'au temps d'arrêt $\tau = \min[T_A, T_B]$, compte combien de fois l'événement R se réalise et divise par n ,

$$\hat{\gamma} = n^{-1} \sum_{i=1}^n \mathbb{1}_R(X^i).$$

est l'estimateur résultant $\hat{\gamma}$. Puisque la variable $\mathbb{1}_R(X^i)$ est i.i.d et suit la loi de Bernoulli avec une probabilité de succès γ , alors, $\mathbf{Var}[\mathbb{1}_R(X^i)] = \gamma(1 - \gamma)$. Supposons que nous voulions l'erreur relative soit au plus r dans notre estimation. Puisque la variance d'un estimateur MC utilisant n échantillons est égale à :

$$\begin{aligned} \mathbb{V}ar(\hat{\gamma}) &= \frac{n \mathbb{V}ar(\mathbb{1}_R)}{n^2} \\ &= \frac{\gamma(1 - \gamma)}{n}, \end{aligned} \tag{1}$$

cela garantit que nous avons une erreur relative (RE) égal à

$$RE(\hat{\gamma}) := \frac{\sqrt{\mathbb{V}ar(\hat{\gamma})}}{\hat{\gamma}} = \frac{\sqrt{1 - \gamma}}{\sqrt{n\gamma}},$$

Si nous Fixons cette RE à une valeur r , nous aurons besoin de simuler au minimum $n^* = r^{-2}(1 - \gamma)/\gamma \approx r^{-2}/\gamma$ échantillons de processus, qui est un nombre énorme quand γ est très petit. La méthode MC devient très coûteuse en terme du temps et très peu fiable (presque inutile). Par exemple, si $\gamma = 10^{-10}$ et

2.3 La méthode Échantillonnage Préférentiel

si nous voulons que le nombre prévu d'occurrences de cet événement soit d'au moins 100, il faut prendre $n = 10^{12}$.

Pour $n < 10^{10}$, il est probable que nous n'observons pas même une seule occurrence de cet événement. Dans ce cas, non seulement l'estimateur de γ prend la valeur 0 mais la variance empirique aussi, ce qui peut être assez trompeur si nous utilisons la moyenne empirique et la variance empirique pour calculer l'intervalle de confiance sur $\hat{\gamma}$.

Le tableau suivant illustre un exemple pratique de l'augmentation exponentielle de temps de calcul pour $r = 0.01$:

γ	n^*	temps de calcul
0.11	$8,45 \cdot 10^4$	6 sec
$1,45 \cdot 10^{-2}$	$6,36 \cdot 10^5$	58 sec
$4,34 \cdot 10^{-5}$	$2,12 \cdot 10^8$	6 heures
$2,19 \cdot 10^{-9}$	$4,61 \cdot 10^{12}$	3.2 années

Supposons que le chemin menant à B est paramétré par un *paramètre de rareté* noté $L \in \mathbb{R}^+$. Nous entendons par là que la probabilité d'événement rare $\gamma(L)$ diminue quand L augmente. Il existe un certain nombre de techniques personnalisées pour contourner l'inefficacité de la méthode MC en proposant des algorithmes efficaces pour réduire le nombre n^* quand L est grand. Ce qui garantit que l'algorithme est toujours utilisable dans ce cas. Nous appelons ces techniques les *méthodes de simulation d'événements rares*. Deux axes principaux se sont développés dans ce domaine. La méthode de l'échantillonnage préférentiel et la méthode de Branchement Multi-niveau.

2.3 La méthode Échantillonnage Préférentiel

La méthode Échantillonnage Préférentiel fonctionne en modifiant le processus stochastique $(X_t)_{t \geq 0}$ lui-même. En changeant le processus (X_t) , nous pouvons obtenir un processus pour lequel l'événement R a une probabilité supérieure, ce qui rend l'estimateur de MC plus efficace. Pour obtenir un estimateur non biaisé, nous devons mesurer les valeurs de l'échantillon et les prendre en compte dans le changement de distribution de processus stochastique que nous simulons. Ce facteur est également appelé *ratio de vraisemblance*, parce que c'est le rapport de vraisemblance des événements obtenus dans l'ancien processus et le nouveau processus.

La méthode est appelée Échantillonnage Préférentiel car l'événement rare devient plus probable et donc les événements importants sont échantillonnés plus fréquemment. Le problème bien connu de la méthode EP est qu'il est difficile d'obtenir la distribution optimale : le choix du nouveau processus à simuler qui minimisera la variance dans l'estimateur.

Un choix non optimal pourrait rendre la performance de l'estimateur EP encore pire que l'estimateur MC.

2.4 La méthode *Branchement Multi-niveaux*

L'autre technique principale de simulation d'événements rares, celle qui fait l'objet de ce mémoire, tente de créer de nombreuses réalisations de l'événement

rare par une technique de branchement/sélection. Nous définissons la simulation d'un chemin (qui représente l'événement rare) comme une réalisation d'une trajectoire du processus stochastique $(X_t)_{t \geq 0}$ dans l'espace d'états \mathbb{R}^d . Cette méthode ne change pas le système lui-même pour générer des trajectoires plus importantes comme la méthode EP, mais décide lors de l'exécution d'une simulation si c'est une trajectoire prometteuse ou non. Une trajectoire de simulation qui semble importante sera branchée plusieurs fois pour aider à générer plus d'échantillons pour simuler le chemin représentant l'événement rare. Ces chemins sont corrélés car ils partagent un historique commun jusqu'à la dernière "branche", mais ils sont plus susceptibles de simuler l'événement rare que les trajectoires ordinaires.

Comment choisir les niveaux? Quand les décisions? combien de branchements? sont des questions traitées avec soin dans la version classique de la méthode BM. Cependant, dans la version adaptative proposée récemment nous n'avons le praticien n'a besoin de répondre à ces questions. En effet les choix seront fait au fur et à mesure que le système évolue par l'algorithme. Nous allons montrer qu'asymptotiquement la méthode semble très efficace pour simuler γ dans le cas où le praticien ne dispose pas d'informations sur le système.

3 La méthode de Branchement Multi-niveaux Classique

L'idée de base de la méthode Branchement Multi-niveaux Classique (BMC) est de considérer l'événement rare comme intersection d'une séquence emboîtée d'événements. La probabilité de l'événement rare est donc le produit de probabilités conditionnelles, dont chacune peut généralement être estimé beaucoup plus précisément que l'événement rare lui-même.

Considérons le processus de Markov continu à droite $X := (X_t)_{t \geq 0}$ d'espace d'états \mathbb{R}^d , et f une fonction mesurable définie de \mathbb{R}^d dans \mathbb{R} . définissons $Z_t = f(X_t)$ pour tout $t \geq 0$ et supposons $Z_0 \geq 0$. Pour tout *seuil* ou *niveau* $L \geq 0$, notons T_L le premier temps que le processus $Z := \{Z_t ; t \geq 0\}$ atteint l'ensemble $[L, \infty[$ et T_A le premier temps que le processus Z atteint l'ensemble $] - \infty, 0]$. Nous supposons aussi que les deux temps d'arrêts T_A et T_L sont bien définies par rapport au processus X .

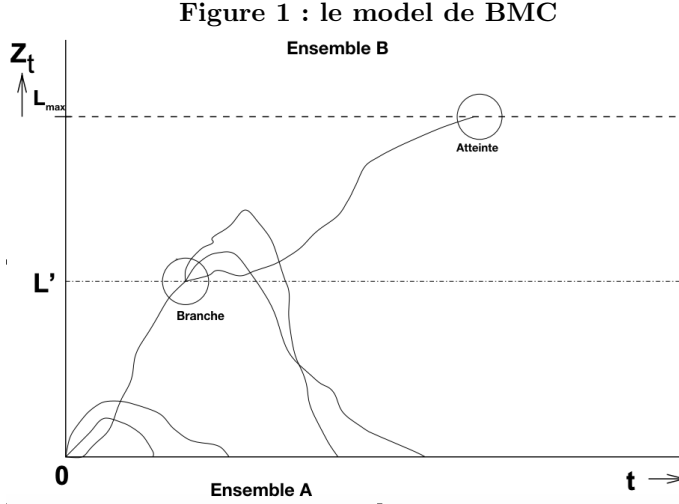
Nous nous intéressons à la probabilité γ de l'événement $D_L := \{T_L < T_A\}$, i.e, la probabilité que Z "franchisse" le niveau supérieur L avant de "franchir" le niveau inférieur 0. Notez que γ dépend de la distribution initiale de X

3.1 Fonction d'importance

Quand le processus $(X_t)_{t \geq 0}$ est unidimensionnel ($X_t \in \mathbb{R}$) nous pouvons mesurer directement les niveaux d'avancement des trajectoires pendant les observations. Cependant, quand le processus est multidimensionnel nous devons projeter le processus au cas unidimensionnel. La méthode BMC est mieux décrite visuellement dans ce cas. Nous imaginons l'espace d'états du problème que nous essayons de résoudre projeté sur un paramètre de valeur réelle que nous appellerons *niveau*. La fonction continue $f : \mathbb{R}^d \rightarrow \mathbb{R}$, faisant cette projection est appelée *fonction d'importance*, et nous donne le *niveau* de l'état actuel X_t

3.2 Étapes et seuils

comme $Z_t = f(X_t)$. L'ensemble représentant l'événement rare est de la forme $B := \{x \in \mathbb{R}^d : f(x) \geq L_{max}\}$ composé des valeurs les plus élevées de f , et $A := \{x \in \mathbb{R}^d : f(x) \leq 0\}$ l'ensemble d'arrêt. Puisque nous avons complètement masqué l'espace d'états \mathbb{R}^d du processus dans f , nous pouvons imaginer la méthode BMC comme des trajectoires dans un plan paramétré avec $Z_t = f(X_t)$ en ordonnées et le temps t en abscisses, comme suit (*Figure 1*).



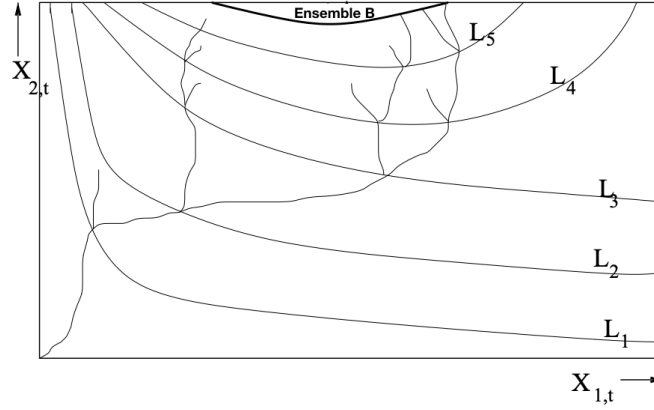
L'évolution d'un chemin de l'échantillon est maintenant très similaire à un modèle de barrage où nous observons le niveau de l'eau sur le barrage, l'événement rare étant l'inondation parce que l'eau montait plus haut que la hauteur du barrage L_{max} .

3.2 Étapes et seuils

Comme il est difficile d'atteindre le plus haut niveau L_{max} en une seule fois, la version BMC consiste à diviser le domaine naturel des variables des niveaux (Z_t) , $[0, L_{max}]$ en m intervalles disjoints, $[L_i, L_{i+1}]$ avec $i \in \{1, \dots, m-1\}$, en choisissant, $0 = L_0 < L_1 < \dots < L_{m-1} < L_m = L_{max}$ comme *seuils*. Notez que nous ne comptons pas L_0 comme *seuil*, mais simplement un niveau d'arrêt. Cela nous permet d'identifier m *seuils* et m *étapes* : l'*étape* i signifie l'évolution de Z_t de $[L_{i-1}, L_i[$ à $[L_i, L_{i+1}[$.

Les m intervalles sont ensuite numérotés par l'indice $i \in \mathbb{N}$ comme suit : $I_i = [L_{i-1}, L_i]$. Dans la figure 1 nous avons choisi $m = 2$ et $L' = L_1$ pour garder la représentation aussi simple que possible. Nous présentons une version avec plusieurs seuils et un espace d'états bidimensionnel $(X_{1,t}, X_{2,t})_{t \geq 0}$ dans la *Figure 2*

Figure 2 : le model de BMC bidimensionnel



Ainsi, la méthode génère toujours un arbre : un chemin se comporte comme un arbre se ramifiant en petites branches.

Le niveaux L_i est appelé seuil puisqu'il régit l'entrée dans la prochaine *étape*. Le nombre de seuils, m , et les seuils L_i eux-mêmes doivent être choisi de telle sorte qu'un trajectoire de l'échantillon franchissant un seuil ait une probabilité notable d'atteindre le seuil suivant. il doit être clair que partir d'une étape à l'étape suivante ne doit pas être un *événement rare*, sinon il n'y a aucun avantage à utiliser la version BMC

3.3 Les Branchements

En atteignant un seuil, disons L_i , l'état du processus doit être sauvegardé et n_i copies identiques à état sauvegardé sont créées, chacune progressant indépendamment après le *branchement*. Ceci est fait pour chaque *seuil* atteint ; i.e, chaque *chemin* qui atteint le plus haut niveau L_{max} aura été branché exactement $m - 1$ fois. Encore une fois, il faut prendre soin de choisir le nombre de *branchements* n_i de manière appropriée ; un nombre trop petit générera un nombre très réduit de trajectoires identiques, et cela entraînera la disparition des trajectoires après un certain nombre d'étapes. Un facteur de branchement n_i trop grand provoquera trop de trajectoires atteignant le prochain seuil et entraînera un nombre de trajectoire par *étape* à effet "boule de neige". Il en résultera un temps de simulation total plus important, et spécialement lors des dernières étapes.

3.4 Départ et arrêt

Nous supposons en outre que nous commençons au niveau $L_0 = 0$ et désignons les états de départ qui génèrent l'échantillon des trajectoires indépendants au cours de la première *étape*. Dans la deuxième étape les trajectoires qui ont la même "racine", i.e, ceux qui ont été générés à partir des états de départ identiques doivent être considérés comme dépendants car ils partagent un historique commun, durant la première étape bien qu'ils progressent indépendamment après le *branchement*, leurs propriétés stochastiques seront influencées par l'état de départ commun. Le critère d'arrêt pour un trajectoire de l'échantillon peut être choisi librement, tant qu'il est indépendant de l'échantillon futur. un exemple

typique; le temps d'arrêt T_A qui marque l'instant du premier retour au *niveau* zéro.

3.5 Analyses

Définissons R_i ; $i = 1, 2, \dots, m$; le nombre d'atteintes au seuil i ($R_0 := 1$), et p_i ; $i = 1, 2, \dots, m$ la probabilité qu'un trajectoire partant d'une branche du seuil inférieure de l'étape i d'atteindre le seuil supérieur avant de s'arrêter. Le nombre de trajectoires initiales partants de l'état initial x_0 à l'instant $t = 0$ est égal à n_0 ; le nombre de *branchements* d'un trajectoire qui réussit à atteindre l'étape i sera n_i , $i = 1, 2, \dots, m$.

Nous introduisons le variable aléatoire indiquant les temps auxquels le processus Z_t atteint les seuils L_i ,

$$T_i = \inf \{t \geq 0, Z_t \geq L_i\}$$

et $T_i := \infty$ quand le seuil i n'est pas atteint avant T_A .

On définit la *distribution d'entrée* comme la distribution d'état du système à l'instant où le seuil L_i est atteint pour la première fois. En d'autres termes, la *distribution d'entrée* à l'étape $i + 1$ est le variable aléatoire $S_i := X_{T_i}$.

Il est évident que la probabilité de succès (atteindre le seuil L_{i+1}) d'une trajectoire à l'étape $i + 1$ dépendra de l'état d'entrée S_i . Notons-le comme une fonction :

$$p_{i+1}(s) := \mathbb{E}[I_{i+1}(s)] \quad , \quad s \in \mathcal{S}(S_i)$$

où $\mathcal{S}(S_i)$ l'espace d'états de S_i , et $I_{i+1}(s)$ un variable de Bernoulli indiquant si une trajectoire partant de l'état s à l'étape $i + 1$, atteint le prochain seuil L_{i+1} . Nous prévoyons évidemment que la probabilité de réussite sera égale à p_{i+1} .

$$\mathbb{E}[p_{i+1}(S_i)] = p_{i+1} \quad (2)$$

3.5.1 Distribution de nombre de succès

A l'étape $i + 1$ nous avons R_i états d'entrée, notons les $S_i^{(1)}, S_i^{(2)}, \dots, S_i^{(R_i)}$. Pour chacun de ces états d'entrée, nous simulons n_i trajectoires et regardons s'ils atteindront le seuil suivant, notons le résultat de ces observations comme vecteur de variables aléatoires de Bernoulli $\{I_{i+1}^{(1)}(S_i^{(1)}), I_{i+1}^{(2)}(S_i^{(1)}) \dots I_{i+1}^{(n_i)}(S_i^{(R_i)})\}$. on obtient le nombre d'atteinte au prochain seuil :

$$R_{i+1} = \sum_{j=1}^{R_i} \sum_{k=1}^{n_i} I_{i+1}^{(k)}(S_i^{(j)}),$$

La somme indiquant le nombre de *réussites* d'un état d'entrée est $G_{i+1}(S_i^{(j)}) := \sum_{k=1}^{n_i} I_{i+1}^{(k)}(S_i^{(j)})$, avec, $j \in 1, \dots, R_i$ peut être vu comme une distribution binomiale de paramètres $(n_i, p_{i+1}(S_i^{(j)}))$. Notons-le $G_{i+1}(S_i^{(j)}) \sim \mathbf{Bin}(n_i, p_{i+1}(S_i^{(j)}))$. La distribution de R_{i+1} devient donc ;

$$R_{i+1} \sim \sum_{k=1}^{R_i} \mathbf{Bin}(n_i, p_{i+1}(S_i^{(k)})) \quad (3)$$

3.5.2 Estimateur

L'estimateur de Monte Carlo pour la probabilité de succès à l'étape i est :

$$\hat{p}_i = \frac{R_i}{n_{i-1}R_{i-1}}, \quad \text{si et seulement si, } R_{i-1} > 0, \quad i \in 1, \dots, m \quad (4)$$

puisque c'est la proportion des échantillons de l'étape i qui ont réussi à atteindre le seuil supérieur. Notons que si $R_i = 0$ pour un certain $0 < i < m$, l'estimateur $\hat{p}_i = 0$, et tous les R_j , $j > i$, ultérieurs seront également égales à zéro, ce qui rend l'estimateur \hat{p}_j indéfini pour $j > i$, puisque il n'y a pas d'observations des étapes $j > i$, l'estimateur p_j donc n'existe pas dans ce cas. Nous supposons dans la suite $R_i > 0$, $i = 1, \dots, m$ car nos estimations seront indéfinies sinon. Notons \mathcal{F}_k la filtration associée au processus stochastique $(R_i)_{i=1}^k$. l'estimateur \hat{p}_i et non biaisé puisque, d'après (2), (3), (4),

$$\begin{aligned} \mathbb{E}[R_{i+1} | \mathcal{F}_i] &= \mathbb{E}[R_{i+1} | R_i] = \mathbb{E}_{S_i} \left[\mathbb{E} \left[R_{i+1} \mid R_i, S_i^{(1)}, S_i^{(2)}, \dots, S_i^{(R_i)} \right] \right] \\ &= \mathbb{E}_{S_i} \left[\mathbb{E} \left[\sum_{k=1}^{R_i} \text{Bin}(n_i, p_{i+1}(S_i^{(k)}) \mid R_i, S_i^{(1)}, S_i^{(2)}, \dots, S_i^{(R_i)} \right] \right] \\ &= \mathbb{E}_{S_i} \left[\sum_{k=1}^{R_i} n_i p_{i+1}(S_i^{(k)}) \right] = n_i p_{i+1} R_i \end{aligned} \quad (5)$$

et en appliquant cela à \hat{p}_i on obtient,

$$\mathbb{E}[\hat{p}_i | \mathcal{F}_{i-1}] = \mathbb{E} \left[\frac{R_i}{n_{i-1}R_{i-1}} \mid R_{i-1} \right] = \frac{\mathbb{E}[R_i | R_{i-1}]}{n_{i-1}R_{i-1}} = \frac{n_{i-1}p_i R_{i-1}}{n_{i-1}R_{i-1}} = p_i \quad (6)$$

en rendant (6) inconditionné à \mathcal{F}_{i-1} ,

$$\mathbb{E}[\hat{p}_i] = \mathbb{E}[\mathbb{E}[\hat{p}_i | \mathcal{F}_{i-1}]] = \mathbb{E}[p_i] = p_i. \quad (7)$$

Donc \hat{p}_i est bien un estimateur non biaisé de p_i .

Soit D_i l'événement " le seuil L_i est atteint pendant la simulation";

$$D_i := \{T_i < T_A\}$$

Notons aussi que $D_0 = \mathbb{R}^d$, puisque $T_0 = 0 < T_A$. Nous remarquons que l'événement rare est $\{T_B < T_A\} = D_m$, et $D_{i+1} \subset D_i$, $i = 0, 1, 2, \dots, m-1$, donc nécessairement pour tout j satisfaisant $m \geq j > i$,

$$\mathbb{P}(D_j \mid D_i) = \frac{\mathbb{P}(D_j \cap D_i)}{\mathbb{P}(D_i)}$$

et comme $j > i$, $D_j \cap D_i = D_j$, donc,

$$\mathbb{P}(D_j \mid D_i) = \frac{\mathbb{P}(D_j)}{\mathbb{P}(D_i)} = \prod_{k=i+1}^j p_k \quad (8)$$

On peut en déduire :

$$\gamma = \mathbb{P}(D_m \mid D_0) = \prod_{i=1}^m p_i. \quad (9)$$

introduisons,

$$\hat{\gamma} = \prod_{i=1}^m \hat{p}_i = \prod_{i=1}^m \frac{R_i}{n_{i-1} R_{i-1}} = \frac{R_m}{\prod_{i=0}^{m-1} n_i} \quad (10)$$

l'estimateur de la probabilité "atteindre le seuil ultime en partant de x_0 ". Pour toute paire d'indices $(i, j) \in \{1, \dots, m\}^2$, $i \neq j$, (supposons $i > j$ sans perte de généralité)

$$\mathbb{E}[\hat{p}_i \hat{p}_j] = \mathbb{E}[\mathbb{E}[\hat{p}_i \hat{p}_j \mid \mathcal{F}_{i-1}]] = \mathbb{E}[\hat{p}_j \mathbb{E}[\hat{p}_i \mid \mathcal{F}_{i-1}]] = \mathbb{E}[\hat{p}_j p_i] = p_i \mathbb{E}[\hat{p}_j] = p_i p_j \quad (11)$$

nous pouvons finalement déduire par récurrence sur (11) que

$$\mathbb{E}[\hat{\gamma}] = \mathbb{E}\left[\prod_{i=1}^m \hat{p}_i\right] = \prod_{i=1}^m p_i$$

et d'après (9),

$$\mathbb{E}[\hat{\gamma}] = \prod_{i=1}^m p_i = \gamma \quad (12)$$

L'estimateur $\hat{\gamma}$ de l'événement rare γ est donc non biaisé.

3.5.3 Branchement Fixé Versus Effort Fixé

Il existe plusieurs options pour instaurer les branchements des trajectoires atteignant le seuil L_i à l'étape $i + 1$. Par exemple, nous pouvons brancher chacune des R_i trajectoires qui atteint le niveau L_i par c_i copies, pour un entier positif fixe c_i . Ainsi, chaque entier $N_i = c_i R_i$ est aléatoire, cette option est appelé *branchement fixé* (*Fixed Splitting*). Si nous voulons que le nombre de branchements de chaque trajectoire soit c_i , où, $c_i = \lfloor c_i \rfloor + \delta$, et $0 \leq \delta < 1$, alors nous supposons que le nombre de *branchements* à l'étape actuelle (étape $i + 1$) est $\lfloor c_i \rfloor + 1$ avec probabilité δ , et $\lfloor c_i \rfloor$ avec probabilité $1 - \delta$.

La seconde option, appelée *effort fixé*, consiste à fixer à priori chaque N_i et simuler à chaque étape la bonne quantité de branchement pour atteindre cette valeur, soit d'une manière aléatoire : choisir aléatoirement les N_i états de départ avec remise parmi les R_i disponibles, ou d'une manière déterministe, en branchant les R_i états d'entrée comme suit : $c_i = \lfloor \frac{N_i}{R_i} \rfloor$ et $d_i = N_i \bmod R_i$, on choisit aléatoirement d_i états d'entrée parmi les R_i disponibles sans remise, et chaque état choisie est branché $c_i + 1$ fois, les états d'entrée qui ne sont pas été choisis seront branchés exactement c_i fois.

Les estimateurs $\hat{\gamma}_1$ et $\hat{\gamma}_2$ de γ par les options *branchement fixé* et *effort fixé* respectivement restent des estimateurs non biaisés selon la sous section précédente, mais elles diffèrent en termes de variance. Garvels et Kroese (1998) ont conclut de leur analyse et des expérience empiriques que l'option *effort fixé* est plus performante que l'option *branchement fixé*, principalement parce qu'elle réduit la variance du nombre de trajectoires simulées à chaque étape. Il se trouve cependant, qu'avec des facteurs de branchement optimaux, cela n'est pas toujours vrai et c'est la méthode *branchement fixé* qui devient meilleur

3.5.4 Analyse de la variance pour un paramètre simplifié

Nous présentons dans la suite une analyse grossière de la variance des deux options *effort fixé* et *branchement fixé* dans le cadre des paramètres idéales (Glasserman et al, 1998)

Commençons par l'option *effort fixé* avec les paramètres idéaux :

$$N_0 = N_1 = \dots = N_{m-1} = n$$

et \hat{p}_i une variable aléatoire indépendant qui suit une loi binomiale de paramètres n et $p = \gamma^{\frac{1}{m}}$. Alors, pour $m > 1$,

$$\mathbf{Var}[\hat{\gamma}] = \mathbf{Var}[\hat{p}_1 \dots \hat{p}_m] = \mathbf{Var} \left[\prod_{i=1}^m \hat{p}_i \right] = \mathbb{E} \left[\prod_{i=1}^m \hat{p}_i^2 \right] - \mathbb{E} \left[\prod_{i=1}^m \hat{p}_i \right]^2,$$

par (11) et (12),

$$\begin{aligned} &= \prod_{i=1}^m \mathbb{E}[\hat{p}_i^2] - \gamma^2 = \prod_{i=1}^m (\mathbf{Var}[\hat{p}_i] + \mathbb{E}[\hat{p}_i]^2) - \gamma \\ &= \left(p^2 + \frac{p(1-p)}{n} \right)^m - p^{2m}, \end{aligned}$$

en utilisant la formule du binôme de Newton,

$$= \frac{mp^{2m-1}(1-p)}{n} + \frac{m(m-1)p^{2m-2}(1-p)^2}{2n^2} + \dots + \frac{(p(1-p))^m}{n^m}.$$

Si nous supposons de plus

$$n \gg \frac{(m-1)(1-p)}{p}, \quad (13)$$

le premier terme $\frac{mp^{2m-1}(1-p)}{n} \simeq \frac{m\gamma^{2-\frac{1}{m}}}{n}$ domine la dernière expression. la variance de l'estimateur de MC (1) est égale à $\frac{\gamma(1-\gamma)}{n} \simeq \frac{\gamma}{n}$. Pour illustrer l'énorme réduction de variance, supposons $\gamma = 10^{-20}$, $m = 20$, $p = \frac{1}{10}$ et $n = 1000$. alors la variance de MC est égale à 10^{-23} , tandis que $\frac{mp^{2m-1}(1-p)}{n} \simeq 1.8 \times 10^{-41}$. Cette simplification n'est pas très précise, car généralement les \hat{p}_i ne sont pas indépendants, et il est difficile d'obtenir $p_i = \gamma^{\frac{1}{m}}$ pour tout i , mais cela donne une idée de l'ordre de grandeur du potentiel réduction de variance.

3.5.5 Efficacité

Un estimateur non biaisé $\hat{\gamma}$ de γ est *d'erreur relative* bornée si $\lim_{\gamma \rightarrow 0^+} \mathbf{RE}[\hat{\gamma}] < \infty$, donc :

$$\begin{aligned} \lim_{\gamma \rightarrow 0^+} \frac{\sqrt{\mathbf{Var}[\hat{\gamma}]}}{\gamma} &< \infty \\ \lim_{\gamma \rightarrow 0^+} \frac{\sqrt{\mathbb{E}[\hat{\gamma}^2] - \gamma^2}}{\gamma} &< \infty \end{aligned}$$

ce qui implique,

$$\lim_{\gamma \rightarrow 0^+} \mathbb{E}[\hat{\gamma}^2] - \gamma^2 = 0$$

3.5 Analyses

donc,

$$\lim_{\gamma \rightarrow 0^+} \log(\mathbb{E}[\hat{\gamma}^2]) = 2 \log(\gamma)$$

d'où

$$\lim_{\gamma \rightarrow 0^+} \frac{\log(\mathbb{E}[\hat{\gamma}^2])}{\log(\gamma)} = 2 \quad (14)$$

Pour analyser le coût de calcul d'un estimateur $\hat{\gamma}$, il est courant de considérer l'*efficacité* de l'estimateur définie comme suit :

$$\mathbf{Eff}[\hat{\gamma}] := \frac{1}{\mathbf{Var}[\hat{\gamma}]K[\hat{\gamma}]},$$

où $K[\hat{\gamma}]$ est le temps nécessaire pour calculer $\hat{\gamma}$. Pour optimiser le coût de calcul il faut donc trouver un estimateur non biaisé, avec une plus grande efficacité que celui disponible auparavant.

On dit qu'un estimateur $\hat{\gamma}$ admet une *efficacité relative* strictement supérieure à zéro si, $\lim_{\gamma \rightarrow 0^+} \gamma^2 \mathbf{Eff}[\hat{\gamma}] > 0$, et *asymptotiquement efficace* si,

$$\lim_{\gamma \rightarrow 0^+} \frac{\log(K[\hat{\gamma}]\mathbb{E}[\hat{\gamma}^2])}{\log(\gamma)} = 2.$$

Une condition suffisante pour que la dernière équation se réalise, est de vérifier (14) et $\frac{\log(K[\hat{\gamma}])}{\log(\gamma)} = 0$, en effet,

$$\lim_{\gamma \rightarrow 0^+} \frac{\log(K[\hat{\gamma}]\mathbb{E}[\hat{\gamma}^2])}{\log(\gamma)} = \frac{\log(K[\hat{\gamma}])}{\log(\gamma)} + \frac{\log(\mathbb{E}[\hat{\gamma}^2])}{\log(\gamma)} = 0 + 2 = 2$$

Le temps CPU pour simuler les trajectoires ou la durée d'utilisation d'une unité centrale pour le traitement du processus à chaque *étape* est proportionnel à n , donc le temps CPU total est proportionnel à $K[\hat{\gamma}] = nm$. La majeure partie de cette procédure consiste à simuler les n trajectoires retournant l'ensemble d'arrêt A à chaque étape. Ainsi, sous l'hypothèse (13), l'efficacité de l'estimateur de la méthode *Branchement Multi-niveaux Classique* est approximativement proportionnel à :

$$\mathbf{Eff}[\hat{\gamma}] = \frac{n}{mp^{2m-1}(1-p)nm} \simeq \frac{\gamma^{-2+\frac{1}{m}}}{m^2}$$

en dérivant le dernier terme par rapport à m on obtient $2m + \log(\gamma)$, ainsi, l'efficacité maximum est atteinte en prenant $m = \frac{-\log(\gamma)}{2}$, et, $p^m = \gamma = e^{-2m}$, ainsi, $p = e^{-2}$. Garvels et Kroes (1998) obtinrent ce résultat. L'erreur relative quadratique dans ce cas est approximativement égale à :

$$\begin{aligned} SRE[\hat{\gamma}] &= RE[\hat{\gamma}]^2 = \frac{\mathbf{Var}[\hat{\gamma}]}{\gamma^2} \\ &\simeq \frac{\gamma^{2-\frac{1}{m}}m\gamma^{-2}}{n} = \frac{e^2m}{n} = \frac{-e^2\log(\gamma)}{2n} \end{aligned}$$

et l'*efficacité relative* est proportionnelle à :

$$\mathbf{REff}[\hat{\gamma}] = \gamma^2 \mathbf{Eff}[\hat{\gamma}] \simeq \frac{\gamma^2 \gamma^{-2+\frac{1}{m}}}{m^2} = (em)^{-2} = \left[\frac{e}{2} \log(\gamma) \right]^{-2},$$

toujours sous l'hypothèse (13).

Considérons maintenant $\gamma \rightarrow 0$ et p fixe, alors $m \rightarrow \infty$, et l'hypothèse (13) n'est donc plus possible. Dans ce cas, l'erreur relative quadratique SRE croît vers l'infini, et l'efficacité relative converge vers 0 à échelle logarithmique. la limite dans (14) dans ce cas est :

$$\begin{aligned} \lim_{\gamma \rightarrow 0^+} \frac{\log(\mathbb{E}[\hat{\gamma}^2])}{\log(\gamma)} &= \lim_{\gamma \rightarrow 0^+} \frac{\log(p^2 + \frac{p(1-p)}{n})^m}{\log(\gamma)} \\ &= \lim_{\gamma \rightarrow 0^+} \frac{\frac{1}{m} \log(p^2 + \frac{p(1-p)}{n})^m}{\frac{1}{m} \log(\gamma)} = \lim_{\gamma \rightarrow 0^+} \frac{\log(p^2 + \frac{p(1-p)}{n})}{\log(\gamma^{\frac{1}{m}})} \\ &= \lim_{\gamma \rightarrow 0^+} \frac{-\log(p^2 + \frac{p(1-p)}{n})}{-\log(p)} < 2. \end{aligned}$$

ainsi, la condition (14) n'est pas vérifiée, cet estimateur n'est pas tout à fait asymptotiquement efficace, mais presque quand n est très grand.

Étudions à présent l'option *branchement fixé* toujours dans le cadre des paramètres idéaux avec $N_0 = n$, $p_i = p = \gamma^{\frac{1}{m}}$, pour tout $i \in \{1, \dots, m-1\}$; et le nombre de *branchements* de chaque trajectoire atteignant le seuil L_i à chaque étape $i+1$ est fixé à $c_i = \frac{1}{p}$ i.e, $N_i = R_i c_i = \frac{R_i}{p}$. Alors, $\{N_i, i \in 1, \dots, m-1\}$; est un processus de *branchement fixé* et l'estimateur $\hat{\gamma}$ devient :

$$\hat{\gamma} = \hat{p}_1 \dots \hat{p}_m = \frac{R_1}{N_0} \frac{R_2}{N_1} \dots \frac{R_m}{N_{m-1}} = \frac{p^{m-1} R_m}{n}$$

donc,

$$\mathbf{Var}[\hat{\gamma}] = \frac{p^{2m-2}}{n^2} \mathbf{Var}[R_m]$$

et d'après la théorie des processus de branchement standard (Harris 1963) :

$$\mathbf{Var}[R_m] = nm(1-p)p$$

Par conséquent

$$\mathbf{Var}[\hat{\gamma}] = \frac{m(1-p)p^{2m-1}}{n} \quad (15)$$

l'erreur relative quadratique associé à cet estimateur :

$$SRE[\hat{\gamma}] = RE[\hat{\sigma}]^2 = \frac{\mathbf{Var}[\hat{\gamma}]}{\gamma^2} = \frac{m(1-p)p^{2m-1}}{np^{2m}} = \frac{m(1-p)}{np}$$

est non bornée dans ce cas aussi, cependant, la limite dans (14) est égale à :

$$\begin{aligned} \lim_{\gamma \rightarrow 0^+} \frac{\log(\mathbb{E}[\hat{\gamma}^2])}{\log(\gamma)} &= \frac{\log\left(\frac{m(1-p)\gamma^2}{np} + \gamma^2\right)}{\log(\gamma)} \\ &= \lim_{\gamma \rightarrow 0^+} \frac{-2m \log p - \log\left(1 + \frac{m(1-p)}{np}\right)}{-m \log p} = \frac{-2m \log p}{-m \log p} = 2 \end{aligned}$$

3.5 Analyses

le temps de simulation associé a cet estimateur est : $K[\hat{\gamma}] = N_0c + N_1c + \dots + N_{m-1}c = c(N_0 + N_1 + \dots + N_{m-1})$ donc,

$$\lim_{\gamma \rightarrow 0^+} \frac{\log(K[\hat{\gamma}])}{\log \gamma} = \lim_{\gamma \rightarrow 0^+} \left(\frac{\log c}{\log \gamma} + \frac{N_0 + N_1 + \dots + N_{m-1}}{\log \gamma} \right) = 0 + 0 = 0$$

la condition suffisante est bien vérifiée, l'estimateur de BMC avec l'option *branchement-fixé* est *asymptotiquement efficace*.

À chaque *étape* i on simule n trajectoires, Le temps de calcul dans l'algorithme BMC est donc d'ordre de grandeur $O(n)$, puisque le nombre des étape est constant et le nombre de branchements est fini.

L'option *branchement fixé* semble asymptotiquement meilleure sous les conditions idéales, cependant cette efficacité est incertaine quand le choix du nombre de branchements doit-être fixé, en effet, un nombre de branchement c très grand causera un très grand nombre de trajectoire à simuler, donc le temps de calcul explose. Inversement un nombre de branchements c très petit entraînera une variance très grande car peu de trajectoires simulées atteindront B . Dans la pratique, étant donné que nombre de branchements optimal c_{opt} est inconnu, l'option *effort-fixé* est communément privilégié.

3.5.6 Implémentation

L'option *effort-fixé* présente un inconvénient qui est son coût élevé en terme de mémoire en comparaison avec l'option *branchement-fixé*. En effet, elle utilise une approche de "largeur" pour son implémentation : à chaque étape $i - 1$ toutes les trajectoires doivent être simulés jusqu'à ce qu'ils atteignent le seuil L_i ou l'ensemble A avant de connaître le facteur de branchement à cette *étape*, ce qui entraîne une mémoire trop élevée quand les N_i sont très grands. Cependant, l'option *branchement-fixé* opte pour une approche de "profondeur" : chaque trajectoire est simulée jusqu'à ce qu'elle atteigne l'ensemble A ou le seuil ultime L_{max} (donc l'ensemble B), ainsi, sa dernière branche (qui a comme état de départ le dernier seuil le plus élevé atteint par la trajectoire simulée) est simulée entièrement, et le procédé suivi la même façon au prochain plus haut seuil, et ainsi de suite. Cette approche récursive nécessite de mémoriser au moins une trajectoire par *étape*, le procédé s'arrête sinon.

Algorithme Branchement Multi-niveaux classique : branchement fixé

Exige : Condition initiale x_0 , fonction d'importance f , les seuils $0 = L_0 < L_1 < \dots < L_m = L_{max}$, les coefficients de branchement n_1, \dots, n_{m-1} , et N_0 l'échantillon initiales ;

Initialisation : $X_0^1, \dots, X_0^{N_0}$ i.i.d à partir de x_0

pour i variant de 1 à $m - 1$

pour j variant de 1 à N_{i-1}

Simuler le trajectoire j jusqu'à le seuil $Z_t^j > L_i$ ou le temps d'arrêt τ

fin pour

Rejeter les trajectoires qui n'ont pas atteints le seuil L_i

Brancher n_i fois les trajectoires qui ont atteints le seuil L_i

Noter N_i le nombre de trajectoires total résultant

Renommer les trajectoires disponibles de 1 à N_i

```

fin pour
pour  $j$  variant de 1 à  $N_{m-1}$ 
    Simuler le trajectoire  $j$  jusqu'à le temps d'arrêt  $\tau$ 
fin pour

```

Estimer la probabilité de l'événement rare par (10) :

$$\hat{p}_{msb} = \frac{1}{N_0} \frac{1}{\prod_{i=1}^{m-1} n_i} \sum_{j=1}^{N_{m-1}} \mathbb{1}_R(X_\tau^j)$$

fin algorithme

4 La version Branchement Multi-niveaux Adaptative

La version adaptative de la méthode BM proposé par Cérou et Guyader, 2007, est largement utilisée dans la pratique : la performance de l'estimateur dans la version classique dépend en effet du choix des seuils L_i , ce choix ne sont pas évidents si aucune information supplémentaire sur le système n'est disponible . Pour rendre l'algorithme de BM encore plus facile et pratique à utiliser, l'idée est de définir les niveaux de manière adaptative au fur et à mesure que l'algorithme évolue : Au lieu de choisir les niveaux à atteindre à priori, on choisit le nombre k de trajectoires à rejeter sur n , ce qui équivaut à choisir $u = n - k$ trajectoires à garder, et chaque niveau correspond à une étape (itération de l'algorithme) et un quantile empirique des niveaux maximaux atteints par l'ensemble de trajectoires. L'algorithme résultant, appelé *algorithme de branchement multi-niveaux adaptative*

Nous gardons les mêmes notations de la partie, précédente : $(X_t)_{t \geq 0}$ un processus de Markov à temps continu, d'espace d'états \mathbb{R}^d , A et B deux sous ensembles disjoints de \mathbb{R}^d et x_0 dans le voisinage de A , f une fonction d'importance telle que $B := \{x \in \mathbb{R}^d : f(x) \geq L_{max}\}$ et $A := \{x \in \mathbb{R}^d : f(x) \leq 0\}$. Soit $Z_t = f(X_t)$ et $z_0 = Z_0 = f(X_0)$. Le but est d'estimer $\gamma = \mathbb{P}[T_B < T_A]$ quand cet événement est rare.

4.1 Description de l'algorithme

La structure de l'algorithme est la suivante : pour atteindre la barrière d'intérêt $\{Z_t \geq L \mid 0 < t < T_A\}$, qui signifie $X_t \in B$, nous procédons en plusieurs étapes. Sauf la première et la dernière, toutes les étapes sont équivalentes. Dans ce qui suit, n désigne le nombre de trajectoires qui sont simulées à l'instant $t = 0$, et u désigne le nombre de trajectoires qui ne sont pas rejeter d'une étape à l'autre.

- *Pas 1* : Simuler n trajectoires $(X_t^j)_{t \geq 0}$ i.i.d suivant la loi du processus $(X_t)_{t \geq 0}$ et de condition initiale commune : $X_0^j = x_0 \forall j \in \{1, \dots, n\}$, donc,

$$Z_0^j = f(X_0^j) = f(x_0) = z_0 \quad \forall j \in \{1, \dots, n\},$$

attendre jusqu'à que toutes les trajectoires finissent dans l'ensemble d'arrêt A , ceci nécessite un temps T_0^j pour chaque trajectoire avec $\mathbb{E}[T_A^j] = \mathbb{E}_{x_0}[T_A] < \infty$.

4.1 Description de l'algorithme

soit :

$$M_{n,j}^1 = \sup_{0 \leq t \leq T_A^j} Z_t^j, \quad j \in \{1, \dots, n\}$$

ordonner $(M_{n,1}^1, M_{n,2}^1, \dots, M_{n,n}^1)$ dans un ordre croissant comme suit,

$$M_{n,\varepsilon(1)}^1 \leq \dots \leq M_{n,\varepsilon(n-u)}^1 \leq \dots \leq M_{n,\varepsilon(n)}^1,$$

où ε est une permutation sur $\{1, \dots, n\}$

garder en mémoire la quantité :

$$\hat{q}_1 = M_{n,(n-u)}^1$$

• *Pas 2* : Maintenir $(M_{n,\varepsilon(n-u+1)}^1, M_{n,\varepsilon(n-u+2)}^1, \dots, M_{n,\varepsilon(n)}^1)$ inchangés et notons les simplement $(M_{n,n-u+1}^2, M_{n,n-u+2}^2, \dots, M_{n,n}^2)$. Rejeter les $(n-u)$ trajectoires auxquelles correspond $(M_{n,\varepsilon(1)}^1, \dots, M_{n,\varepsilon(n-u)}^1)$ et simuler ensuite $(n-u)$ nouvelles trajectoires (branchement) $(X_t^j)_{t \geq 0}$ à partir du point initial $x_0^{(1)}$ tel que $f(x_0^{(1)}) = \hat{q}_1$. Attendre jusqu'à que toutes les $(n-u)$ trajectoires rejoignent l'ensemble A : pour la j -ième trajectoire cela nécessite T_A^j avec $\mathbb{E}[T_A^j] = \mathbb{E}_{x_0^{(1)}}[T_A] < \infty$. Pour chaque $j \in \{1, \dots, n\}$, notons

$$M_{n,j}^2 = \sup_{0 \leq t \leq T_A^j} Z_t^j, \quad j \in \{1, \dots, n\}$$

réordonner $(M_{n,1}^2, M_{n,2}^2, \dots, M_{n,n}^2)$ dans un ordre croissant comme suit :

$$M_{n,\varepsilon(1)}^2 \leq \dots \leq M_{n,\varepsilon(n-u)}^2 \leq \dots \leq M_{n,\varepsilon(n)}^2,$$

où ε est une permutation sur $\{1, \dots, n\}$

garder en mémoire la quantité :

$$\hat{q}_2 = M_{n,(n-u)}^2$$

• répéter la procédure jusqu'à l'itération \hat{N} telle que, $\hat{q}_{\hat{N}+1} \geq L_{max}$.

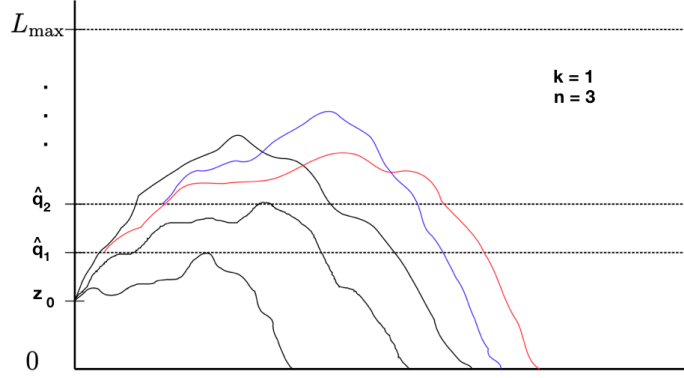
Il existe une proportion $\hat{r} > 0$ parmi l'échantillon $(M_{n,1}^{\hat{N}}, M_{n,2}^{\hat{N}}, \dots, M_{n,n}^{\hat{N}})$ supérieur ou égale à L_{max} , cette proportion correspond aussi aux trajectoires atteignant l'ensemble B par branchement, donc à l'événement rare. Telles trajectoires sont appelées *chemins*.

• Calculer l'estimateur de probabilité $\hat{\gamma}_n$ de γ

Notons $p = \frac{u}{n} = \frac{n-k}{n} = 1 - \frac{k}{n}$, alors

$$\hat{\gamma}_n = \hat{r} p^{\hat{N}} = \hat{r} \left(1 - \frac{k}{n}\right)^{\hat{N}}$$

Figure 3 : le Model BMA



Cette estimateur est non biaisé. On peut le démontrer d'une manière analogue au cas *BMC*, où les seuils L_i sont remplacés par les quantiles empiriques \hat{q}_i d'ordre $\frac{k}{n}$ et le nombre de trajectoires simulées à chaque *étape* i , $n_i R_i$, devient constant et est égal à n .

Remarque : Le nombre k de trajectoires rejetées peut devenir aléatoires si plus d'une trajectoire ont le même maximum à l'étape i . Il existe une astuce qui permet de traiter différemment le cas d'égalité dans l'ordre statistique $M_{n,\varepsilon(1)}^i \leq \dots \leq M_{n,\varepsilon(n)}^i$, ce qui peut entraîner à l'étape i de rejeter un nombre k_i de trajectoires tel que, $k_i > k$. La variante est la suivante :

- Premièrement, imposer un ordre total arbitrairement parmi les scores des trajectoires $M_{n,\varepsilon(1)}^i \leq \dots \leq M_{n,\varepsilon(n)}^i$ en le choisissant uniformément au hasard parmi les scores compatibles.

- Deuxièmement, écarter exactement les k premières trajectoires par rapport à cet ordre, au lieu de $k_i > k$ trajectoires dans le où on ne le précise pas à l'algorithme.

- Troisièmement et dernièrement, les trajectoires restantes à dupliquées à l'étape sont soit branchées avant le premier temps qu'ils entrent dans $\{f \geq \hat{q}_i\}$ soit le premier temps qu'ils entrent dans $\{f > \hat{q}_i\}$ (si ce dernier existe). Notez qu'une règle de décision doit être choisi ici.

il est important de prendre en compte les valeurs multiples des maxima de f le long de l'ensemble des trajectoires. Si aucune attention n'est faite, alors l'estimateur résultant pourrait être biaisé.

4.2 Convergence de l'estimateur

Nous sollicitations à présent montrer que l'estimateur $\hat{\gamma}_n$ est convergent (consistant), i.e, il converge vers la vraie valeur γ pour n assez grand.

Théorème 1. *L'estimateur $\hat{\gamma}_n$ est convergent. Autrement dit,*

$$\hat{\gamma}_n \xrightarrow[n \rightarrow +\infty]{} \gamma \quad \text{presque sûrement.}$$

Avant tout, remarquons grâce à la continuité de la *fonction d'importance* f et la continuité des trajectoires du processus de Markov $(X_t)_{t \geq 0}$ qu'à l'étape l ,

4.2 Convergence de l'estimateur

et pour relever $\hat{q}_l = M_{n,(n-u)}^l$ sachant \hat{q}_{l-1} , les variables aléatoires $(M_{n,j}^l)_{1 \leq j \leq n}$ sont i.i.d et suivent la loi :

$$\begin{aligned} M_{n,j}^l &\sim \mathfrak{D} \left(\sup_{0 \leq t \leq T_A} f(X_t^{x_0}) \mid \sup_{0 \leq t \leq T_A} f(X_t^{x_0}) \geq x_0^{(l-1)} \right) \\ &\sim \mathfrak{D} \left(\sup_{0 \leq t \leq T_A} Z_t^{z_0} \mid \sup_{0 \leq t \leq T_A} Z_t^{z_0} \geq \hat{q}_{l-1} \right) \end{aligned}$$

Avec la convention $\hat{q}_0 = f(x_0) = Z_0 = z_0$, on peut écrire simplement :

$$\begin{aligned} M_{n,j}^l &\sim \mathfrak{D} \left(\sup_{0 \leq t \leq T_A} f(X_t^{x_0^{(l-1)}}) \mid x_0^{(l-1)} \right) \\ &\sim \mathfrak{D} \left(\sup_{0 \leq t \leq T_A} Z_t^{\hat{q}_{l-1}} \mid \hat{q}_{l-1} \right) \end{aligned}$$

Notons pour la suite $M = \sup_{0 \leq t \leq T_A} Z_t^{\hat{q}_0}$ et F sa fonction de répartition que l'on suppose continue. Soit \mathbf{F} la fonction à deux variables (q_1, q_2) définie comme suit :

$$\mathbf{F}(q_1, q_2) = \mathbb{P}(M \leq q_2 \mid M > q_1)$$

Comme F est continue :

$$\mathbf{F}(q_1, q_2) = \frac{\mathbb{P}(q_1 < M \leq q_2)}{\mathbb{P}(M > q_1)} = \frac{F(q_2) - F(q_1)}{1 - F(q_1)}$$

Ainsi, $M_{n,j}^l$ admet $\mathbf{F}(\hat{q}_{l-1}, \cdot)$ comme fonction de répartition puisque, $\mathbf{F}(\hat{q}_{l-1}, \cdot) = \mathbb{P}(M_{n,j}^l \leq \cdot \mid M_{n,j}^l > \hat{q}_{l-1})$. Les variables aléatoires $U_{n,1}^l, \dots, U_{n,n}^l$ définies comme

$$U_{n,j}^l = \mathbf{F}(\hat{q}_{l-1}, M_{n,j}^l), \quad (16)$$

sont identiquement distribuées avec la loi uniforme sur $[0, 1]$ et indépendants pour chaque étape l . En fait, la preuve de la convergence de l'estimateur repose principalement sur cette simple argument.

Rappel 1 : *Un tableau triangulaire de variables aléatoires $(U_{n,j}, 1 \leq j \leq n)_{n>0}$ est de la forme :*

$$\begin{array}{ccccccc} & U_{1,1} & & & & & \\ U_{2,1} & & U_{2,2} & & & & \\ U_{3,1} & & U_{3,2} & & U_{3,3} & & \\ \dots & & \dots & & \dots & & \\ U_{n,1} & & U_{n,2} & & U_{n,3} & \dots & U_{n,n} \\ \dots & & \dots & & \dots & \dots & \dots \end{array}$$

$(U_{n,j}, 1 \leq j \leq n)_{n>0}$ est dit identiquement distribué si les variables aléatoires ont la même loi, de rangées indépendantes si les variables aléatoires dans la même ligne sont mutuellement indépendantes.

Rappel 2 : Inégalité DKW (Dvoretzky-Kiefer-Wolfowitz)

Soit (U_1, \dots, U_n) des variables aléatoires *i.i.d* de fonction de répartition \mathbf{G} et de fonction de répartition empirique \mathbb{G}_n , alors,

$$\mathbb{P}(\|\mathbb{G}_n - \mathbf{G}\|_\infty > \alpha) \leq 2 e^{-2n\alpha^2}, \quad \forall \alpha > 0.$$

Pour démontrer ce théorème, il nous faut deux lemmes simples :

Lemme 1. Soit $(U_{n,j}, 1 \leq j \leq n)_{n>0}$ Un tableau triangulaire de variables aléatoires identiquement distribué de rangées indépendantes suivant la loi uniforme sur $[0, 1]$. soit $u_n = \lfloor np \rfloor$, alors,

$$U_{n,(n-u_n)} \xrightarrow[n \rightarrow +\infty]{p.s.} q = 1 - p$$

Preuve. Notons \mathbb{G}_n la fonction de répartition empirique de $(U_{n,j})_{1 \leq j \leq n}$, et \mathbf{G} la fonction de répartition de la loi uniforme sur $[0, 1]$ que n'est d'autre que l'identité sur $[0, 1]$. Remarquons d'abord, puisque $\mathbf{G}(x) = x \ \forall x \in [0, 1]$, alors, $\|\mathbb{G}_n - \mathbf{G}\|_\infty = \|\mathbb{G}_n^{-1} - \mathbf{G}\|_\infty$, et par la méthode de la transformée inverse on a :

$$U_{n,(n-u_n)} = \mathbb{G}_n^{-1}\left(\frac{n-u_n}{n}\right) = \mathbb{G}_n^{-1}\left(1 - \frac{u_n}{n}\right)$$

comme $q \in [0, 1]$, alors, $\mathbf{G}(q) = q$, donc,

$$\begin{aligned} |U_{n,(n-u_n)} - q| &= |\mathbb{G}_n^{-1}\left(1 - \frac{u_n}{n}\right) - \mathbf{G}(q)| \\ &\leq \|\mathbb{G}_n^{-1} - \mathbf{G}\|_\infty + \left|p - \frac{u_n}{n}\right| \end{aligned} \quad (17)$$

En utilisant l'inégalité DKW,

$$\mathbb{P}\left(\|\mathbb{G}_n - \mathbf{G}\|_\infty > \frac{1}{\log n} - \left|p - \frac{u_n}{n}\right|\right) \leq 2 e^{-2n\left(\frac{1}{\log n} - \left|p - \frac{u_n}{n}\right|\right)^2},$$

avec $\frac{1}{\log n} > \left|p - \frac{u_n}{n}\right|$, ce qui implique en tenant compte du (17),

$$\mathbb{P}\left(|U_{n,(n-u_n)} - q| > \frac{1}{\log n}\right) \leq 2 e^{-2n\left(\frac{1}{\log n} - \left|p - \frac{u_n}{n}\right|\right)^2},$$

comme $-1 \leq np - \lfloor np \rfloor \leq 1 \Leftrightarrow \frac{-1}{n} \leq p - \frac{\lfloor np \rfloor}{n} \leq \frac{1}{n} \Leftrightarrow \left|p - \frac{u_n}{n}\right| \leq \frac{1}{n}$, alors

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(|U_{n,(n-u_n)} - q| > \frac{1}{\log n}\right) = 0$$

donc $U_{n,(n-u_n)} \xrightarrow[n \rightarrow +\infty]{p.s.} q = 1 - p$ □

Lemme 2. Soit $(V_{n,j}, 1 \leq j \leq n)_{n>0}$ Un tableau triangulaire de variables aléatoires de rangées indépendantes et identiquement distribuées. Soit \mathbf{H}_n la fonction de répartition du n -ième ligne, et \mathbb{H}_n sa fonction de répartition empirique alors :

$$\|\mathbb{H}_n - \mathbf{H}_n\|_\infty \xrightarrow[n \rightarrow +\infty]{p.s.} 0$$

4.2 Convergence de l'estimateur

Preuve. C'est une simple application de l'inégalité de *DKW*, en effet,

$$\mathbb{P} \left(\|\mathbb{H}_n - \mathbf{H}_n\|_\infty > \frac{1}{\log n} \right) \leq 2 e^{-2n \left(\frac{1}{\log n} \right)^2},$$

et comme $\frac{1}{\log n} < \frac{1}{n}$, on déduit, $\lim_{n \rightarrow +\infty} \mathbb{P} \left(\|\mathbb{H}_n - \mathbf{H}_n\| > \frac{1}{\log n} \right) = 0 \quad \square$

Preuve du théorème : Tout d'abord, nous remarquons que pour tout l , $\widehat{q}_l = M_{n,n-u}^l$ donc $\mathbf{F}(\widehat{q}_{l-1}, \widehat{q}_l) = \mathbf{F}(\widehat{q}_{l-1}, M_{n,n-u}^l)$ et par (16), $\mathbf{F}(\widehat{q}_{l-1}, M_{n,n-u}^l) = U_{n,n-u}^l$. Une application directe du lemme 1 nous permet de déduire que,

$$\mathbf{F}(\widehat{q}_{l-1}, \widehat{q}_l) \xrightarrow[n \rightarrow +\infty]{p.s} q = 1 - p \quad (18)$$

Nous constatons aussi,

$$\prod_{k=1}^l (1 - \mathbf{F}(\widehat{q}_{k-1}, \widehat{q}_k)) = \prod_{k=1}^l \left(1 - \frac{F(\widehat{q}_k) - F(\widehat{q}_{k-1})}{1 - F(\widehat{q}_{k-1})} \right) = \prod_{k=1}^l \frac{1 - F(\widehat{q}_k)}{1 - F(\widehat{q}_{k-1})}$$

et comme par définition de F , $F(\widehat{q}_0) = 0$ alors,

$$= 1 - F(\widehat{q}_l) = \mathbb{P}(Z_t^{z_0} > \widehat{q}_l | \widehat{q}_l)$$

Nous allons montrer ensuite par récurrence que pour toute étape l :

$$\left| \prod_{k=1}^l (1 - \mathbf{F}(\widehat{q}_{k-1}, \widehat{q}_k)) - p^l \right| \xrightarrow[n \rightarrow +\infty]{p.s} 0 \quad (19)$$

Le cas $l = 1$ est vraie d'après (18). Supposons (19) est vrai pour un certain $l \geq 1$, comme le produit de deux variable aléatoires converge *p.s* vers le produit des limites alors

$$\left| \prod_{k=1}^{l+1} (1 - \mathbf{F}(\widehat{q}_{k-1}, \widehat{q}_k)) - p^{l+1} \right| \xrightarrow[n \rightarrow +\infty]{p.s} 0$$

Nous nous concentrons maintenant sur la dernière étape. Supposons d'abord que $\frac{\log \gamma}{\log p}$ n'est pas un nombre entier. Posons $N = \lfloor \frac{\log \gamma}{\log p} \rfloor$. Ainsi nous avons *p.s* pour n assez grand,

$$\prod_{k=1}^{N+1} [(1 - \mathbf{F}(\widehat{q}_{k-1}, \widehat{q}_k))] < \gamma < \prod_{k=1}^N [(1 - \mathbf{F}(\widehat{q}_{k-1}, \widehat{q}_k))]$$

donc,

$$1 - F(\widehat{q}_{N+1}) < \gamma < 1 - F(\widehat{q}_N)$$

qui est équivalent à

$$\mathbb{P}(M > \widehat{q}_{N+1}) < \mathbb{P}(M > L) < \mathbb{P}(M > \widehat{q}_N)$$

ce qui implique,

$$\widehat{q}_N < L < \widehat{q}_{N+1}$$

4.2 Convergence de l'estimateur

donc *p.s* pour n assez grand, l'algorithme s'arrête après $\hat{N} = N$ itérations. Soit $\mathbb{F}_{n,N}$ la fonction de répartition empirique de $\{M_{n,j}^N, 1 \leq j \leq n\}$, d'après le lemme 2, nous avons *p.s*,

$$|\mathbf{F}(\hat{q}_N, L) - \mathbb{F}_{n,N}(L)| \leq \|\mathbf{F}(\hat{q}_N, \cdot) - \mathbb{F}_{n,N}\|_\infty \xrightarrow[n \rightarrow +\infty]{p.s.} 0$$

D'autre part, en utilisant la définition de N

$$\begin{aligned} \prod_{k=1}^N [(1 - \mathbf{F}(\hat{q}_{k-1}, \hat{q}_k)) \cdot (1 - \mathbf{F}(\hat{q}_{k-1}, L))] &= 1 - F(L) = \mathbb{P}(M \geq L) \\ &= \mathbb{P}\left(\sup_{0 \leq t \leq T_A} Z_t^{\hat{q}_0} \geq L\right) = \gamma \end{aligned}$$

et d'après (19) on a,

$$\lim_{n \rightarrow \infty} \prod_{k=1}^N [(1 - \mathbf{F}(\hat{q}_{k-1}, \hat{q}_k))] = p^N \quad p.s.,$$

donc

$$\lim_{n \rightarrow \infty} [1 - \mathbb{F}_{n,N}(L)] = \frac{\gamma}{p^N} \quad p.s.$$

et par définition de \hat{r}

$$\mathbb{F}_{n,N}(L) = \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{\{M_{n,j}^N \leq L\}} \Leftrightarrow 1 - \mathbb{F}_{n,N}(L) = \hat{r}$$

nous obtenons

$$\lim_{n \rightarrow \infty} \hat{\gamma}_n = \lim_{n \rightarrow \infty} p^N \hat{r} = \lim_{n \rightarrow \infty} p^N [1 - \mathbb{F}_{n,N}(L)] = p^N \frac{\gamma}{p^N} = \gamma$$

Donc l'estimateur est convergent.

Nous considérons maintenant le cas où $N = \frac{\log \gamma}{\log p}$ est un nombre entier. D'après (19) on a encore,

$$\lim_{n \rightarrow \infty} \prod_{k=1}^N [(1 - \mathbf{F}(\hat{q}_{k-1}, \hat{q}_k))] = p^N = \gamma \quad p.s.,$$

Donc pour n assez grand, l'algorithme s'arrête après $\hat{N} = N$ ou $\hat{N} = (N + 1)$ étape. Nous avons donc deux cas à considérer, $\hat{q}_N \geq L$ ou $\hat{q}_N < L$. l'estimateur γ_n peut s'écrire :

$$\gamma_n = p^{N-1} [1 - \mathbb{F}_{n,N}(L)] \mathbb{1}_{\{\hat{q}_N \geq L\}} + p^N [1 - \mathbb{F}_{n,N+1}(L)] \mathbb{1}_{\{\hat{q}_N < L\}}$$

En utilisant le lemme 2 :

$$\lim_{n \rightarrow \infty} [1 - \mathbb{F}_{n,N-1}(L)] = 1 - \mathbf{F}(\hat{q}_{N-1}, L) = p$$

et

$$\lim_{n \rightarrow \infty} [1 - \mathbb{F}_{n,N}(L)] = 1 - \mathbf{F}(\hat{q}_N, L) = 1$$

On obtient finalement

$$\begin{aligned} |\hat{\gamma}_n - \gamma| &\leq |p^{N-1} [1 - \mathbb{F}_{n,N}(L)] \mathbb{1}_{\{\hat{q}_N \geq L\}} \\ &\quad + |p^N [1 - \mathbb{F}_{n,N+1}(L)] \mathbb{1}_{\{\hat{q}_N < L\}}| \end{aligned}$$

où le terme de gauche tend *p.s* vers 0. Ce qui achève la preuve.

4.3 Complexité et Implémentation

Nous avons déjà vu dans la section précédente que l'implémentation l'algorithme BMC coûte $O(n)$. Le coût de simulations dans l'algorithme BMA est un peu plus élevé par rapport à l'algorithme BMC. En effet, au première étape on simule n trajectoires donc le coût de calcul est d'ordre de grandeur $O(n)$, à cela se rajoute le coût de trouver les maxima de ces n trajectoires qui est également d'ordre de grandeur $O(n)$, ensuite il indispensable d'ordonner ces maxima (en utilisant la méthode de dichotomie par exemple) pour garder les u trajectoires de meilleurs scores, le coût de ce tri est un variable aléatoires d'espérance d'ordre de grandeur $O(n \log n)$, et comme le nombre de étapes est fini, le coût total de l'algorithme BMA est d'ordre de grandeur $O(n \log n)$. en terme de variance Cérou et Guyader, 2007, les auteurs initiateurs de l'algorithme BMA ont démontré rigoureusement que,

$$\mathbf{Var}(\hat{\gamma}_n) = \gamma^2 \left(N \frac{1-p}{p} - \frac{1-r}{r} \right), \quad \text{où } r = \gamma p^{-N}.$$

Pour simplifié les calculs supposons, $\gamma = p^{-N}$ alors,

$$\mathbf{Var}(\hat{\gamma}_n) = \frac{N(1-p)p^{2N-1}}{n}$$

qui est similaires à la variance de l'algorithme BMC avec l'option *branchements-fixé* (15) en utilisant les paramètres idéals.

Algorithme Branchement Multi-niveaux Adaptative

Exige : Point initiale x_0 , fonction d'importance f , taille de l'échantillon n , le nombre de trajectoires à rejeter à chaque étape, le niveau ultime L_{max}

```

Initialisation :  $X_0^1, \dots, X_0^n$  i.i.d à partir de  $x_0$ 
 $m \leftarrow 0$  (indexation des étapes)
pour  $i$  variant de 1 à  $n$ 
    Simuler chaque trajectoire jusqu'à son temps d'arrêt  $\tau_i$ 
poser  $M_i \leftarrow \max_{0 \leq s \leq \tau_i} f(X_s)$ 
fin pour
trier les  $M_i$  dans un ordre croissant,  $M_{\varepsilon(1)} \leq \dots \leq M_{\varepsilon(n)}$ 
poser  $L \leftarrow M_{\varepsilon(k)}$  le niveau actuel
tant que  $L < L_{max}$ 
     $m \leftarrow m + 1$ 
    Rejeter toutes les trajectoires  $i$  telles que  $M_i \leq L$ 
    soit  $k_m$  le nombre de trajectoires rejetées ( $k_m \geq k$ )
    définir  $I_m$  comme ensemble d'indices des trajectoires restantes
    pour  $i \in \{1, \dots, n\} \setminus I_m$ 
        tirer aléatoirement un index dans  $I_m$ 
        Brancher la trajectoire correspondante à partir du premier temps où elle entre dans  $\{f > L\}$ 
        à partir de ce temps simuler le trajectoire branché jusqu'à son temps d'arrêt  $\tau_i$ 
    remplacer le trajectoire  $i$  par le nous trajectoires branché
    posez  $M_i \leftarrow \max_{0 \leq s \leq \tau_i} f(X_s^i)$  fin pour

```

4.4 Sur la fonction d'importance

trier les M_i dans un ordre croissant, $M_{\varepsilon(1)} \leq \dots \leq M_{\varepsilon(n)}$
poser $L \leftarrow M_{\varepsilon(k)}$ le niveau actuel
fin tant que

poser $M = m$ le nombre total d'itérations
Estimer la probabilité de l'événement rare par

$$\hat{p}_{msa} = \left\{ \prod_{m=1}^M \frac{n - k_m}{n} \right\} \times \frac{1}{n} \sum_{i=1}^n \mathbb{1}_R(X_{\tau_i}^i)$$

4.4 Sur la fonction d'importance

Une des questions clé de l'efficacité de la méthode multilevel splitting (dans les deux versions adaptative et classique) est le choix de la fonction d'importance f , car elle intervient conjointement avec les seuils L_i dans la détermination des probabilités p_i dans la version classique et mesure le fait que nous se rapprochent de l'événement rare dans la version adaptative. En pratique, un choix judicieux doit être établi par le praticien, néanmoins nous verrons dans le domaine de la dynamique moléculaire, et sous certaines conditions, qu'un choix judicieux pour obtenir des bonnes estimations n'est pas obligatoire. Si le processus de Markov $(X_t)_{t \geq 0}$ est unidimensionnel d'espace d'états $\mathcal{X} \subset \mathbb{R}$, $\gamma(l)$ est croissante par rapport à l , et si $A =]-\infty, 0]$ et $B = [L, +\infty[$ pour un certain L constant, on peut considérer simplement $f(x) = x$. Dans ce cas, le i -ème seuil est atteint lorsque la trajectoire atteint la valeur L_i dans la version classique, et l'état du processus à l'instant t indique le norme de cette état. Dans le cas d'un espace d'états multidimensionnel, le le choix de f est beaucoup plus difficile, l'idée de base est d'utiliser une fonction réelle pour mesurer, même approximativement, à quel point le processus peut être "proche" de l'événement rare R . Idéalement, $f(x) = f^*(x) = q^*(x)$, où :

$$q^*(x) = \mathbb{P}(X_\tau \in R \mid X_0 = x)$$

est appelée la fonction committor.

Il est très difficile d'utiliser f^* dans les situations pratiques car q^* est inconnu et son calcul est intraitable, sinon, le problème d'estimer l'événement rare serait résolu. Souvent dans ce cas, la fonction d'importance utilisée est imprécise, résultant souvent d'heuristiques proposées par les praticiens, par "imprécise" nous voulons dire que la meilleure variance asymptotique pour l'estimateur MSA de γ est atteint en prenant $f = f^*$, qui est malheureusement impossible dans la plupart des situations pratiques.

4.5 Conclusion

Contrairement à l'algorithme BMC qui nécessite une connaissance approfondie du domaine pour pouvoir bien placé les seuils intermédiaires, la version adaptative est capable de simuler un événement rare dont on ne possède pas les informations sur son domaine, puisque l'algorithme utilise les connaissances accumulées au cours de la simulation pour déterminer à la volée quelle trajectoires doit être branchée et à quel moment. l'estimateur étant non biaisé et convergent, algorithme au surplus n'a pas de perte de variance en comparaison

avec la version classique dans son cas idéal, ce qui veut dire, aucun algorithme classique peut faire mieux que l'algorithme adaptative en terme de variance asymptotique. Nous pouvons aussi utiliser une extension de l'algorithme BMA pour estimer le temps nécessaire pour qu'une trajectoire atteigne l'ensemble B , donc le temps nécessaire pour la réalisation de l'événement rare. Ceci est particulièrement important dans le domaine de la dynamique moléculaire.

5 Application de l'algorithme BMA dans la dynamique moléculaire

5.1 Cadre de l'application

La dynamique moléculaire est une technique de simulation numérique sur ordinateur pour étudier les mouvements physiques des atomes et des molécules. Ces particules sont mises à interagir sous certaines conditions (température, énergie potentielle,...) ce qui donne une idée de l'évolution dynamique du système. La méthode a été initialement développée dans le domaine de la physique théorique à la fin des années 50, mais elle est aujourd'hui appliquée principalement à la physique chimique, à la science des matériaux et à la modélisation de biomolécules. Si la modélisation de particules et sous certaines conditions peut être relativement abordable en utilisant une simple méthode de Monte Carlo, un changement de conditions (e.g température) peut entraîner une rareté d'observations et donc l'inefficacité de la méthode MC. Cependant, dans ce cas, l'algorithme BMA devient une alternative très bénéfique et donne des résultats très proches de la méthode MC.

Nous allons adapter cet algorithme pour calculer la probabilité qu'une particule partant d'un état métastable (A un ouvert de \mathbb{R}^d en respectant les précédentes notations) atteigne un autre état métastable (B un autre ouvert de \mathbb{R}^d tel que $A \cap B = \emptyset$) un tel itinéraire est appelée *chemin réactif*. Nous allons voir en particulier qu'en basse température, et grâce à une adaptation spécifique de l'algorithme BMA, la méthode BM donne de bonnes estimations du temps dans le cas unidimensionnel et dans le cas bidimensionnel. Pour ce dernier, l'algorithme BMA semble être fiable même pour un choix très simple de fonction d'importance, appelée *coordonnée de réaction* dans la dynamique moléculaire.

Paramétrage

Chaque particule du système suit un processus stochastique $(X_t)_{t \geq 0}$ solution de l'équation différentielle :

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\beta^{-1}}dW_t \quad (20)$$

appelée équation de Langevin. Où la fonction, $V : \mathbb{R}^d \rightarrow \mathbb{R}$ décrit le potentiel d'interaction de la particule, ∇ est le gradient de sorte que $-\nabla V(X_t)$ est la force appliquée sur la particule à partir du potentiel. $\beta = \frac{1}{k_B T}$ est le bruit blanc, avec, $k_B = 1,38064852 \times 10^{-23} \text{ J.K}^{-1}$, T la température et W un mouvement brownien standard.

Dans l'équation (20) nous considérons que les particules ne subissent à aucune

5.2 Adaptation de l'algorithme à la dynamique moléculaire

accélération, cette modélisation des particules est connu sous le nom "la dynamique de Langevin sur-amortie". La mesure canonique d'équilibre est :

$$d\mu = Z^{-1} e^{-\beta V(x)} dx, \quad (21)$$

où $Z = \int_{\mathbb{R}^d} e^{-\beta V(x)} dx$.

A et B sont deux états métastables, i.e, se sont des états de l'espace \mathbb{R}^d dans lesquelles l'énergie potentiel d'interaction est à son minimum local, par conséquent la force appliquée sur la particule au voisinage d'un état métastable à partir du potentiel est très proche de zéro. À ce temps, la particule subit presque uniquement au bruit associé au mouvement brownien. Dans ce cas, il est très difficile de générer des chemins réactifs (trajectoires allant de A à B). La difficulté vient du fait qu'une particule à l'équilibre reste typiquement très longtemps autour d'un état métastable avant de sauter à un autre.

Comme mentionné dans la première section, dans le cas où $(X_t)_{t \geq 0}$ est unidimensionnelle, on n'a pas besoin de faire intervenir une coordonnée de réaction (elle est tout simplement égale à l'identité). Cependant, dans le cas où $(X_t)_{t \geq 0}$ est multidimensionnelle, le choix d'une coordonnée de réaction h est nécessaire pour évaluer l'évolution de la particule. Elle doit vérifier les hypothèses suivantes :

$$h : \mathbb{R}^d \longrightarrow \mathbb{R}$$

telle que,

$$\begin{aligned} |\nabla h| &\neq 0, \quad A \subset \{x \in \mathbb{R}^d, h(x) < z_{min}\} \text{ et} \\ B &\subset \{x \in \mathbb{R}^d, h(x) > z_{max}\} \end{aligned} \quad (22)$$

où $z_{min} < z_{max}$ deux nombre réels donnés (ont le même rôle que 0 et L_{max} dans les notations précédentes). Notons aussi

$$\Sigma_z = \{x \in \mathbb{R}^d, h(x) = z\}$$

la sous variété des configurations à une valeur z fixe de h .

Pour que l'algorithme donne des résultats fiables, Σ_{min} et Σ_{max} doivent être "proches" respectivement de A et B . "Proche" signifie que la plupart des particules partantes de Σ_{min} (resp. de Σ_{max}) finissent dans A (resp. dans B). Ultérieurement nous allons tester l'algorithme dans le cas unidimensionnel, et également dans le cas bidimensionnel avec simplement la distance euclidienne par rapport à un point référence dans A comme coordonnée de réaction.

5.2 Adaptation de l'algorithme à la dynamique moléculaire

Le but de cette partie est d'estimer la probabilité de produire une trajectoire réactive, et exploiter l'algorithme BMA pour calculer une estimation du temps associé à la transition d'un chemin réactif. L'implémentation de l'algorithme nécessite un traitement particulier pendant la procédure d'initialisation, en effet, il existe une condition initiale x_0^j appartenant à l'ensemble A pour chaque trajectoire $(X_t^j)_{t \geq 0}$ de l'échantillon à l'étape 1, contrairement à l'implémentation précédente de l'algorithme BMA, où toutes les n trajectoires de l'étape 1 ont une

condition initiale x_0 commune, dans cette application il faut choisir de générer les n points initiales $(x_0^j)_{1 \leq j \leq n}$ selon la mesure canonique (21) à condition d'appartenir à A . Nous traitons aussi particulièrement la procédure d'initialisation pour estimer le temps de transition d'un chemin réactif. Le reste des étapes est exactement identique à l'algorithme BMA vu dans la section précédente. Les trajectoires d'équilibre sont ceux obtenus avec les conditions initiales $(x_0^j)_{1 \leq j \leq n}$ réparties selon μ et satisfaisant l'équation (20).

Implémentation de l'algorithme BMA pour la dynamique moléculaire

L'implémentation de l'algorithme se fait en deux parties ; la partie initialisation (étapes de (1) à (3)) et la partie d'itérations (de (4) à (8)) comme suit :

- (1) Générer un ensemble de n conditions initiales $(x_0^j)_{1 \leq j \leq n}$ selon la mesure μ conditionnellement à $x_0^j \in A, \forall j \in \{1, \dots, n\}$
- (2) Générer n particules, avec pour chaque particule j , x_0^j comme condition initiale et satisfaisant l'équation (20). Simuler leurs trajectoires jusqu'à le temps d'arrêt

$$\sigma^j = \inf\{t \geq 0, h(X_t^j) \geq z_{min}\}$$

- (3) Continuer la simulation des trajectoires des particules jusqu'à le temps d'arrêt

$$\tau^j = \inf\{t \geq \sigma, X_t^j \in A \cup B\}$$

à la fin de la procédure d'initialisation nous avons n trajectoire à l'équilibre $(X_t^j)_{1 \leq t \leq \tau^j}$, qui abandonnent A , et ensuite soit retournent à A (très souvent) soit atteignent B (très rarement), conditionnellement au fait, $\sup_{t \geq \sigma} h(X_t^j) > z_{min}$. Notons les trajectoires de ces particules issues de cette étape

$$(X_t^{1,1})_{1 \leq t \leq \tau^1}, \dots, (X_t^{1,n})_{1 \leq t \leq \tau^n}$$

où les $\tau^{1,j} = \tau^j$ les temps d'arrêts associés.

A partir de cette étape on applique ordinairement la version BMA.

- (4) Calculer le maximum pour chaque trajectoire :

$$z^{k,j} = \sup_{0 \leq t \leq \tau^{k,j}} h(X_t^{k,j})$$

- (5) Ordonner les valeurs $(Z^{k,j})_{1 \leq j \leq n}$ dans un ordre croissant

$$z^{k,\varepsilon^k(1)} \leq z^{k,\varepsilon^k(2)} \leq \dots \leq z^{k,\varepsilon^k(n)}$$

où ε^k est une permutation sur $\{1, \dots, n\}$. Pour simplifier les notations posons : $j^k = \varepsilon^k(1) = \operatorname{argmin} z^{k,j}$, le quantile empirique est

$$q^k = z^{k,\varepsilon^k(1)} = \min_{j \in [1, \dots, n]} z^{k,n}$$

- (6) Rejeter la trajectoire qui a la valeur de coordonnée de réaction la plus faible $(X_t^{k,j^k})_{0 \leq t \leq \tau^{k,j^k}}$, et considérer les trajectoires à l'itération suivante comme suit :

5.3 Illustration Numérique : MC versus BMA

- Pour tout $j \neq j^k$, garder la j -ième trajectoire inchangé :
 $\tau^{k+1,j} = \tau^{k,j}$ et $(X_t^{k+1,j})_{0 \leq t \leq \tau^{k+1,j}} = (X_t^{k,j})_{0 \leq t \leq \tau^{k,j}}$;
- Brancher une nouvelle trajectoire dans 3 étape :
 - (i) Choisir aléatoirement (avec une loi uniforme) $i_k \in \{1, \dots, j^k - 1, k^k + 1, \dots, n\}$
 - (ii) Garder $X_t^{k+1,j^k} = X_t^{k,i_k}$ pour tout $t \in [0, \sigma_k]$ où

$$\sigma_k = \inf\{t \geq 0, h(X_t^{k,i_k}) \geq q^k\}$$

- (iii) Générer une particule pour compléter la trajectoire jusqu'à le temps d'arrêt

$$\tau^{k+1,j^k} = \inf\{t \geq \sigma_k, X_t^{k+1,j^k} \in A \cup B\}$$

- (7) Retourner à (4) avec k devient $k+1$ jusqu'à $q^k \geq z_{max}$.

À la fin chaque itération k , nous avons n trajectoires à l'équilibre, qui partent de A et finissent soit en A ou en B . conditionnellement au fait $\sup_{0 \leq t \leq \tau^{k,j}} h(X_t^j) \geq q^k$.

Le nombre total d'itération est définit comme

$$k_{max} = \sup\{k \geq 1, q^k \leq z_{max}\}$$

À la fin de l'algorithme, tous les trajectoires traversent la sous variété $\Sigma_{q_{max}}$. Comme k_{max} est l'index de la dernière itération pour laquelle le quantile empirique q^k est inférieur à z_{max} , et comme $\Sigma_{z_{max}}$ est "très proche" de zéro par hypothèse, la plupart de ces trajectoires finissent dans B

- (8) Retenir que les trajectoire ayant atteint B . Notant r la proportion de telles trajectoires. L'algorithme BMA nous permet de conclure,

$$\hat{\alpha}_n = r \left(1 - \frac{1}{n}\right)^{k_{max}}$$

donne une estimation du probabilité de l'événement rare " observer une trajectoire réactive", puisque on ne rejette qu'une seule trajectoire à chaque étape, $k = 1$

5.3 Illustration Numérique : MC versus BMA

Dans cette partie le but est d'utiliser l'algorithme BMA pour estimer le temps de transition des particules entre deux états métastables A et B d'un système simple, ie, le temps de transition d'un chemin réactif.

Nous allons d'abord valider l'algorithme BMA via des résultats numériques en les comparant avec les résultats numérique de la méthode MC, dans le cas où le processus $(X_t)_{0 \leq t}$ est unidimensionnel (donc le potentiel d'interaction V est unidimensionnel) dans un premier temps, et ensuite dans le cas où le processus $(X_t)_{0 \leq t}$ est bidimensionnel (donc le potentiel d'interaction V est bidimensionnel).

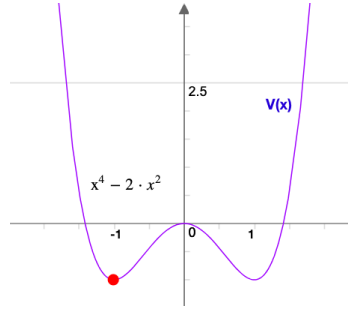
5.4 Cas unidimensionnel

5.4.1 Durée du trajectoire réactive

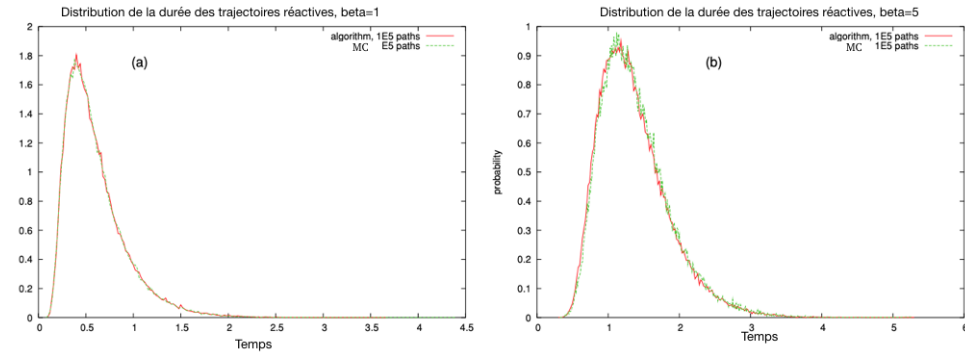
Définissons le potentiel unidimensionnel V comme suit :

$$V(x) = x^4 - 2x^2.$$

qui admet deux minima locaux, -1 et $+1$, correspondants respectivement au deux états métastables A et B . Comme on est dans un cas unidimensionnel il est inutile de choisir une coordonnée de réaction, nous gardons simplement l'identité $h(x) = x$. Le schéma suivant présente le potentiel V et une particule (rouge) à l'état métastable $A = -1$,



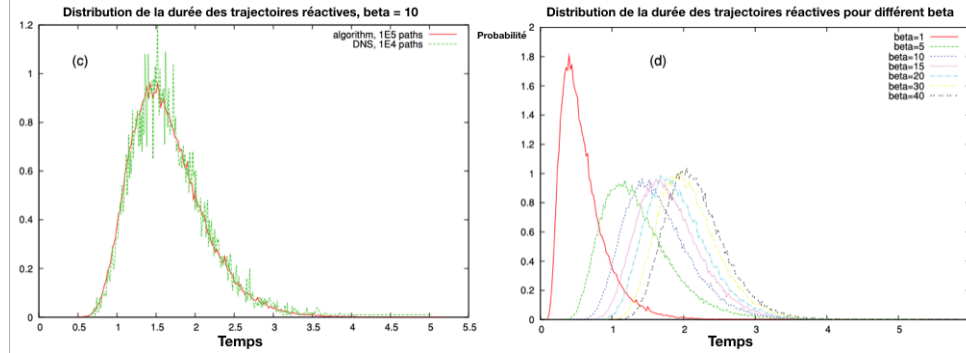
Pour valider l'algorithme, nous calculons d'abord numériquement la courbe de distribution de la durée d'une trajectoire réactive, i.e, la distribution du temps nécessaire pour une particule de se déplacer de z_{min} à $B = \{1\}$ pour les deux méthodes, MC et BMA, pour différentes valeurs de β . les deux courbes suivantes représente la distribution de la durée d'une trajectoire réactives pour $\beta = 1$ et $\beta = 5$,



Nous constatons que pour β petit, ce qui signifie une température élevée, le bruit dans l'équation (20) devient plus important, donc les trajectoires réactives sont moins rares, les deux méthodes MC et BMA donnent des résultats très proches et très précis. On peut en déduire que l'algorithme BMA donne des bonnes estimations même quand l'événement est moins rare et calculable avec une simple méthode de MC.

Pour les valeurs de β élevée, i.e, pour une température basse, nous obtenons les courbes suivantes :

5.4 Cas unidimensionnel

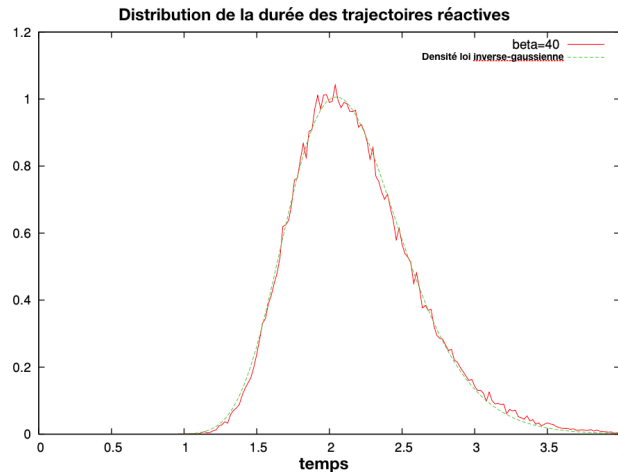


Pour $\beta = 10$ la température T est très basse, ce qui signifie le bruit dans l'équation (20) et encore très faible, la méthode MC montre ses limites, les estimations sont imprécises sur la courbe de gauche (c) pour la méthode MC. En revanche, l'algorithme BMA donne des résultats très satisfaisant même sous les températures très basses (courbe(d)) sous lesquelles la méthode MC cesse de fonctionner.

Pour s'assurer que l'algorithme BMA donne des bonnes estimations de la durée des trajectoires réactives, la loi inverse-gaussienne est un moyen fiable pour comparer les estimations, car le temps en lequel un processus stochastique atteint une valeur fixée est de loi inverse-gaussienne. Sa densité de probabilité est donnée par :

$$f_{\mu,\lambda}(x) = \left(\frac{\lambda}{2\pi x^3} \right) e^{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}} \mathbb{1}_{x>0}$$

où μ et λ sont deux paramètres positives. Ainsi pour qu'un processus unidimensionnel $Y_t = vt + \sigma W_t$ réalise un premier passage par la valeur $\alpha > 0$, le temps $T_\alpha = \inf\{t > 0 \mid Y_t = \alpha\}$ est aléatoire et suit la loi inverse-gaussienne de $\left(\frac{\alpha}{v}, \frac{\alpha^2}{\sigma^2}\right)$. La figure ci-dessous compare la densité de la loi inverse-gaussienne et la l'histogramme des trajectoires réactives BMA.



Sur ce schéma nous constatons aussi un bon accord entre l'algorithme BMA et la densité de la loi inverse gaussienne, ce qui confirme la performance de l'algorithme en terme de durée de trajectoire réactive à basse température.

5.4.2 Temps de calcul

Le tableau suivant représente le temps de calcul nécessaire pour l'apparition d'un trajectoire réactives pour la méthode MC et l'algorithme BMA

β	\hat{p}_{BMA}	temps MC	temps BMA
1	1.03×10^{-1}	2s	2s
10	2.04×10^{-5}	140 min 5 s	5s
15	1.78×10^{-7}	92000	7min52 s
20	1.33×10^{-9}		8 min 36 s
40	5.82×10^{-18}		10 min 09 s

l'algorithme permet un gain de temps de calcul très important, pour $\beta = 15$ le ratio entre la méthode MC et l'algorithme est d'ordre 1000.

5.4.3 Temps de transition

Nous souhaitant à présent estimer le temps T pour qu'une particule se déplace de l'état métastable $A = \{1\}$ à l'état métastable $B = \{-1\}$. Une telle trajectoire est appelé *temps de transition*, le temps d'un chemin de transition est supérieur à la durée de trajectoire réactive, puisque une trajectoire réactive est une partie du chemin de transition. pour obtenir une observation du temps de transition T , on lance des particules une par une jusque que la I -ième particule atteint le point critique B , le temps T est alors

$$T = \sum_{i=1}^{I-1} (T_1^i + T_2^i) + (T_1^I + T_3^I)$$

où

T_1^i : le temps pour que la i -ème lancée atteint z_{min} en partant de A .

T_2^i : le temps pour que la i -ème lancée atteint A en partant de z_{min} .

T_3^I : le temps pour que la I -ème lancée atteint B en partant de z_{min} .

Ce temps est une variable aléatoire *i.i.d*, puisque la loi de transition de chaque particule vérifie la propriété de Markov et l'équation (20). L'espérance de T est égale à,

$$\mathbb{E}[T] = \mathbb{E} \left[\sum_{i=1}^{I-1} (T_1^i + T_2^i) \right] + \mathbb{E}[T_1^I + T_3^I]$$

et comme pour $i \geq 1$, T_1^i (*resp.* T_2^i) sont *i.i.d* et I suit une loi géométrique, $\mathbb{P}(I = i) = (1 - p)^{i-1}p$, où p est la probabilité d'atteindre B en partant de z_{min} , donc $\mathbb{E}(I) = \frac{1}{p}$, ainsi

$$\begin{aligned} \mathbb{E}[T] &= (\mathbb{E}[I] - 1)\mathbb{E}[T_1 + T_2] + \mathbb{E}[T_1 + T_3] \\ &= \left(\frac{1}{p} - 1 \right) \mathbb{E}[T_1 + T_2] + \mathbb{E}[T_1 + T_3] \end{aligned} \quad (23)$$

$\mathbb{E}[T_1 + T_2]$ n'est pas un événement rare car par hypothèse, z_{min} , est très proche de $A = \{-1\}$, donc on peut l'estimer avec une simple MC. $\mathbb{E}[T_1 + T_3]$ qui correspond au temps moyen pour une trajectoire partant de A d'atteindre B sans

5.5 Cas bidimensionnel

retourner de nouveau à A , est obtenu par une légère modification de l'algorithme l'algorithme BMA précédent : à l'étape 2 de la procédure d'initialisation, nous ne rejetons pas le début de la trajectoire pour obtenir le temps intégral de A à B . pour estimer p on peut utiliser encore l'algorithme BMA en prenant tout simplement z_{min} condition initiale.

Le tableau suivant présente une comparaison entre la méthode MC et et l'algorithme avec différentes valeurs de β et de z_{min}

β	z_{min}	$\mathbb{E}[T] (MC)$	$\mathbb{E}[T] (BMA)$	I.C	erreur (%)
1	-0.9	3.634	3.911	[3.798, 4.029]	7.620
1	-0.6	3.634	3.533	[3.466, 3.602]	2.775
7	-0.9	1405	1380	[1316, 1449]	1.808
7	-0.6	1405	1400	[1358, 1444]	0.370

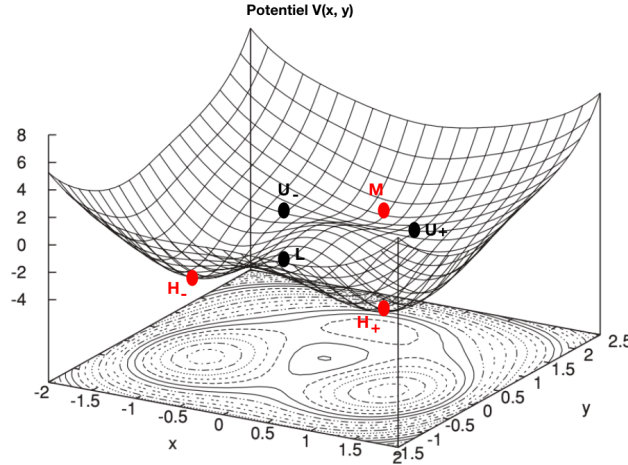
Nous pouvons constater que l'erreur diminue considérablement quand la température diminue mais aussi quand z_{min} augmente. Donc le choix de z_{min} infecte les estimations, plus il est proche de la région critique A plus l'erreur augmente.

5.5 Cas bidimensionnel

Pour illustrer le cas du processus bidimensionnel, nous considérons le potentiel de deux variables suivant,

$$V(x, y) = 3e^{-x^2 - (y - \frac{1}{3})^2} - 3e^{-x^2 - (y - \frac{5}{3})^2} - 5e^{-(x-1)^2 - y^2} - 5e^{-(x+1)^2 - y^2} + 0,2x^4 + 0,2\left(y - \frac{1}{3}\right)^4$$

Sa représentation graphique,



Ce potentiel possède trois minima locaux, H_- , H_+ , et M . Nous allons nous intéresser à la durée de passage des particules de la boule euclidienne de centre H_- et rayon r_1 à la boule euclidienne de centre H_+ et de rayon r_2 représentant respectivement les états métastables A et B , où la force appliquée à ces particules est très faible dans ces deux endroits. Cependant nous pouvons remarquer préalablement que si la température est très basse, ce qui signifie $\beta \rightarrow +\infty$, le

5.5 Cas bidimensionnel

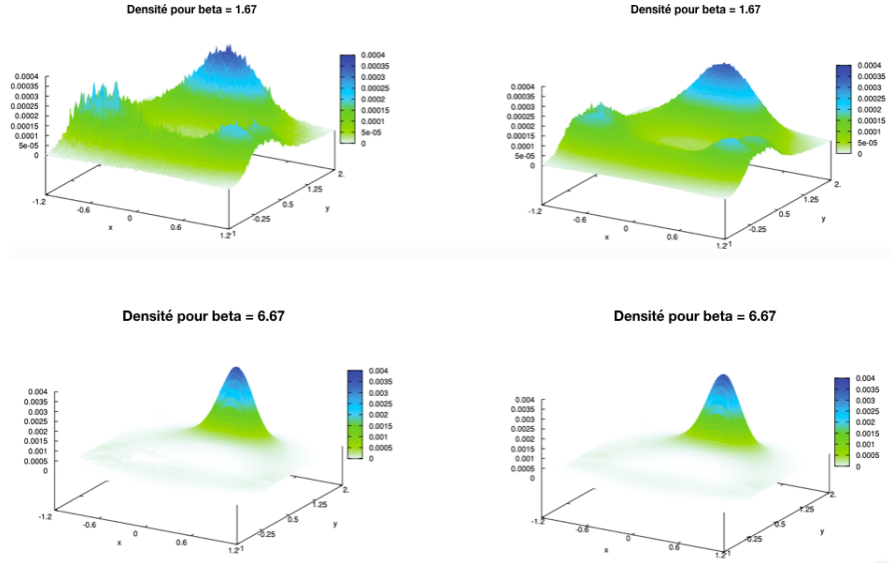
bruit dans l'équation (20) est encore très faible, les trajectoires réactives favorisent un passage par le canal supérieur (H_-, U_-, M, U_+, H_+) où U_-, U_+ , sont des points-selles, car la barrière de l'énergie potentielle est relativement inférieure par rapport à la barrière de l'énergie potentielle du canal supérieur (H_-, L, H_+). Alors que si la température est élevée, donc β est encore proche de zéro, les trajectoires réactives favorisent un passage par le canal inférieur (H_-, L, H_+), où L est un point-selle, car dans ce cas la particule traverse un seul point-selle pour atteindre l'état métastable B contre deux points-selles pour le canal supérieur.

5.5.1 Insensibilité au choix de la coordonnée de réaction

Afin de relever numériquement l'effet du choix de la coordonnée de réaction sur la pertinence du résultat, nous allons tracer la densité ρ des positions conditionnellement d'être dans une trajectoire réactive. Nous testons l'algorithme avec deux régimes de température : régime de température basse avec $\beta = 6.67$ et régime de température élevé avec $\beta = 1.67$, et ce, pour deux coordonnées de réactions différentes :

$$h_1(x, y) = x, \quad h_2(x, y) = \|(x, y) - H_-\|$$

où $\|\cdot\|$ est la distance euclidienne. Pour le paramétrage numérique nous prenons ; $r_1 = r_2 = 0.05$, $z_{min} = -0.9$ (resp. $z_{max} = 0.9$) pour la coordonnée de réaction ε_1 , $z_{min} = 0.05$ (resp. $z_{max} = 1.5$) pour la coordonnée de réaction ε_2 et $n = 10^5$ particules. Nous obtenons les résultats suivants :



La colonne de gauche correspond à la coordonnée de réaction h_1 et la colonne de droite à celle de h_2 . Les résultats confirment les prédictions dans les deux régimes. On observe que les résultats numériques sont légèrement meilleurs avec h_2 (qui est en fait plus proche de la fonction committor) mais h_1 donne des estimations très proches de h_2 .

Pour illustrer avec des résultats numériques les deux deux tab

Ainsi nous concluons que le choix de la coordonnée de réaction n'infecte pas les estimations : on a besoin de choisir une coordonnée de réaction proche de la fonction committor pour obtenir des bons résultats numériques.

5.5.2 Temps de transition

L'équation (23) reste valable dans le cas bidimensionnel (ou pour des dimensions supérieures) en cas de vérifier ces deux hypothèses :

- (i) Le nombre de lancers I est *i.i.d* et suit une loi géométrique de paramètre p (où p est la probabilité d'atteindre l'état métastable B avant A en partant de $\Sigma_{z_{min}}$). Cela veut dire que pour une trajectoire partant de ∂A et atteignant $\Sigma_{z_{min}}$, la probabilité d'atteindre ∂B ne dépend pas des points de contact dans $\Sigma_{z_{min}}$
- (ii) Si $\nu_{\partial A}$ et $\nu_{\partial \Sigma_{z_{min}}}$ sont respectivement la distribution des points initiaux dans A et la distribution des points d'atteinte dans $\Sigma_{z_{min}}$, les points d'atteinte des trajectoires partants de A suivant $\nu_{\partial A}$, sont distribués suivant $\nu_{\partial \Sigma_{z_{min}}}$ et les points d'atteinte des trajectoires partants de $\Sigma_{z_{min}}$ suivant $\nu_{\Sigma_{z_{min}}}$, sont distribués suivant $\nu_{\partial A}$ dans A .

Le tableau suivant présente les résultats obtenus pour l'estimation du temps de transition T par la méthode MC et l'algorithme BMS pour le processus bidimensionnel $V(x, y)$ avec la coordonnée de réaction h_2 , $z_{min} = 0.1$, et $z_{max} = 1.5$ et pour différentes valeurs de n .

n $\times 10^3$	$\mathbb{E}(T_1 + T_2)$ (MC)	$\mathbb{E}(T_1 + T_3)$ (BMA)	$\mathbb{E}(T)$ (BMA)	I.C $\mathbb{E}(T)$	$\mathbb{E}(T)$ (MC)	Erreur Relative %
2	2.72×10^{-1}	1.43	24.66	[21.63, 28.38]	26.54	7.61%
10	2.73×10^{-1}	1.56	27.14	[26.44, 27.52]	26.50	2.38%
50	2.74×10^{-1}	1.54	26.66	[26.45, 26.73]	26.53	0.48%
100	2.73×10^{-1}	1.52	26.41	[26.36, 26.46]	26.50	0.34%

Le tableau montre une cohérence entre les méthodes MC et BMC. L'erreur relative sur le temps de transition d'un chemin réactif diminue quand le nombre de particules simulées est augmenté ce qui montre l'efficacité de l'algorithme même pour estimer le temps de transition des particules.

6 Conclusion

Nous nous sommes d'abord intéressés à la méthode de Branchement Multi-niveaux Classique, qui est apparue dans la littérature scientifique comme une technique prometteuse pour estimer un événement rare, nous avons étudié les deux options associées à cette méthode, branchement-fixé et effort-fixé, de variance nettement trop petite par rapport à la variance de la méthode Monte Carlo. Cependant, même sous les conditions idéales les deux options ont montré leurs limites (quand le nombre de seuils tend vers l'infini pour l'option effort-fixé et quand les coefficients de branchement sont mal choisis pour l'option branchement-fixé) surtout au niveau pratique, quand le domaine est complexe,

la fixation des niveau n'est pas facile, et des fois impossible. Pour répondre à ces obstacles pratique la version Branchement Multi-niveaux adaptative est une alternative très puissante, puisque elle demande aucun paramétrage préalable et son estimateur est convergent. Nous avons testé la version adaptative dans le domaine de la dynamique moléculaire et la performance des résultat numérique ont confirmés les résultats théoriques. L'intérêt de la la version adaptative est qu'elle a permis toutefois d'estimer avec excellence le temps de transition d'une particule entre deux états métastables.

Néanmoins la version adaptative présente elle-même des limites. En effet, même si elle est théoriquement non biaisé, la discrétisation du temps continu dans la pratique peut entraîner un estimateur biaisé, et malgré que les performances de la version adaptative sont très proche de celles de la version classique en terme de variance, à présent, pas des résultats théoriques en terme de variance asymptotique.

Références

- [1] Frédéric Cérou & Arnaud Guyader (2007) Adaptive Multilevel Splitting for Rare Event Analysis, *Stochastic Analysis and Applications*, 25 :2, 417-443
- [2] Charles-Edouard Bréhier, Maxime Gazeau, Ludovic Goudenège, Tony Lelièvre, and Mathias Rousset, Unbiasedness of some generalized adaptive multilevel splitting algorithms
- [3] Paul Glasserman, Philip Heidelberger, Perwez Shahabuddin, Multilevel Splitting for Estimating Rare Event Probabilities.
- [4] Frédéric Cérou, Arnaud Guyader, Tony Lelièvre, and David Pommier, A multiple replica approach to simulate reactive trajectories (2012)
- [5] Arnaud Guyader, Nicolas Hengartner, Eric Matzner-Løbe, Simulation and Estimation of Extreme Quantiles and Extreme Probabilities