

Simulation d'Événements Rares : Application de l'algorithme Branchement Multi-niveaux Adaptatif

Mohammed Hajji

Université Paris Est Marne-La-Vallée

05/09/2019

Définition d'événement rare

Définition d'événement rare

$X = (X_t)_{t \geq 0}$: Processus stochastique à temps continu d'espace d'états \mathbb{R}^d

Définition d'événement rare

$X = (X_t)_{t \geq 0}$: Processus stochastique à temps continu d'espace d'états \mathbb{R}^d

B : Sous ensemble d'états

Définition d'événement rare

$X = (X_t)_{t \geq 0}$: Processus stochastique à temps continu d'espace d'états \mathbb{R}^d

B : Sous ensemble d'états

$R := \{X \in B\}$: Événement rare,

Définition d'événement rare

$X = (X_t)_{t \geq 0}$: Processus stochastique à temps continu d'espace d'états \mathbb{R}^d

B : Sous ensemble d'états

$R := \{X \in B\}$: Événement rare,

- $\mathbb{P}(R)$ est très proche de 0 mais n'égale pas 0. Disons,

$$0 < \mathbb{P}(R) \leq 10^{-9}.$$

Définition d'événement rare

$X = (X_t)_{t \geq 0}$: Processus stochastique à temps continu d'espace d'états \mathbb{R}^d

B : Sous ensemble d'états

$R := \{X \in B\}$: Événement rare,

- $\mathbb{P}(R)$ est très proche de 0 mais n'égale pas 0. Disons,

$$0 < \mathbb{P}(R) \leq 10^{-9}.$$

- Nous supposons que nous connaissons comment une pseudo réalisation de X peut se produire.

Définition d'événement rare

$X = (X_t)_{t \geq 0}$: Processus stochastique à temps continu d'espace d'états \mathbb{R}^d

B : Sous ensemble d'états

$R := \{X \in B\}$: Événement rare,

- $\mathbb{P}(R)$ est très proche de 0 mais n'égale pas 0. Disons,

$$0 < \mathbb{P}(R) \leq 10^{-9}.$$

- Nous supposons que nous connaissons comment une pseudo réalisation de X peut se produire.
- Il est très important de connaître cette probabilité ou de générer des réalisations typiques.

Définition d'événement rare

$X = (X_t)_{t \geq 0}$: Processus stochastique à temps continu d'espace d'états \mathbb{R}^d

B : Sous ensemble d'états

$R := \{X \in B\}$: Événement rare,

- $\mathbb{P}(R)$ est très proche de 0 mais n'égale pas 0. Disons,

$$0 < \mathbb{P}(R) \leq 10^{-9}.$$

- Nous supposons que nous connaissons comment une pseudo réalisation de X peut se produire.
- Il est très important de **connaître cette probabilité** ou de **générer des réalisations typiques**.
- **Différent d'une valeur extrême** : Un modèle que nous pouvons simuler \neq approche statistique, basé sur les données disponibles

Pourquoi une simple MC ne marche pas ?

Notons $\gamma = \mathbb{P}(R)$

Pourquoi une simple MC ne marche pas ?

Notons $\gamma = \mathbb{P}(R)$

La méthode de Monte Carlo : Simuler n copies X^1, \dots, X^n suivant la loi de X , compter le nombre de succès

$$\hat{\gamma}_{MC} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_R(X^i)$$

Pourquoi une simple MC ne marche pas ?

Notons $\gamma = \mathbb{P}(R)$

La méthode de Monte Carlo : Simuler n copies X^1, \dots, X^n suivant la loi de X , compter le nombre de succès

$$\hat{\gamma}_{MC} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_R(X^i)$$

$$\mathbb{E}[\hat{\gamma}_{MC}] = \gamma; \text{Var}(\hat{\gamma}_{MC}) = \frac{\gamma(1-\gamma)}{n}; RE(\hat{\gamma}) := \frac{\sqrt{\text{Var}(\hat{\gamma})}}{\gamma} = \frac{\sqrt{1-\gamma}}{\sqrt{n\gamma}},$$

Pourquoi une simple MC ne marche pas ?

Notons $\gamma = \mathbb{P}(R)$

La méthode de Monte Carlo : Simuler n copies X^1, \dots, X^n suivant la loi de X , compter le nombre de succès

$$\hat{\gamma}_{MC} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_R(X^i)$$

$$\mathbb{E}[\hat{\gamma}_{MC}] = \gamma; \text{Var}(\hat{\gamma}_{MC}) = \frac{\gamma(1-\gamma)}{n}; RE(\hat{\gamma}) := \frac{\sqrt{\text{Var}(\hat{\gamma})}}{\gamma} = \frac{\sqrt{1-\gamma}}{\sqrt{n\gamma}},$$

Si nous fixons cette RE à une valeur r , nous aurons besoin de simuler au minimum $n^* = r^{-2}(1-\gamma)/\gamma \approx r^{-2}/\gamma$ copies de processus. Ainsi, $n^* \xrightarrow{\gamma \rightarrow 0} +\infty$,

Exemple

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

Alternatives ?

Alternatives ?

- Échantillonnage Préférentiel

Alternatives ?

● Échantillonnage Préférentiel

- Changer la loi du processus X lui-même.
- Obtenir une loi pour laquelle l'événement R a une probabilité supérieur, ce qui rend l'estimateur MC plus efficace.

Alternatives ?

● Échantillonnage Préférentiel

- Changer la loi du processus X lui-même.
- Obtenir une loi pour laquelle l'événement R a une probabilité supérieur, ce qui rend l'estimateur MC plus efficace.
- Mesurer les valeurs de l'échantillon et les prendre en compte dans le changement de distribution. (ratio de vraisemblance).

Alternatives ?

● Échantillonnage Préférentiel

- Changer la loi du processus X lui-même.
- Obtenir une loi pour laquelle l'événement R a une probabilité supérieur, ce qui rend l'estimateur MC plus efficace.
- Mesurer les valeurs de l'échantillon et les prendre en compte dans le changement de distribution. (ratio de vraisemblance).

→ Il est difficile d'obtenir la distribution optimale

Alternatives ?

● Échantillonnage Préférentiel

- Changer la loi du processus X lui-même.
- Obtenir une loi pour laquelle l'événement R a une probabilité supérieur, ce qui rend l'estimateur MC plus efficace.
- Mesurer les valeurs de l'échantillon et les prendre en compte dans le changement de distribution. (ratio de vraisemblance).

→ Il est difficile d'obtenir la distribution optimale

● Branchement Multi-niveaux

Alternatives ?

● Échantillonnage Préférentiel

- Changer la loi du processus X lui-même.
- Obtenir une loi pour laquelle l'événement R a une probabilité supérieur, ce qui rend l'estimateur MC plus efficace.
- Mesurer les valeurs de l'échantillon et les prendre en compte dans le changement de distribution. (ratio de vraisemblance).

→ Il est difficile d'obtenir la distribution optimale

● Branchement Multi-niveaux

- Créer de nombreuses réalisations de l'événement rare par une technique de branchement/sélection

Alternatives ?

● Échantillonnage Préférentiel

- Changer la loi du processus X lui-même.
- Obtenir une loi pour laquelle l'événement R a une probabilité supérieur, ce qui rend l'estimateur MC plus efficace.
- Mesurer les valeurs de l'échantillon et les prendre en compte dans le changement de distribution. (ratio de vraisemblance).

→ Il est difficile d'obtenir la distribution optimale

● Branchement Multi-niveaux

- Créer de nombreuses réalisations de l'événement rare par une technique de branchement/sélection
- Deux versions : Version Classique et Version Adaptative

Alternatives ?

● Échantillonnage Préférentiel

- Changer la loi du processus X lui-même.
- Obtenir une loi pour laquelle l'événement R a une probabilité supérieur, ce qui rend l'estimateur MC plus efficace.
- Mesurer les valeurs de l'échantillon et les prendre en compte dans le changement de distribution. (ratio de vraisemblance).

→ Il est difficile d'obtenir la distribution optimale

● Branchement Multi-niveaux

- Créer de nombreuses réalisations de l'événement rare par une technique de branchement/sélection
- Deux versions : Version Classique et Version Adaptative
 - ★ Version classique : Un paramétrage préalable est nécessaire

Alternatives ?

● Échantillonnage Préférentiel

- Changer la loi du processus X lui-même.
- Obtenir une loi pour laquelle l'événement R a une probabilité supérieur, ce qui rend l'estimateur MC plus efficace.
- Mesurer les valeurs de l'échantillon et les prendre en compte dans le changement de distribution. (ratio de vraisemblance).

→ Il est difficile d'obtenir la distribution optimale

● Branchement Multi-niveaux

- Créer de nombreuses réalisations de l'événement rare par une technique de branchement/sélection
- Deux versions : Version Classique et Version Adaptative
 - ★ Version classique : Un paramétrage préalable est nécessaire
 - ★ Version adaptative : Pas de paramétrage (boite noire)

Exemples d'utilisation

- La gestion du trafic aérien

Exemples d'utilisation

- La gestion du trafic aérien
- Collision satellite et débris

Exemples d'utilisation

- La gestion du trafic aérien
- Collision satellite et débris
- File d'attente

Exemples d'utilisation

- La gestion du trafic aérien
- Collision satellite et débris
- File d'attente
- Marche aléatoire

Exemples d'utilisation

- La gestion du trafic aérien
- Collision satellite et débris
- File d'attente
- Marche aléatoire
- Dynamique moléculaire

Paramètres de bases

▷ $X := (X_t)_{t \geq 0}$ un processus stochastique de Markov, continu à droite, à valeurs dans l'espace \mathbb{R}^d , et $(\mathcal{F}_t)_{t \geq 0}$ sa filtration naturelle.

Paramètres de bases

- ▷ $X := (X_t)_{t \geq 0}$ un processus stochastique de Markov, continu à droite, à valeurs dans l'espace \mathbb{R}^d , et $(\mathcal{F}_t)_{t \geq 0}$ sa filtration naturelle.
- ▷ 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 "proche" de A .

Paramètres de bases

- ▷ $X := (X_t)_{t \geq 0}$ un processus stochastique de Markov, continu à droite, à valeurs dans l'espace \mathbb{R}^d , et $(\mathcal{F}_t)_{t \geq 0}$ sa filtration naturelle.
- ▷ 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 "proche" de A .
- ▷ A est un événement récurrent pour X , i.e, $\mathbb{P}(X \text{ retourne à } A) \approx 1$

Paramètres de bases

- ▷ $X := (X_t)_{t \geq 0}$ un processus stochastique de Markov, continu à droite, à valeurs dans l'espace \mathbb{R}^d , et $(\mathcal{F}_t)_{t \geq 0}$ sa filtration naturelle.
- ▷ 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 "proche" de A .
- ▷ A est un événement récurrent pour X , i.e, $\mathbb{P}(X \text{ retourne à } A) \approx 1$
- ▷ $T_A = \inf \{t > 0 : X_t \in A\}$ et $T_B = \inf \{t > 0 : X_t \in B\}$

Paramètres de bases

- ▷ $X := (X_t)_{t \geq 0}$ un processus stochastique de Markov, continu à droite, à valeurs dans l'espace \mathbb{R}^d , et $(\mathcal{F}_t)_{t \geq 0}$ sa filtration naturelle.
- ▷ 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 "proche" de A .
- ▷ A est un événement récurrent pour X , i.e, $\mathbb{P}(X \text{ retourne à } A) \approx 1$
- ▷ $T_A = \inf \{t > 0 : X_t \in A\}$ et $T_B = \inf \{t > 0 : X_t \in B\}$

Le but est d'estimer

$$\gamma = \mathbb{P}[T_B < T_A].$$

$R = \{T_B < T_A\}$ est donc l'événement rare.

Branchement Multi-niveaux Classique

▷ Soit $f : \mathbb{R}^d \longrightarrow \mathbb{R}$, une fonction continue, telle que, $Z_t = f(X_t)$.

$A := \{x \in \mathbb{R}^d : f(x) \leq 0\}$ et $B := \{x \in \mathbb{R}^d : f(x) \geq L_{\max}\}$.

Branchement Multi-niveaux Classique

▷ Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}$, une fonction continue, telle que, $Z_t = f(X_t)$.

$A := \{x \in \mathbb{R}^d : f(x) \leq 0\}$ et $B := \{x \in \mathbb{R}^d : f(x) \geq L_{\max}\}$.

→ f est appelée **la fonction d'importance**

Branchement Multi-niveaux Classique

▷ Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}$, une fonction continue, telle que, $Z_t = f(X_t)$.

$A := \{x \in \mathbb{R}^d : f(x) \leq 0\}$ et $B := \{x \in \mathbb{R}^d : f(x) \geq L_{\max}\}$.

→ f est appelée **la fonction d'importance**

▷ Comme il est difficile d'atteindre le seuil supérieur L_{\max} en une seule fois, on divise le domaine $[0, L_{\max}]$ en m sous intervalles disjoints $[L_i, L_{i+1}[$, $i \in 1, \dots, m-1$.

Branchement Multi-niveaux Classique

▷ Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}$, une fonction continue, telle que, $Z_t = f(X_t)$.

$A := \{x \in \mathbb{R}^d : f(x) \leq 0\}$ et $B := \{x \in \mathbb{R}^d : f(x) \geq L_{\max}\}$.

→ f est appelée **la fonction d'importance**

▷ Comme il est difficile d'atteindre le seuil supérieur L_{\max} en une seule fois, on divise le domaine $[0, L_{\max}]$ en m sous intervalles disjoints $[L_i, L_{i+1}[$, $i \in 1, \dots, m-1$.

▷ $T_i = \inf \{t \geq 0, Z_t \geq L_i\}$; $D_i := \{T_i < T_A\}$

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

Branchement Multi-niveaux Classique

▷ Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}$, une fonction continue, telle que, $Z_t = f(X_t)$.

$A := \{x \in \mathbb{R}^d : f(x) \leq 0\}$ et $B := \{x \in \mathbb{R}^d : f(x) \geq L_{\max}\}$.

→ f est appelée **la fonction d'importance**

▷ Comme il est difficile d'atteindre le seuil supérieur L_{\max} en une seule fois, on divise le domaine $[0, L_{\max}]$ en m sous intervalles disjoints $[L_i, L_{i+1}[$, $i \in 1, \dots, m-1$.

▷ $T_i = \inf \{t \geq 0, Z_t \geq L_i\}$; $D_i := \{T_i < T_A\}$

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

▷ L'événement rare est une intersection de la séquence emboîtée d'événements $D_1 \subset D_2 \subset \dots \subset D_m$

Branchement Multi-niveaux Classique

▷ Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}$, une fonction continue, telle que, $Z_t = f(X_t)$.

$A := \{x \in \mathbb{R}^d : f(x) \leq 0\}$ et $B := \{x \in \mathbb{R}^d : f(x) \geq L_{\max}\}$.

→ f est appelée **la fonction d'importance**

▷ Comme il est difficile d'atteindre le seuil supérieur L_{\max} en une seule fois, on divise le domaine $[0, L_{\max}]$ en m sous intervalles disjoints $[L_i, L_{i+1}[$, $i \in 1, \dots, m-1$.

▷ $T_i = \inf \{t \geq 0, Z_t \geq L_i\}$; $D_i := \{T_i < T_A\}$

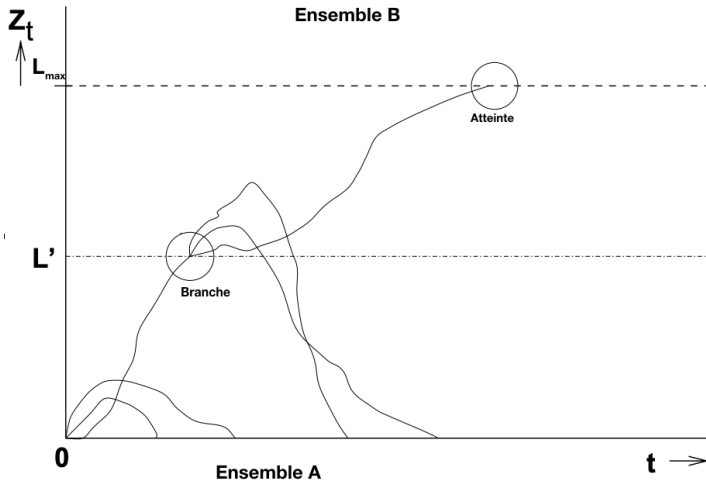
D_i est l'événement "le seuil L_i est atteint pendant la simulation".

▷ L'événement rare est une intersection de la séquence emboîtée d'événements $D_1 \subset D_2 \subset \dots \subset D_m$

▷ L'événement rare est donc : $R := \{T_B < T_A\} := D_m$

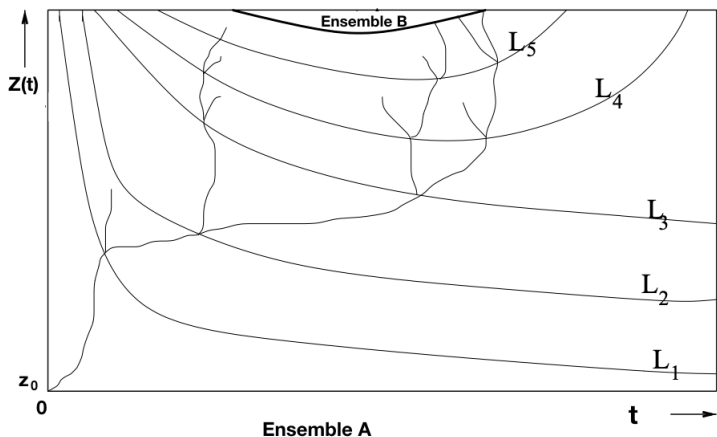
La version BMC

Nombre de trajectoires simulées à $t = 0$ est $n = 3$, coefficient de branchement $n_1 = 2$



La version BMC

Nombre de trajectoires simulées à $t = 0$ est $n = 1$, coefficient de branchement $n_i = 2$



Paramétrage de la version BMC

Paramétrage de la version BMC

- Il faut prendre soin de choisir le coefficient de branchements n_i à chaque seuil de manière appropriée :

Paramétrage de la version BMC

- Il faut prendre soin de choisir le coefficient de branchements n_i à chaque seuil 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.

Paramétrage de la version BMC

- Il faut prendre soin de choisir le coefficient de branchements n_i à chaque seuil 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."

Paramétrage de la version BMC

- Il faut prendre soin de choisir le coefficient de branchements n_i à chaque seuil 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 faut bien choisir où placer les niveaux.

Analyses

- R_i le nombre d'atteintes au seuil i . ($R_0 := 1$)

Analyses

- R_i le nombre d'atteintes au seuil i . ($R_0 := 1$)
- p_i la probabilité qu'une trajectoire partante d'une branche du seuil inférieure de l'étape i d'atteindre le seuil supérieur avant de s'arrêter.

Analyses

- R_i le nombre d'atteintes au seuil i . ($R_0 := 1$)
- p_i la probabilité qu'une trajectoire partante d'une branche du seuil inférieure de l'étape i d'atteindre le seuil supérieur avant de s'arrêter.
- 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 la variable aléatoire $S_i := X_{T_i}$.

$$R_{i+1} = \sum_{j=1}^{R_i} \sum_{k=1}^{n_i} I_{i+1}^{(k)}(S_i^{(j)}) \sim \sum_{k=1}^{R_i} \text{Bin} \left(n_i, p_{i+1}(S_i^{(k)}) \right)$$

$I_{i+1}(s)$ 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}

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

$$\hat{p}_i = \frac{R_i}{n_{i-1} R_{i-1}}, \text{ c'est un estimateur non biaisé de } p_i.$$

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

$$\hat{p}_i = \frac{R_i}{n_{i-1} R_{i-1}}, \text{ c'est un estimateur non biaisé de } p_i.$$

- γ est une séquence d'événements emboîtés $(D_i)_{0 \leq i \leq m}$ donc :

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

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

$$\hat{p}_i = \frac{R_i}{n_{i-1} R_{i-1}}, \text{ c'est un estimateur non biaisé de } p_i.$$

- γ est une séquence d'événements emboîtés $(D_i)_{0 \leq i \leq m}$ donc :

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

$$\hat{\gamma}_{BMC} = \prod_{i=1}^m \hat{p}_i = \prod_{i=1}^m \frac{R_i}{n_{i-1} R_{i-1}} = \frac{R_m}{\prod_{i=0}^{m-1} n_i}$$

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

L'estimateur $\hat{\gamma}_{BMC}$ est non biaisé

Branchement-fixé Vs Effort-fixé

- Deux options pour instaurer les branchements des trajectoires atteignant le seuil L_i à l'étape $i + 1$

Branchement-fixé Vs Effort-fixé

- Deux options pour instaurer les branchements des trajectoires atteignant le seuil L_i à l'étape $i + 1$
 - ▷ **Branchement-fixé** : Brancher chacune des R_i trajectoires qui ont atteint le seuil L_i par c_i copies, pour un entier positif fixe c_i . Ainsi, chaque entier $N_i = c_i R_i$ est aléatoire.

Branchement-fixé Vs Effort-fixé

- Deux options pour instaurer les branchements des trajectoires atteignant le seuil L_i à l'étape $i + 1$
 - ▷ **Branchement-fixé** : Brancher chacune des R_i trajectoires qui ont atteint le seuil L_i par c_i copies, pour un entier positif fixe c_i . Ainsi, chaque entier $N_i = c_i R_i$ est aléatoire.
 - ▷ **Effort-fixé** : Fixer à priori chaque N_i et simuler à chaque étape la bonne quantité de branchement pour atteindre cette valeur.

Branchement-fixé Vs Effort-fixé

- Deux options pour instaurer les branchements des trajectoires atteignant le seuil L_i à l'étape $i + 1$
 - ▷ **Branchement-fixé** : Brancher chacune des R_i trajectoires qui ont atteint le seuil L_i par c_i copies, pour un entier positif fixe c_i . Ainsi, chaque entier $N_i = c_i R_i$ est aléatoire.
 - ▷ **Effort-fixé** : Fixer à priori chaque N_i et simuler à chaque étape la bonne quantité de branchement pour atteindre cette valeur.
- Les estimateurs $\hat{\gamma}_{bf}$ et $\hat{\gamma}_{ef}$ de γ respectivement par les options branchement fixé et effort fixé, restent des estimateurs non biaisés, mais elles diffèrent en termes de variance.

Variance Effort-fixé

L'option **Effort-fixé** avec les **paramètres idéales** :

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

et $\hat{p}_i \sim \mathcal{B}(n, p = \gamma^{\frac{1}{m}})$.

Variance Effort-fixé

L'option **Effort-fixé** avec les **paramètres idéales** :

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

et $\hat{p}_i \sim \mathcal{B}(n, p = \gamma^{\frac{1}{m}})$. Alors, pour $m > 1$,

$$\text{Var}[\hat{\gamma}_{ef}] = \frac{mp^{2m-1}(1-p)}{n} + \frac{m(m-1)p^{2m-2}(1-p)^2}{2n^2} + \dots + \frac{(p(1-p))^m}{n^m}.$$

Variance Effort-fixé

L'option **Effort-fixé** avec les **paramètres idéales** :

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

et $\hat{p}_i \sim \mathcal{B}(n, p = \gamma^{\frac{1}{m}})$. Alors, pour $m > 1$,

$$\text{Var}[\hat{\gamma}_{ef}] = \frac{mp^{2m-1}(1-p)}{n} + \frac{m(m-1)p^{2m-2}(1-p)^2}{2n^2} + \dots + \frac{(p(1-p))^m}{n^m}.$$

Si nous supposons de plus

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

le premier terme $\frac{mp^{2m-1}(1-p)}{n} \simeq \frac{m\gamma^{2-\frac{1}{m}}}{n}$ domine la dernière expression

Variance Effort-fixé vs variance MC

$$\bullet \text{Var}[\hat{\gamma}_{ef}] \simeq \frac{m\gamma^{2-\frac{1}{m}}}{n}$$

Variance Effort-fixé vs variance MC

- $\text{Var}[\hat{\gamma}_{ef}] \simeq \frac{m\gamma^{2-\frac{1}{m}}}{n}$
- $\text{Var}[\hat{\gamma}_{MC}] = \frac{\gamma(1-\gamma)}{n} \simeq \frac{\gamma}{n}.$

Variance Effort-fixé vs variance MC

$$\bullet \mathbf{Var}[\hat{\gamma}_{ef}] \simeq \frac{m\gamma^{2-\frac{1}{m}}}{n}$$

$$\bullet \mathbf{Var}[\hat{\gamma}_{MC}] = \frac{\gamma(1-\gamma)}{n} \simeq \frac{\gamma}{n}.$$

▷ Pour illustrer l'énorme réduction de variance, supposons

$\gamma = 10^{-20}$, $m = 20$, $p = \frac{1}{10}$ et $n = 1000$

$\mathbf{Var}[\hat{\gamma}_{MC}] = 10^{-23}$, tandis que $\mathbf{Var}[\hat{\gamma}_{ef}] \simeq 1.8 \times 10^{-41}$.

Variance Effort-fixé vs variance MC

$$\bullet \mathbf{Var}[\hat{\gamma}_{ef}] \simeq \frac{m\gamma^{2-\frac{1}{m}}}{n}$$

$$\bullet \mathbf{Var}[\hat{\gamma}_{MC}] = \frac{\gamma(1-\gamma)}{n} \simeq \frac{\gamma}{n}.$$

▷ Pour illustrer l'énorme réduction de variance, supposons

$$\gamma = 10^{-20}, m = 20, p = \frac{1}{10} \text{ et } n = 1000$$

$$\mathbf{Var}[\hat{\gamma}_{MC}] = 10^{-23}, \text{ tandis que } \mathbf{Var}[\hat{\gamma}_{ef}] \simeq 1.8 \times 10^{-41}.$$

En terme d'Efficacité de $\hat{\gamma}_{ef}$

$$\bullet \mathbf{Eff}[\hat{\gamma}_{ef}] := \frac{1}{\mathbf{Var}[\hat{\gamma}_{ef}]K[\hat{\gamma}_{ef}]} \simeq \frac{\gamma^{-2+\frac{1}{m}}}{m^2}, \text{ sous } (\star). \text{ Maximale en } m = \frac{-\log(\gamma)}{2}$$

Variance Effort-fixé vs variance MC

$$\bullet \mathbf{Var}[\hat{\gamma}_{ef}] \simeq \frac{m\gamma^{2-\frac{1}{m}}}{n}$$

$$\bullet \mathbf{Var}[\hat{\gamma}_{MC}] = \frac{\gamma(1-\gamma)}{n} \simeq \frac{\gamma}{n}.$$

▷ Pour illustrer l'énorme réduction de variance, supposons

$$\gamma = 10^{-20}, m = 20, p = \frac{1}{10} \text{ et } n = 1000$$

$$\mathbf{Var}[\hat{\gamma}_{MC}] = 10^{-23}, \text{ tandis que } \mathbf{Var}[\hat{\gamma}_{ef}] \simeq 1.8 \times 10^{-41}.$$

En terme d'Efficacité de $\hat{\gamma}_{ef}$

$$\bullet \mathbf{Eff}[\hat{\gamma}_{ef}] := \frac{1}{\mathbf{Var}[\hat{\gamma}_{ef}]K[\hat{\gamma}_{ef}]} \simeq \frac{\gamma^{-2+\frac{1}{m}}}{m^2}, \text{ sous } (\star). \text{ Maximale en}$$

$$m = \frac{-\log(\gamma)}{2}$$

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

Variance Effort-fixé vs variance MC

$$\bullet \text{Var}[\hat{\gamma}_{ef}] \simeq \frac{m\gamma^{2-\frac{1}{m}}}{n}$$

$$\bullet \text{Var}[\hat{\gamma}_{MC}] = \frac{\gamma(1-\gamma)}{n} \simeq \frac{\gamma}{n}.$$

▷ Pour illustrer l'énorme réduction de variance, supposons

$$\gamma = 10^{-20}, m = 20, p = \frac{1}{10} \text{ et } n = 1000$$

$$\text{Var}[\hat{\gamma}_{MC}] = 10^{-23}, \text{ tandis que } \text{Var}[\hat{\gamma}_{ef}] \simeq 1.8 \times 10^{-41}.$$

En terme d'Efficacité de $\hat{\gamma}_{ef}$

$$\bullet \text{Eff}[\hat{\gamma}_{ef}] := \frac{1}{\text{Var}[\hat{\gamma}_{ef}]K[\hat{\gamma}_{ef}]} \simeq \frac{\gamma^{-2+\frac{1}{m}}}{m^2}, \text{ sous } (\star). \text{ Maximale en } m = \frac{-\log(\gamma)}{2}$$

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

• Considérons maintenant $\gamma \rightarrow 0$ et p fixe, alors $m \rightarrow \infty$, et l'hypothèse (\star) n'est donc plus possible. et *l'efficacité relative* converge vers 0

Variance Effort-fixé vs variance MC

$$\bullet \mathbf{Var}[\hat{\gamma}_{ef}] \simeq \frac{m\gamma^{2-\frac{1}{m}}}{n}$$

$$\bullet \mathbf{Var}[\hat{\gamma}_{MC}] = \frac{\gamma(1-\gamma)}{n} \simeq \frac{\gamma}{n}.$$

▷ Pour illustrer l'énorme réduction de variance, supposons

$$\gamma = 10^{-20}, m = 20, p = \frac{1}{10} \text{ et } n = 1000$$

$$\mathbf{Var}[\hat{\gamma}_{MC}] = 10^{-23}, \text{ tandis que } \mathbf{Var}[\hat{\gamma}_{ef}] \simeq 1.8 \times 10^{-41}.$$

En terme d'Efficacité de $\hat{\gamma}_{ef}$

$$\bullet \mathbf{Eff}[\hat{\gamma}_{ef}] := \frac{1}{\mathbf{Var}[\hat{\gamma}_{ef}]K[\hat{\gamma}_{ef}]} \simeq \frac{\gamma^{-2+\frac{1}{m}}}{m^2}, \text{ sous } (\star). \text{ Maximale en } m = \frac{-\log(\gamma)}{2}$$

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

• Considérons maintenant $\gamma \rightarrow 0$ et p fixe, alors $m \rightarrow \infty$, et l'hypothèse (\star) n'est donc plus possible. et *l'efficacité relative* converge vers 0

• Cet estimateur **n'est pas tout à fait asymptotiquement efficace**

Variance Branchement-fixé

- Un estimateur *asymptotiquement efficace* si,

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

Variance Branchement-fixé

- Un estimateur *asymptotiquement efficace* si,

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

- Sous les **conditions idéales** : $N_0 = n$, $p_i = p = \gamma^{\frac{1}{m}}$, coefficient de branchement fixé à $c = \frac{1}{p}$ i.e, $N_i = R_i c = \frac{R_i}{p}$. Alors

Variance Branchement-fixé

- Un estimateur *asymptotiquement efficace* si,

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

- Sous les **conditions idéales** : $N_0 = n$, $p_i = p = \gamma^{\frac{1}{m}}$, coefficient de branchement fixé à $c = \frac{1}{p}$ i.e, $N_i = R_i c = \frac{R_i}{p}$. Alors

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

$$\text{Var}[\hat{\gamma}_{bf}] = \frac{p^{2m-2}}{n^2} \text{Var}[R_m] = \frac{m(1-p)p^{2m-1}}{n}$$

$$\text{Var}[R_i] = nm(1-p) \text{ (Harris, 1964)}$$

Variance Branchement-fixé

- Un estimateur *asymptotiquement efficace* si,

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

- Sous les **conditions idéales** : $N_0 = n$, $p_i = p = \gamma^{\frac{1}{m}}$, coefficient de branchement fixé à $c = \frac{1}{p}$ i.e, $N_i = R_i c = \frac{R_i}{p}$. Alors

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

$$\text{Var}[\hat{\gamma}_{bf}] = \frac{p^{2m-2}}{n^2} \text{Var}[R_m] = \frac{m(1-p)p^{2m-1}}{n}$$

$$\text{Var}[R_i] = nm(1-p) \text{ (Harris, 1964)}$$

- Le temps de calcul est $K[\hat{\gamma}_{bf}] = c(N_0 + \dots + N_{m-1})$. de plus

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

Variance Branchement-fixé

- Un estimateur *asymptotiquement efficace* si,

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

- Sous les **conditions idéales** : $N_0 = n$, $p_i = p = \gamma^{\frac{1}{m}}$, coefficient de branchement fixé à $c = \frac{1}{p}$ i.e, $N_i = R_i c = \frac{R_i}{p}$. Alors

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

$$\text{Var}[\hat{\gamma}_{bf}] = \frac{p^{2m-2}}{n^2} \text{Var}[R_m] = \frac{m(1-p)p^{2m-1}}{n}$$

$$\text{Var}[R_i] = nm(1-p) \text{ (Harris, 1964)}$$

- Le temps de calcul est $K[\hat{\gamma}_{bf}] = c(N_0 + \dots + N_{m-1})$. de plus

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

- L'estimateur $\hat{\gamma}_{bf}$ est **asymptotiquement efficace**.

Inconvénients

Effort-fixé

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".

Inconvénients

Effort-fixé

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".

Branchement-fixé

Nécessite des choix très précis des coefficients de branchement.

Branchement Multi-niveaux Adaptative

- Proposé par Cérou et Guyader, 2007.

Branchement Multi-niveaux Adaptative

- Proposé par Cérou et Guyader, 2007.
- Largement utilisée dans la pratique, car ne nécessite aucune connaissance du domaine.

Branchement Multi-niveaux Adaptative

- Proposé par Cérou et Guyader, 2007.
- Largement utilisée dans la pratique, car ne nécessite aucune connaissance du domaine.
- Pas de paramétrage particulier.

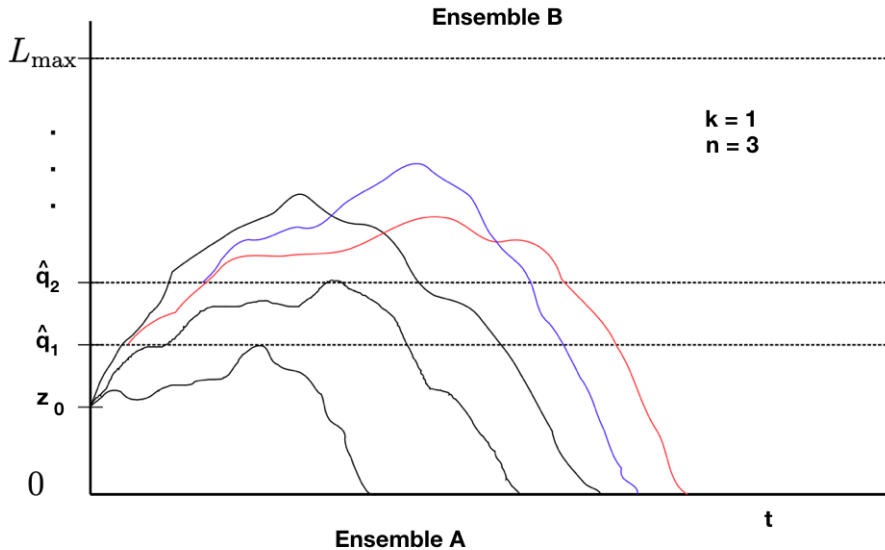
Branchement Multi-niveaux Adaptative

- Proposé par Cérou et Guyader, 2007.
- Largement utilisée dans la pratique, car ne nécessite aucune connaissance du domaine.
- Pas de paramétrage particulier.

Description de la version adaptative

Au lieu de choisir les niveaux à atteindre à priori, on choisit le nombre k de trajectoires à rejeter sur n , et chaque niveau correspond à une étape (itération de l'algorithme) et un quantile empirique des niveaux maximaux atteints par l'ensemble des trajectoires.

Le Modèle BMA



Algorithme

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

Algorithme

- **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 , Soit :

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

Algorithme

- **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 , 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,

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

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

Algorithme

- **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 , 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,

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

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

- Garder en mémoire la quantité :

$$\hat{q}_1 = M_{n,\varepsilon(k)}^1$$

- Pas 2 : Maintenir $\left(M_{n,\varepsilon(k+1)}^1, M_{n,\varepsilon(n-u+2)}^1, \dots, M_{n,\varepsilon(n)}^1 \right)$ inchangés. Notons les $\left(M_{n,n-u+1}^2, M_{n,n-u+2}^2, \dots, M_{n,n}^2 \right)$.

- Pas 2 : Maintenir $\left(M_{n,\varepsilon(k+1)}^1, M_{n,\varepsilon(n-u+2)}^1, \dots, M_{n,\varepsilon(n)}^1\right)$ inchangés. Notons les $\left(M_{n,n-u+1}^2, M_{n,n-u+2}^2, \dots, M_{n,n}^2\right)$.
• Rejeter les k trajectoires auxquelles correspond $\left(M_{n,\varepsilon(1)}^1, \dots, M_{n,\varepsilon(k)}^1\right)$

- Pas 2 : Maintenir $\left(M_{n,\varepsilon(k+1)}^1, M_{n,\varepsilon(n-u+2)}^1, \dots, M_{n,\varepsilon(n)}^1\right)$ inchangés. Notons les $\left(M_{n,n-u+1}^2, M_{n,n-u+2}^2, \dots, M_{n,n}^2\right)$.
 - Rejeter les k trajectoires auxquelles correspond $\left(M_{n,\varepsilon(1)}^1, \dots, M_{n,\varepsilon(k)}^1\right)$
 - Simuler k nouvelles trajectoires (branchement) $(X_t^j)_{t \geq 0}$ à partir du point initial $x_0^{(1)}$ tel que $f\left(x_0^{(1)}\right) = \hat{q}_1$.

- Pas 2 : Maintenir $\left(M_{n,\varepsilon(k+1)}^1, M_{n,\varepsilon(n-u+2)}^1, \dots, M_{n,\varepsilon(n)}^1\right)$ inchangés. Notons les $\left(M_{n,n-u+1}^2, M_{n,n-u+2}^2, \dots, M_{n,n}^2\right)$.
 - Rejeter les k trajectoires auxquelles correspond $\left(M_{n,\varepsilon(1)}^1, \dots, M_{n,\varepsilon(k)}^1\right)$
 - Simuler k nouvelles trajectoires (branchement) $(X_t^j)_{t \geq 0}$ à partir du point initial $x_0^{(1)}$ tel que $f\left(x_0^{(1)}\right) = \hat{q}_1$.
 - Attendre jusqu'à que toutes les k trajectoires rejoignent l'ensemble A : Pour chaque $j \in \{1, \dots, k\}$, notons $M_{n,j}^2 = \sup_{0 \leq t \leq T_A^j} Z_t^j$,

- Pas 2 : Maintenir $\left(M_{n,\varepsilon(k+1)}^1, M_{n,\varepsilon(n-u+2)}^1, \dots, M_{n,\varepsilon(n)}^1\right)$ inchangés. Notons les $\left(M_{n,n-u+1}^2, M_{n,n-u+2}^2, \dots, M_{n,n}^2\right)$.
 - Rejeter les k trajectoires auxquelles correspond $\left(M_{n,\varepsilon(1)}^1, \dots, M_{n,\varepsilon(k)}^1\right)$
 - Simuler k nouvelles trajectoires (branchement) $(X_t^j)_{t \geq 0}$ à partir du point initial $x_0^{(1)}$ tel que $f\left(x_0^{(1)}\right) = \hat{q}_1$.
 - Attendre jusqu'à que toutes les k trajectoires rejoignent l'ensemble A : Pour chaque $j \in \{1, \dots, k\}$, notons $M_{n,j}^2 = \sup_{0 \leq t \leq T_A^j} Z_t^j$,
 - Réordonner $\left(M_{n,1}^2, M_{n,2}^2, \dots, M_{n,n}^2\right)$ dans un ordre croissant comme suit :

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

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

- Garder en mémoire la quantité :

$$\hat{q}_2 = M_{n,\varepsilon(k)}^2$$

- Garder en mémoire la quantité :

$$\hat{q}_2 = M_{n,\varepsilon(k)}^2$$

Pas 3 : Répéter la procédure jusqu'à l'itération \hat{N} telle que,
 $\hat{q}_{\hat{N}+1} \geq L_{max}$.

- Garder en mémoire la quantité :

$$\hat{q}_2 = M_{n,\varepsilon(k)}^2$$

Pas 3 : 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.

- Garder en mémoire la quantité :

$$\hat{q}_2 = M_{n,\varepsilon(k)}^2$$

Pas 3 : 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.

Pas 4 : Calculer l'estimateur de probabilité $\hat{\gamma}_{BMA}$ de γ .
 Notons $p = \frac{n-k}{n} = 1 - \frac{k}{n}$, alors

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

Discussion sur BMA

- L'estimateur $\hat{\gamma}_{BMA}$ est **non biaisé**.

Discussion sur BMA

- L'estimateur $\hat{\gamma}_{BMA}$ est non biaisé.
- L'estimateur $\hat{\gamma}_{BMA}$ est convergent (consistant).

Discussion sur BMA

- L'estimateur $\hat{\gamma}_{BMA}$ est **non biaisé**.
- L'estimateur $\hat{\gamma}_{BMA}$ est **convergent (consistant)**.
- 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 d'informations sur son domaine.

Discussion sur BMA

- L'estimateur $\hat{\gamma}_{BMA}$ est **non biaisé**.
- L'estimateur $\hat{\gamma}_{BMA}$ est **convergent (consistant)**.
- 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 d'informations sur son domaine.
- 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.

Applications dans dynamique moléculaire

- Adaptation de l'algorithme BMA pour estimer la probabilité qu'une particule partante d'un état métastable A , atteint un autre état métastable B . Un tel trajectoire est appelée **chemin réactif**.

Applications dans dynamique moléculaire

- Adaptation de l'algorithme BMA pour estimer la probabilité qu'une particule partante d'un état métastable A , atteint un autre état métastable B . Un tel trajectoire est appelée **chemin réactif**.
- Estimer le temps de transition d'un chemin réactif sous différentes températures.

Applications dans dynamique moléculaire

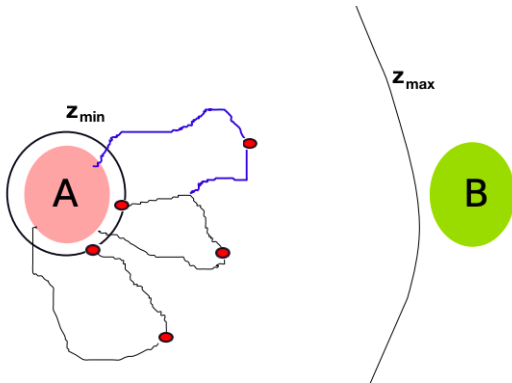
Chaque particule du système suit un processus stochastique $(X_t)_{t \geq 0}$ solution de l'équation différentielle :

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

appelée équation de Langevin. Où,

- La fonction, $V : \mathbb{R}^d \rightarrow \mathbb{R}$ décrit l'énergie potentielle.
- ∇ 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.

Illustration du transition



Initialisation

- Générer un ensemble de n conditions initiales $(x_0^j)_{1 \leq j \leq n}$ selon la mesure μ conditionnellement à $x_0^j \in A, \forall j \in \{1, \dots, n\}$

Initialisation

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

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

Initialisation

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

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

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

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

Initialisation

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

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

- 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, $\sup_{t \geq} h(X_t^j) > z_{min}$.

Itérations : L'algorithme se procède d'une manière normale.

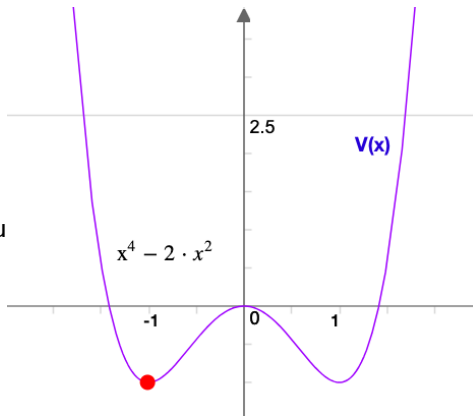
Application : Cas unidimensionnel

Potentiel unidimensionnel

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

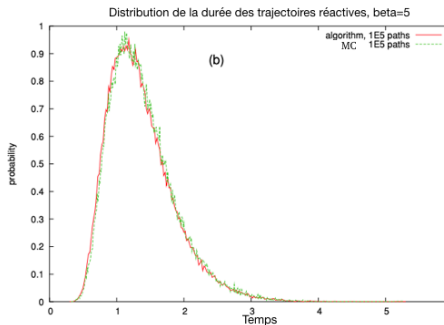
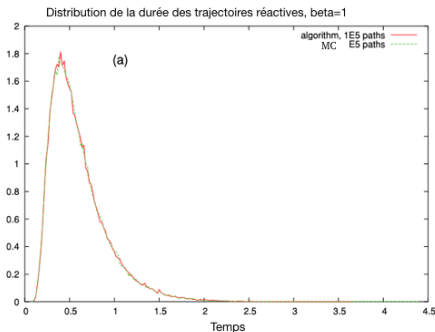
deux minima locaux, -1 et $+1$,
correspondants respectivement au
deux états métastables A et B .

⇒ Pas besoin d'une coordonnée
de réaction



Distribution de la durée des trajectoires réactives

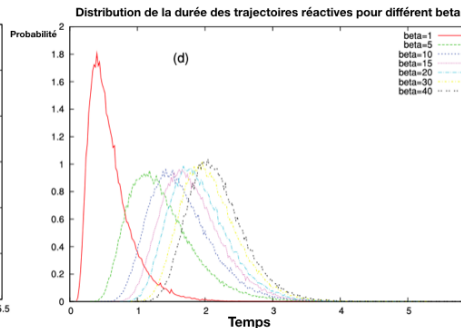
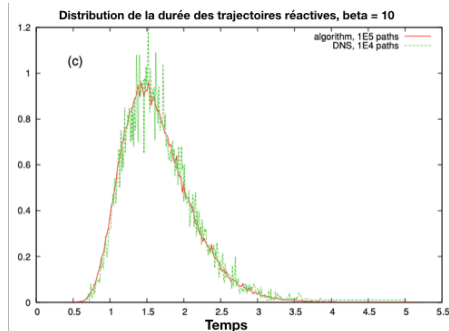
Les deux courbes représente la distribution de la durée des trajectoires réactives pour $\beta = 1$ et $\beta = 5$,



Pour β petit, les deux méthode donnent des résultats très proches et très précis

Distribution de la durée des trajectoires réactives

Les deux courbes représente la distribution de la durée des trajectoires réactives pour $\beta = 10$ à gauche, et différents valeurs de β à droite

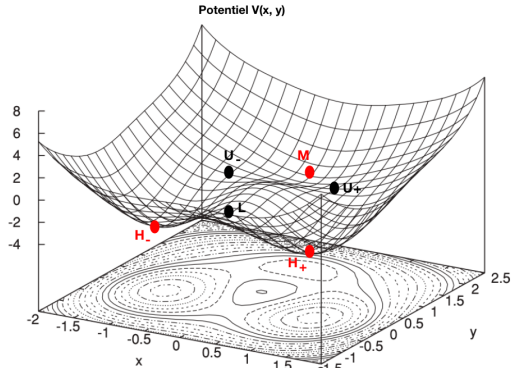


La méthode MC montre ses limites, les estimations sont imprécises. L'algorithme BMA continu de fonctionner.

Application : Cas bidimensionnel

Potentiel bidimensionnel

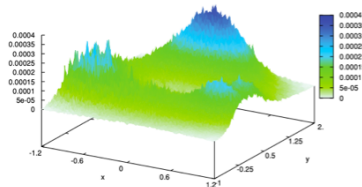
$$V(x, y) = 3e^{-x^2 - (y - \frac{1}{3})^2} - 3e^{-x^2 - (y - \frac{5}{3})^2} - 5e^{-(x-1)^2 - y^2} - 5e^{-(x+1)^2 - y^2} + 0,2x^4 + 0,2\left(y - \frac{1}{3}\right)^4$$



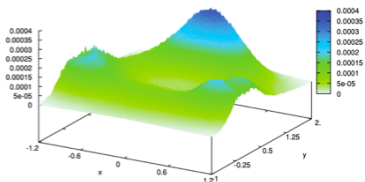
Distributions des positions

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

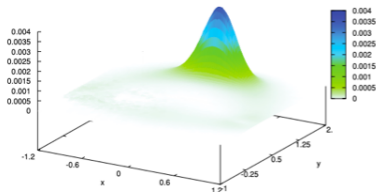
Densité pour beta = 1.67



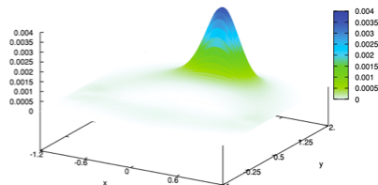
Densité pour beta = 1.67



Densité pour beta = 6.67



Densité pour beta = 6.67



Temps de transition

Pour obtenir une observation du temps de transition T , on lance des particules une par une jusqu'à ce que la I -ième particule atteigne 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 -ième lancée atteigne z_{min} en partant de A .

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

T_3^I : le temps pour que la I -ième lancée atteigne B en partant de z_{min}

Temps de transition

Pour obtenir une observation du temps de transition T , on lance des particules une par une jusqu'à ce que la I -ième particule atteigne 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 -ième lancée atteigne z_{min} en partant de A .

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

T_3^I : le temps pour que la I -ième lancée atteigne B en partant de z_{min}

$$\mathbb{E}[T] = \left(\frac{1}{p} - 1 \right) \mathbb{E}[T_1 + T_2] + \mathbb{E}[T_1 + T_3]$$

Estimation du temps de transition par BMA

n $\times 10^3$	p BMA	$\mathbb{E}(T_1 + T_2)$ MC	$\mathbb{E}(T_1 + T_3)$ BMA	$\mathbb{E}(T)$ BMA	<i>I.C sur</i> $\mathbb{E}(T)$
2	5.37×10^{-8}	$2.694 \cdot 10^{-1}$	14.72	5.02×10^6	$[4.10, 6.46] \times 10^6$
10	4.96×10^{-8}	$2.694 \cdot 10^{-1}$	15.94	5.44×10^6	$[4.68, 6.48] \times 10^6$
50	4.79×10^{-8}	$2.696 \cdot 10^{-1}$	15.75	5.63×10^6	$[5.35, 5.94] \times 10^6$
100	5.03×10^{-8}	$2.698 \cdot 10^{-1}$	14.71	5.36×10^6	$[5.22, 5.50] \times 10^6$

Conclusion

- ⊗ La version de Branchement Multi-niveaux Classique est apparait dans la littérature scientifique comme une technique prometteuse pour estimer un événement rare mais elle a montré ces limite dans la pratique.
- ⊗ 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.
- ⊗ 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 entrainer 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.

Questions / Réponses