UNIVERSITÉ PAUL SABATIER TOULOUSE III

U.F.R. MATHEMATIQUE INFORMATIQUE GESTION

THÈSE

POUR OBTENIR LE GRADE DE

DOCTEUR DE L'UNIVERSITÉ TOULOUSE III

Discipline : Mathématiques
Spécialité : Probabilités

présentée par

Agnès Lagnoux Renaudie

Analyse des modèles de branchement avec duplication des trajectoires pour l'étude des événements rares

---

Directeurs de thèse : Dominique BAKRY et Pascal LEZAUD

Soutenue le 6 décembre 2006 devant le jury composé de Messieurs :

Dominique BAKRY	Université Toulouse III	Directeur
Pierre DEL MORAL	Université Nice Sophia Antipolis	Rapporteur
François LEGLAND	IRISA Rennes	Rapporteur
Pascal LEZAUD	DTI/SDER Toulouse	Directeur
Manuel SAMUELIDES	SUPAERO Toulouse	Examinateur
José VILLÉN ALTAMIRANO	Escuela Universitaria de Informática Madrid	Examinateur

Laboratoire de Statistique et Probabilités
UMR CNRS 5583, Université Paul Sabatier, Toulouse III

REMERCIEMENTS

Je tiens en premier lieu à remercier Dominique Bakry pour m'avoir donné le goût des probabilités lors de son cours de maîtrise et d'agrégation. Son enthousiasme communicatif pour la recherche et son sens de la persuasion (!) m'ont convaincue. Sans lui, il n'y aurait eu ni DEA ni thèse. Je lui en suis profondément reconnaissante. Je le remercie tout particulièrement pour la confiance, l'écoute et la patience qu'il m'a témoignée pendant ces trois années de thèse. Ses profondes connaissances et intuitions ainsi que son humour ont rendu ce travail fort stimulant.

Je tiens également à remercier Pascal Lezaud pour m'avoir proposé ce sujet qui s'est avéré être passionant et tout à fait adapté à ce que j'avais envie de faire. Sa très grande disponibilité, ses conseils et encouragements m'ont été très précieux, notamment dans les moments les plus difficiles. J'ai aussi apprécié sa grande culture lors de nos discussions mathématiques, cinéphiles ou encore littéraires échangées pendant nos missions communes. Je l'en remercie et essaie ardemment de me mettre à son niveau !

Je suis très honorée que Pierre Del Moral et François LeGland aient accepté d'être rapporteurs de ma thèse. Je les remercie pour leur lecture attentive du manuscrit. C'est également un grand honneur que José Villén-Altamirano fasse partie de mon jury. Qu'ils soient tous les trois remerciés d'avoir fait le déplacement à Toulouse pour ma soutenance. Un grand merci aussi à Manuel Samuelides qui m'a fait le plaisir de participer à mon jury.

Merci à Djalil Chafai pour sa bonne humeur et ses encouragements. Je le remercie aussi pour ses bons conseils en informatique ainsi que Sébastien Dejean pour sa disponibilité et son aide pour résoudre mes problèmes de simulation.

Je souhaite saluer les doctorants du LSP, avec qui j'ai passé des moments très agréables, en particulier celles et ceux avec qui j'ai eu le plaisir de partager le bureau 148. Un grand merci pour leur écoute, leurs nombreux conseils et leur entousiasme...

Je tiens aussi à remercier tous mes amis, toulousains ou bien d'ailleurs, pour leur sympathie, leurs attentions dans le *vrai* monde, loin des mathématiques et des soucis de la thèse...

Un mot enfin pour exprimer ma gratitude envers mes parents et mon frère pour leur affection, leur confiance et leurs nombreux encouragements tout au long de mes études et en particulier durant la thèse. Ils ont toujours été là quand j'ai eu besoin d'eux.

Je terminerai avec toi, Eric, qui, jamais les soirs de découragement, n'a hésité à m'écouter et à essayer de se plonger dans mes problèmes probabilistes pour m'aider à trouver une solution ! Je suis heureuse d'avoir partagé cette aventure enrichissante et empreinte de bonheur avec toi. Merci pour ta patience, tes encouragements quotidiens et tout le reste bien sûr...

TABLE DES MATIÈRES

Remerciements 3

1. Introduction et présentation de la thèse 9

1.1. Introduction et motivations 9

1.2. Présentation de la thèse 10

1.3. Plan de la thèse 14

2. Méthodes pour l'étude des événements rares 17

2.1. Théorie des valeurs extrêmes 17

2.2. La simulation des événements rares 23

2.3. Le modèle RESTART 34

2.4. Les algorithmes particulaires 44

3. Modèle de branchement avec duplication des trajectoires 57

3.1. Présentation du modèle 57

3.2. Estimateur de $\mathbb{P}(A)$ 59

3.3. Le modèle de Galton-Watson 61

3.4. Optimisation de l'algorithme 64

3.5. Applications 72

3.6. La fonction génératrice de Z_n 80

3.7. Probabilité d'extinction 88

3.8. Résultats asymptotiques 90

3.9. Lien avec le modèle d'échantillonnage préférentiel 93

3.10. Conclusion 95

4. Transformée de Laplace et intervalles de confiance 97

4.1. Introduction 97

4.2. Premier modèle : déterministe 102

4.3. Second modèle : générer à chaque succès 118

4.4. Troisième modèle : générer un environnement aléatoire 122

4.5. Illustration numérique 137

4.6. Conclusion 139

4.7. Compléments 140

5. Modèles d'apprentissage 143

5.1. Modèle de branchement avec duplication des trajectoires à deux étapes 143

5.2. Modèle Aldous-Vazirani 170

5.3. Un modèle qui garde le système particulaire en vie 178
5.4. Un modèle où les niveaux sont placés durant la simulation 181
5.5. Comparaison des différents modèles 183
5.6. Conclusion 186

6. Application au processus d'Ornstein-Uhlenbeck «modifié» .. 187
6.1. Etude analytique du processus d'Ornstein-Uhlenbeck général 188
6.2. Application au processus d'Ornstein-Uhlenbeck général 195
6.3. Etude analytique du processus d'Ornstein-Uhlenbeck «modifié» .. 198
6.4. Application au processus d'Ornstein-Uhlenbeck «modifié» 199
6.5. Conclusion 202

7. Conclusion 203
Modèle de branchement avec duplication des trajectoires 203
Généralisation à des systèmes multidimensionnels 206

Bibliographie 209

Table des figures 215

Annexe 1 : Preuves-Modèle RESTART (CHAP.2) 217

Annexe 2 : Preuves-Systèmes particulaires (CHAP.2) 225

Annexe 3 : Algorithme MCMC 233

CHAPITRE 1

INTRODUCTION ET PRÉSENTATION DE LA THÈSE

1.1. Introduction et motivations

Dans de nombreux domaines appliqués, l'estimation de la probabilité d'occurrence d'un événement rare est importante, souvent en raison du risque associé à cet événement ; cet événement pouvant être une défaillance catastrophique en fiabilité des systèmes, un risque de collision entre avions, de tremblement de terre... Les événements rares sont caractérisés par des probabilités de l'ordre de 10^{-9} à 10^{-12}. Par exemple, en télécommunication, la probabilité de perte d'un paquet d'informations est inférieure à 10^{-9}.

Diverses voies se présentent pour l'étude de ces risques :
- l'analyse statistique des événements extrêmes qui s'appuie principalement sur les lois asymptotiques extrêmes (Weibull, Frêchet, Gumbel) mais nécessite une longue période d'observation, voir [2] et Section 2.1 pour plus de détails,
- la modélisation qui conduit à estimer la probabilité de l'événement rare, soit par une approche analytique, voir e.g. [53] et Section 2.2.2.2, soit par la simulation, voir Section 2.2.

Dans cette thèse, nous aborderons l'aspect simulation basée sur la méthode de Monte Carlo qui s'appuie sur la loi forte des grands nombres. Cependant, cette méthode de simulation s'avère inefficace pour estimer un événement rare de probabilité 10^{-9} à 10^{-12}, en raison du nombre trop important d'échantillons à générer. Pour résoudre ce problème, de nombreuses méthodes de simulation accélérée ont été proposées. En particulier, une première méthode est un raffinement de Monte Carlo : l'échantillonnage préférentiel. Manuel & José Villén-Altamirano ont développé une autre méthode de simulation [59, 60], dénommée RESTART, qui permet de diminuer de façon drastique la durée de simulation. Une autre approche [18, 20], basée sur des systèmes de particules, permet d'obtenir des résultats théoriques sur la convergence de ce type d'algorithme. Pour une description plus détaillée de ces algorithmes, le lecteur est renvoyé au Chapitre 2.

Les algorithmes particulaires et la méthode RESTART appartiennent à la même classe de modèles : ceux utilisant des multi-trajectoires préférentielles afin d'augmenter le nombre de visites à l'événement rare et s'appuyant sur l'hypothèse qu'il existe des états intermédiaires identifiables visités par la trajectoire, plus souvent que l'événement rare lui-même.

Nous nous intéressons dans cette thèse à l'efficacité réelle des modèles à multi-trajectoires préférentielles en terme de coût de simulation. En raison de leur complexité, l'analyse mathématique directe de tels modèles s'avère impraticable. Cependant, une étude analytique rapide montre que ces modèles reposent tous sur un même «squelette», sorte de modèle simplifié. Nous sommes donc conduits à introduire ce modèle simplifié pour lequel nous pourrons mener une étude précise et obtenir des résultats exacts et non asymptotiques. Cette étude nous permettra, de plus, de dégager une suggestion de procédure à appliquer dans le cadre d'algorithmes plus complexes pour lesquels l'analyse exacte n'a pu être faite. Nous montrerons aussi que ce modèle, malgré sa simplicité relative, s'applique à une grande classe d'exemples et qu'il s'adapte facilement à des applications plus complexes.

1.2. Présentation de la thèse

Comme expliqué dans le paragraphe précédent, nous nous intéressons à un modèle appelé *modèle de branchement avec duplication des trajectoires* (*branching splitting model* en anglais), inspiré des modèles à trajectoires préférentielles. Nous nous plaçons dans le cadre suivant où nous considérons une suite décroissante de $M+1$ ensembles emboîtés B_i menant à l'événement rare A :

$$A = B_{M+1} \subset B_M \subset \cdots \subset B_1.$$

Alors la probabilité d'intérêt $\mathbb{P}(A)$ est donnée par

$$\mathbb{P}(A) = \mathbb{P}(A|B_M)\mathbb{P}(B_M|B_{M-1}) \cdots \mathbb{P}(B_2|B_1)\mathbb{P}(B_1), \tag{1}$$

où chaque probabilité conditionnelle du membre de droite est «moins rare» que l'événement A lui-même. Pour les applications que nous avons en tête, ces probabilités conditionnelles ne sont pas connues explicitement. Par contre, nous savons faire évoluer les particules du niveau B_i vers le niveau suivant B_{i+1} (par ex. comportement markovien).

Dans ce modèle, une apparition plus fréquente de A est réalisée en dupliquant le processus en R_i sous-processus dès qu'il entre dans une région B_i où sa chance d'atteindre l'événement rare est plus grande. On privilégie ainsi les trajectoires favorables.

Nous souhaitons tout d'abord déterminer les paramètres optimaux de l'algorithme (Chapitre 3) : probabilités de transition P_i (correspondant à la position des niveaux), nombres de retirage R_i, nombre N de particules générées au départ, nombre de niveaux M. En considérant que cet algorithme peut être représenté par N processus de branchement Z_{M+1} de Galton-Watson indépendants, nous pouvons exprimer la variance de l'estimateur, noté $\widehat{P}$, en fonction des paramètres de l'algorithme. Guidés par l'heuristique présentée dans [60, 61], nous déduisons un algorithme optimal en minimisant la variance de l'estimateur pour un budget fixé (coût de la simulation), défini comme le nombre moyen de particules générées lors de la simulation, où chaque particule est pondérée par une fonction de coût dépendant du niveau de départ de la particule.

L'optimisation de l'algorithme suggère de prendre toutes les probabilités de transition égales (nous noterons cette valeur commune P_0) et les nombres de retirage égaux à l'inverse de cette constante. Ce résultat n'est pas surprenant puisqu'il signifie que les processus de branchement considérés sont des processus de Galton-Watson critiques. En d'autres termes, les valeurs optimales sont choisies de telle façon à avoir un certain équilibre entre une réduction de la variance due à des nombres de retirage trop faibles et une croissance exponentielle du coût de la simulation due à des nombres de retirage trop élevés. Voir le Chapitre 3 pour les détails.

Nous nous intéressons ensuite (Chapitre 4) à la précision de l'algorithme en dérivant une borne supérieure à la quantité

$$\mathbb{P}(|\widehat{P} - \mathbb{P}(A)|/\mathbb{P}(A) \geq \alpha) \tag{2}$$

La borne de Chebycheff étant trop grossière, nous utilisons une borne de Chernoff basée sur la transformée de Laplace du processus renormalisé

$$W_{M+1} := Z_{M+1}/\mathbb{E}(Z_{M+1}).$$

Ainsi, nous avons besoin d'estimées de la transformée de Laplace de W_{M+1} dépendant de l'itérée n-ième d'une fonction notée ψ, dont nous ne pouvons établir une expression explicite. En pratique, le nombre optimal M de niveaux n'est pas très grand et par conséquent, des estimées asymptotiques ne seraient pas assez précises. Ce nombre optimal étant dérivé justement à partir de ces bornes, une approche numérique s'avère impraticable aussi. C'est pourquoi, nous cherchons des bornes supérieures et inférieures explicites pour un nombre de niveaux M donné. Remarquons que la démarche consistant à optimiser la variance de l'estimateur à coût fixé et ensuite à étudier des intervalles de confiance faisant intervenir la transformée de Laplace de ce même estimateur est validée par le fait qu'asymptotiquement minimiser la variance de l'estimateur à coût fixé revient à minimiser la transformée de Laplace de l'estimateur à coût fixé.

Afin d'obtenir des bornes de l'itérée $M+1$-ième de la fonction ψ, nous procédons de la façon suivante : plutôt que de borner ψ grâce à une seule fonction, nous en utilisons plusieurs pour obtenir des résultats plus précis. Nous choisissons ces fonctions «bornes» dans les groupes de Lie de faible dimension des fonctions affines et des homographies, fonctions monotones dont les itérées n-ièmes peuvent être établies explicitement.

En pratique, l'ajustement des probabilités de transition proches des valeurs optimales peut être fait lors d'une phase d'apprentissage. La proportion du coût consacrée à cette première phase est donnée au Chapitre 5. Mais il apparaît rapidement que même dans le cas où les probabilités de transition sont proches des valeurs optimales, le fait que le nombre de retirage n'est pas nécessairement un entier détruit rapidement la précision de l'algorithme : lorsque la valeur optimale du nombre de retirage n'est pas un entier, un choix naturel est de prendre l'entier le plus proche. Mais quelle que soit la stratégie choisie, la criticalité du processus de Galton-Watson sous-jacent est perdue et la perte de précision conséquente est significative.

Le Chapitre 4 est donc consacré à l'étude de différentes stratégies visant à surmonter ce problème. Guidés par [3], nous introduisons deux algorithmes qui choisissent aléatoirement les nombres de retirage. Plus précisément, posons $k = \lfloor 1/P_0 \rfloor$, P_0 représentant la valeur commune des P_i. Au lieu d'utiliser une unique valeur déterministe pour chaque nombre de retirage, nous considérons les R_i comme des variables aléatoires i.i.d. de Bernoulli sur $\{k, k+1\}$ de paramètre $p = \mathbb{P}(R_1 = k)$. Dans un premier modèle, nous générons une variable aléatoire de Bernoulli R chaque fois qu'une particule, partant du niveau i-1, atteint le niveau i. Un second modèle, consiste à générer un environnement aléatoire décrit par la suite $(R_1, R_2, .., R_M)$ de M variables aléatoires de Bernoulli i.i.d. R_i sur $\{k, k+1\}$ de paramètre p. Le paramètre p est déduit dans chacun des modèles par la même approche d'optimisation que précédemment. Pour le second modèle, le problème est plus complexe et ne permet d'obtenir qu'une solution asymptotique.

Comme nous l'avons déjà mentionné auparavant, en pratique, les probabilités de transition sont généralement inconnues. Il n'est donc pas possible de déduire les nombres de retirage optimaux. Rappelons aussi que, pour les applications que nous avons en tête, bien que ces probabilités conditionnelles ne soient pas connues explicitement, nous savons faire évoluer les particules du niveau B_i vers le niveau suivant B_{i+1}. Supposons, de plus, que les probabilités de transition P_i sont dans un compact $[a, b]$, $0 < a < b < 1$. Une telle hypothèse est essentielle : sinon, rien ne peut être fait algorithmiquement. En pratique, cette supposition est généralement implicite mais nul ne la mentionne.

Pour résoudre ce problème, nous proposons, dans le Chapitre 5, un algorithme en deux étapes : la première est une phase d'apprentissage pour laquelle on génère ρ_N particules. L'algorithme suit la procédure du modèle de branchement avec duplication des trajectoires classique avec des nombres de retirage choisis arbitrairement au départ. Durant la seconde phase, nous générons $N - \rho_N$ particules qui évoluent comme dans la première phase mais ici, nous utilisons des estimateurs des nombres de retirage optimaux obtenus durant la phase d'apprentissage et suivant le schéma optimal proposé au Chapitre 3. Compte tenu de la complexité des formules, nous menons seulement une étude asymptotique lorsque le coût C tend vers l'infini. Le fait de supposer que les probabilités de transition sont bornées inférieurement et supérieurement implique que le coût par particule est borné et l'étude se ramène donc à une étude asymptotique lorsque N, le nombre total de particules générées, tend vers l'infini. Une analyse précise montre que nous devons consacrer asymptotiquement $\mu_s C^{2/3}$ particules à la phase d'apprentissage contre $C/C_{opt} - \mu_s C^{2/3}$ à la seconde phase, où C_{opt} est une constante définie au Chapitre 5 et μ_s se déduit de l'optimisation de l'algorithme et peut être exprimé explicitement ; c-à-d que si on suppose que le nombre ρ_N de particules générées dans la phase d'apprentissage a asymptotiquement la forme $\mu_\alpha(C)C^{1-\alpha}$, on doit prendre $\alpha = 1/3$. Remarquons que, N étant linéaire en C, consacrer $\mu_s C^{2/3}$ particules à la première phase revient à lui consacrer $\lambda_s N^{2/3}$ particules, pour un certain λ_s dépendant de μ_s.

Ce résultat provient directement d'une annulation observée pendant l'optimisation qui est elle-même une conséquence du choix particulier des R_i, solutions d'un problème d'optimisation et conduit à prendre $\alpha = 1/3$. Si nous n'avions pas choisi les R_i selon le critère optimal proposé au Chapitre 3, il n'y aurait pas eu l'annulation précédemment mentionnée et le α optimal aurait simplement été égal à $1/2$, comme l'on peut s'y attendre a priori. Ceci souligne l'importance du choix des probabilités de transition proches des valeurs optimales et des nombres de retirage selon le critère optimal déjà mentionné. Ce résultat insiste aussi sur l'intérêt de mettre en place un algorithme adaptatif qui évaluerait à chaque étape les probabilités de transition et réajusterait les niveaux avant chaque phase.

1.3. Plan de la thèse

Le chapitre suivant est consacré à la présentation des différentes techniques utilisées pour étudier les événements rares : l'étude analytique (pages 17 à 23) puis la modélisation basée sur des méthodes accélérées de Monte-Carlo des pages 23 à 34. La fin du Chapitre 2 est consacrée à la présentation de deux modèles à trajectoires préférentielles : le modèle RESTART (pages 34 à 44) et les systèmes particulaires (à partir de la page 44).

Dans le Chapitre 3, nous présentons d'abord le modèle de branchement avec duplication des trajectoires (pages 57 à 64) puis nous dérivons les paramètres optimaux de l'algorithme par minimisation de la variance de l'estimateur sous une contrainte de coût (pages 64 à 72). De nombreuses applications ainsi qu'une étude détaillée de la transformée de Laplace de l'estimateur sont aussi présentées dans ce chapitre.

Le Chapitre 4 est consacré à l'étude de la transformée de Laplace de l'estimateur de la probabilité de l'événement rare afin de dégager des intervalles de confiance précis (pages 97 à 118) puis à la comparaison de deux stratégies à adopter lorsque les nombres de retirage ne sont pas entiers.

Dans le Chapitre 5, nous présentons et étudions (pages 143 à 170) un modèle à deux étapes qui permet de contourner un des problèmes inhérents au modèle de branchement avec duplication des trajectoires : le choix optimal des niveaux. Puis, (à partir de la page 170), nous présentons et comparons au modèle à deux étapes d'autres modèles existant déjà.

Nous illustrons ce modèle, dans le Chapitre 6, avec l'étude d'excursions rares d'un processus d'Ornstein-Uhlenbeck «modifié» et comparons les résultats obtenus à ceux donnés par les modèles présentés dans le chapitre précédent.

Le dernier chapitre conclut ce travail et dégage quelques pistes à étudier par la suite.

CHAPITRE 2

MÉTHODES POUR L'ÉTUDE DES ÉVÉNEMENTS RARES

Nous présentons dans ce chapitre différentes techniques visant à étudier les événements rares. Quelques pistes par l'étude analytique sont données en Section 2.1 et dans la Section 2.2, nous dressons un rapide panorama des méthodes basées sur la modélisation. Dans les deux dernières sections, nous présentons plus en détail le modèle RESTART et les algorithmes particulaires, méthodes reposant directement sur le modèle de branchement avec duplication des trajectoires étudié dans cette thèse.

2.1. Théorie des valeurs extrêmes

Nous évoquons rapidement dans cette section les principaux résultats sur la théorie des valeurs extrêmes présentés dans [2].

2.1.1. Premiers résultats dans $\mathbb{R}$. —

2.1.1.1. Motivation. — Etant donnée une diffusion sur $\mathbb{R}$

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t,$$

on s'intéresse à l'étude de la variable aléatoire $M_t = \sup_{s\in[0,t]} X_s$, et plus précisément, on souhaite estimer, pour b grand, $\mathbb{P}(M_t \leq b)$.
Introduisons le temps aléatoire

$$T_b = \inf\{t : X_t \geq b\},$$

alors $\mathbb{P}(M_t \leq b) = \mathbb{P}(T_b \geq t)$.

Choisissons b suffisamment grand pour que $\mathbb{P}(M_t \leq b)$ soit très petit (événement rare) et introduisons l'ensemble aléatoire dans $\mathbb{R}^+$

$$S_b = \{t : X_t \geq b\}.$$

Cet ensemble correspond aux excursions du processus X_t au-dessus du niveau b. On a alors équivalence entre les trois événements suivants :

$$\{M_t \leq b\}, \quad \{T_b \geq t\}, \quad S_b \cap [0, t] = \emptyset.$$

Par conséquent, si nous sommes en mesure de déterminer la probabilité du dernier événement, nous aurons obtenu la probabilité des deux premiers. Cette démarche est celle développée dans [2].

2.1.1.2. Premier exemple : Cas des variables aléatoires i.i.d.— Soit $(X_i)_{i\geq 1}$ une suite de v.a.r. i.i.d. de fonction de répartition commune F. Soit $M_n = \sup_{1\leq i\leq n} X_i$, alors

$$\mathbb{P}(M_n \leq x) = F^n(x) \underset{x \text{ grand}}{\simeq} \exp(-n\bar{F}(x)),$$

où $\bar{F}(x) = 1 - F(x)$.

Définition 2.1.1. — *S'il existe des constantes $a_n > 0$ et b_n telles que*

$$\lim_{n\to\infty} \mathbb{P}\big(a_n(M_n - b_n) \leq x\big) = \lim_{n\to\infty} F^n(a_n^{-1}x + b_n) = G(x),$$

on dira que M_n appartient au domaine d'attraction *de G que l'on notera par $M_n \in \mathcal{D}(G)$.*

Si $G(x)$ existe et est non dégéneré, on peut montrer que G est de l'une des trois formes suivantes (où x peut être remplacé par $ax + b$)

$$\begin{cases} \text{type I} & G(x) = \exp(-e^{-x}), \qquad x \in \mathbb{R} \\ \text{type II} & G(x) = \exp(-x^{-\alpha}), \qquad x > 0, \alpha > 0 \\ \text{type III} & G(x) = \exp(-(-x)^{-\alpha}), \qquad x \leq 0, \alpha < 0. \end{cases}$$

On démontre le lemme suivant :

Lemme 2.1.1. — *Soit $(u_n)_{n\geq 1}$ une suite de constantes et $0 \leq \tau < \infty$. Alors*

$$\lim_{n\to\infty} \mathbb{P}(M_n \leq u_n) = e^{-\tau} \Leftrightarrow \lim_{n\to\infty} n\bar{F}(u_n) = \tau$$

En particulier, si $M_n \in \mathcal{D}(G)$, alors $u_n = a_n^{-1}x + b_n$ et $\tau = -\log G(x)$.

Exemples 2.1.1. — 1. *Soit* (X_n) *une suite de v.a.r. i.i.d. uniformes sur* $[-1, 0]$. *Alors, on vérifie aisément que*

$$nM_n \xrightarrow[\mathcal{D}]{} \xi,$$

où ξ *est de type III avec* $\alpha = 1$. *Ainsi,* $a_n = n$ *et* $b_n = 0$.

2. *Soit* (X_n) *une suite de v.a.r. i.i.d. telle que* $\bar{F}(x) \simeq Ax^{-\alpha}$ *quand* $x \to \infty$. *Alors* $(An)^{-1/\alpha}M_n \xrightarrow[\mathcal{D}]{} \xi$, *où* ξ *est de type II. Ainsi,* $a_n = (An)^{-1/\alpha}$ *et* $b_n = 0$.
3. *Soit* (X_n) *une suite de v.a.r. i.i.d. telle que* $\mathbb{P}(X_1 > x) \simeq Aexp(-bx)$. *Alors* $b(M_n - \frac{1}{b}log(An)) \xrightarrow[\mathcal{D}]{} \xi$, *où* ξ *est de type I. Ainsi,* $a_n = b$ *et* $b_n = \frac{1}{b}log(An)$.

2.1.1.3. Deuxième exemple : Cas des variables aléatoires stationnaires. — Soit (X_i) une suite de v.a.r. stationnaires. Posons

$$F_{i_1,\cdots,i_n}(u) = \mathbb{P}(X_{i_1} \le u, \cdots, X_{i_n} \le u)$$

et

$$\alpha_{n,l} = \max_{1\le i_1<\cdots<i_p<j_1<\cdots<j_q\le n, j_1-i_p\le l} \left\{ |F_{i_1,\cdots,i_p,j_1,\cdots,j_q}(u_n) - F_{i_1,\cdots,i_p}(u_n)F_{j_1,\cdots,j_q}(u_n)| \right\}.$$

On dira que la condition $D(u_n)$ est vérifiée par la suite (X_i) si pour une certaine suite (l_n) telle que $l_n = o(n)$, on a $\lim_{n\to\infty} \alpha_{n,l_n} = 0$. Cette propriété traduit une certaine indépendance asymptotique.

On obtient alors le lemme suivant :

Lemme 2.1.2. — *Soit* (u_n) *une suite de constantes telle que pour tout* n, *la condition* $D(u_n)$ *soit vérifiée. Soit* (k_n) *une suite telle que* $k_n = o(n)$, $k_nl_n = o(n)$ *et* $k_n\alpha_{n,l_n} \to 0$. *Alors*

$$\mathbb{P}(M_n \le u_n) - \mathbb{P}^{k_n}(M_{r_n} \le u_n) \to 0, \quad \textit{quand } n \to \infty,$$

où $r_n = [n/k_n]$.

Ce lemme s'interprète de la façon suivante : on peut diviser $\{1, \cdots, n\}$ en $k_n = o(n)$ intervalles de longueur r_n et ne pas tenir compte des intervalles de longueur l_n situés à droite des précédents (leur longueur tendant vers 0), de telle manière que pour $n \to \infty$, M_n est approché par le plus grand des extrema sur chaque intervalle restant ; ces extrema ayant une certaine dépendance mesurée par α_{n,l_n}.

Dans le cas où M_n appartient au domaine d'attraction de G et vérifie la condition $D(a_n^{-1}x + b_n)$, avec x tel que $0 < G(x) < 1$, alors on en déduit que $G(x)$ est de l'un des 3 types précités.

2.1.2. Premiers résultats dans $\mathbb{R}^2$. —

2.1.2.1. Motivation et méthode. — Soit $(X_t, t \in \mathbb{R}^2)$ un processus stationnaire. On s'intéresse à la v.a.r.

$$M_T = \sup_{t\in[0,T]^2} X_t, \qquad T \text{ grand.}$$

Nous voulons estimer pour b grand $\mathbb{P}(M_T < b)$. Nous allons procéder de la même façon que dans le paragraphe 2.1.1.1 et écrire l'événement $M_T < b$ sous une autre forme. Posons $S_b = \{t : X_t \geq b\}$, alors

$$\mathbb{P}(M_T < b) = \mathbb{P}(S_b \cap [0,T]^2 = \emptyset).$$

Par conséquent, si nous sommes en mesure de déterminer la probabilité du deuxième événement, nous aurons obtenu celle du premier.

Introduisons maintenant les processus de mosaïque, outils qui nous permettront d'étudier l'événement $S_b \cap [0,T]^2 = \emptyset$.

2.1.2.2. Processus de mosaïque sur $\mathbb{R}^2$. — Soit $\mathcal{C}$ un sous-ensemble de $\mathbb{R}^2$ tiré au hasard parmi une classe de sous-ensembles $B_1, \cdots, B_k$ avec probabilités $p_1, \cdots, p_k$. On peut imaginer que les valeurs les plus probables de $\mathcal{C}$ sont des ensembles localisés près de 0.

Posons $C = \text{aire}(\mathcal{C})$, ainsi C est une v.a.r. On supposera en outre que $\mathbb{P}(C > 0) = 1$ et $\mathbb{E}(C) < \infty$. Etant donné un sous-ensemble B et $x \in \mathbb{R}^2$, on notera

$$x + B = \{x + y : y \in B\}$$

<u>Rappel</u> : Construction d'un processus ponctuel de Poisson d'intensité $\mu = K * m$ où m est la mesure de Lebesgue restreinte à $E \subset \mathbb{R}^2$ [15].

- on choisit $n \in \mathbb{N}$ (nombre de points) suivant une loi de Poisson de paramètre $\lambda = Km(E) = \mu(E)$.
- on choisit aléatoirement n points y de E suivant la loi $\nu = \mu/\mu(E)$ (loi uniforme sur E).

Soit λ une constante strictement positive, on définit le *processus de mosaïque* S de façon suivante :

1. on génère une suite de points $y \in \mathbb{R}^2$ selon un processus ponctuel de Poisson de taux λ par unité de surface,
2. puis pour chaque y, on choisit un C_y distribué comme $\mathcal{C}$, indépendamment de y,
3. enfin on pose $S = \bigcup(y + C_y)$.

On dira que y est un *centre* et $y + C_y$ un *paquet*. Le processus S est donc la réunion de paquets i.i.d. de centres distribués selon un processus de Poisson. λ est appelé le *taux de paquets*.

Lemme 2.1.3. — *Pour tout $x \in \mathbb{R}^2$, soit N_x le nombre de paquets du processus S contenant x. Alors N_x suit une loi de Poisson de paramètre $\lambda E(C)$.*

Démonstration. — On a $N_x = \sum_{i=1}^{M(\omega)} 1_{x \in y+C_y}(\omega) = \sum_{i=1}^{M(\omega)} X_i(\omega)$.
Notons f_Y et ϕ_Y les fonctions génératrice et caractéristique de la variable Y. Alors,

$$\begin{aligned}
\phi_{N_x}(t) &= \mathbb{E}(e^{itN_x}) = \int \exp(it \sum_{i=1}^{M(\omega)} X_i(\omega)) dP(\omega) \\
&= \sum_{n \in \mathbb{N}} \int 1_{M=n} \exp(it \sum_{i=1}^{M(\omega)} X_i(\omega)) dP(\omega) \\
&= \sum_{n \in \mathbb{N}} \mathbb{E}(1_{M=n} \exp(it \sum_{i=1}^{n} X_i)) = \sum_{n \in \mathbb{N}} \mathbb{E}(1_{M=n}) \mathbb{E}(\exp(it \sum_{i=1}^{n} X_i)) \\
&= \sum_{n \in \mathbb{N}} \mathbb{P}(M = n) \mathbb{E}(\exp(it \sum_{i=1}^{n} X_i)) = \sum_{n \in \mathbb{N}} \mathbb{P}(M = n) \mathbb{E} \phi_X(t)^n \\
&= f_M(\phi_X(t)) = e^{\lambda(\phi_X(t)-1)}
\end{aligned}$$

puisque M suit une loi de poisson. Or $\phi_X(t) = (e^{it} - 1)\mathbb{E}(C) + 1$ donc

$$\phi_{N_x}(t) = e^{\lambda \mathbb{E}(C)(e^{it}-1)}$$

□

Définition 2.1.2. — *On dira que le processus de mosaïque S est* clairsemé *s'il couvre une faible proportion du plan $\mathbb{R}^2$ c-à-d si la probabilité $p := \mathbb{P}(x \in S)$ à x fixé est très petit.*

Dans ce cas, on a

$$p = \mathbb{P}(N_x \geq 1) = 1 - \exp\{-\lambda \mathbb{E}(C)\} = \lambda \mathbb{E}(C) + O(p^2).$$

D'autre part, bien que les paquets peuvent se recouvrir, dans le cas d'un processus clairsemé, on a $\mathbb{P}(N_x \geq 2 | N_x \geq 1) = O(p)$, et il est donc possibe de négliger ce recouvrement en première approximation.

Soit A un sous-ensemble fixé dans $\mathbb{R}^2$ tel que son intérieur ne soit pas localisé près de sa frontière. Soit S un processus de mosaïque clairsemé. On peut alors approximer $S \cap A$ par l'union des paquets de S dont le centre est dans A. Ainsi

$$\begin{aligned} \text{aire}(S \cap A) &\underset{\text{en loi}}{\simeq} \sum_{i=1}^{M} C_i, \\ \mathbb{P}(S \cap A = \emptyset) &\simeq \exp(-\lambda \text{aire}(A)), \end{aligned}$$

où les C_i sont des copies i.i.d. de $\mathcal{C}$ et M est une v.a.r. de loi de Poisson de paramètre $\lambda\text{aire}(A)$, indépendante de $\mathcal{C}$.

2.1.2.3. Démarche. — Soit $(X_t, t \in \mathbb{R}^2)$ un processus stationnaire. On s'intéresse à la v.a.r.

$$M_T = \sup_{t \in [0,T]^2} X_t, \qquad T \text{ grand.}$$

Posons $S_b = \{t : X_t \geq b\}$, alors

$$\mathbb{P}(M_T < b) = \mathbb{P}(S_b \cap [0, T]^2 = \emptyset).$$

Pour b grand, S_b est un ensemble aléatoire clairsemé. Supposons alors que S_b puisse être approché par un processus de mosaïque clairsemé de taux λ_b pour une distribution des paquets $\mathcal{C}_b$. Alors,

$$\mathbb{P}(S_b \cap [0, T]^2 = \emptyset) \simeq \exp(-\lambda_b T^2).$$

Supposons en outre que nous connaissions la distribution marginale de X_t, et notons par

$$p_b := \mathbb{P}(X_t \geq b) = \mathbb{P}(x \in S_b).$$

Alors, $p_b \simeq \lambda_b \mathbb{E}(C_b)$, d'où

$$\mathbb{P}(M_T < b) \simeq \exp\left(-\frac{p_b T^2}{\mathbb{E}(C_b)}\right).$$

Enfin, pour obtenir la valeur de $\mathbb{P}(M_T < b)$, il nous faut pouvoir soit estimer $\mathbb{E}(C_b)$, soit λ_b. Différentes heuristiques sont présentées dans [2] pour estimer λ_b.

2.2. La simulation des événements rares

Nous présentons, dans cette section, différentes méthodes de simulation des événements rares. Tout d'abord, nous rappelons la méthode de Monte Carlo, base de la simulation des événements rares, puis nous présentons différentes techniques d'accélération ainsi qu'une approche analytique.

2.2.1. La méthode de Monte Carlo.

2.2.1. La méthode de Monte Carlo. — Cette méthode de calcul approché d'intégrales repose sur l'interprétation de l'intégrale comme l'espérance d'une certaine variable aléatoire.

2.2.1.1. Principe. — On cherche à calculer

$$I = \int_{[0,1]^d} g(u_1, \dots, u_d) du_1 \dots du_d.$$

On remarque que si $U_1, \dots, U_d$ sont des v.a. i.i.d. uniformes sur $[0, 1]$, on a $I = \mathbb{E}(g(U_1, \dots, U_d))$.

Soit alors (X_n) une suite de v.a. i.i.d. de même loi que $g(U_1, \dots, U_d)$. Par la loi forte des grands nombres, si $\mathbb{E}(|X_1|) < \infty$,

$$\frac{1}{n}(X_1 + \cdots + X_n) \underset{n\to\infty}{\longrightarrow} \mathbb{E}(f(U_1, \dots, U_d)) = I$$

La moyenne empirique est donc, pour n grand, une «bonne» approximation de I.

On généralise aisément ce résultat au calcul d'une intégrale dans $\mathbb{R}^d$

$$I = \int_{\mathbb{R}^d} f(x_1, \dots, x_d) g(x_1, \dots, x_d) dx_1 \dots dx_d$$

où f est la densité d'une v.a. X par rapport à la mesure de Lebesgue sur R^d.

Si $g(X)$ est intégrable, $I = \mathbb{E}(g(X))$ sera approchée par une réalisation de

$$\widehat{I}_n := \frac{1}{n}(g(X_1) + \cdots + g(X_n)),$$

où les X_i sont i.i.d. de même loi que X.

2.2.1.2. Vitesse de convergence. — L'erreur (aléatoire) que l'on commet est

$$\epsilon_n(\omega) = \mathbb{E}(g(X)) - \frac{1}{n}\left[g(X_1(\omega)) + \cdots + g(X_n(\omega))\right] = I - \widehat{I}_n.$$

Si la variance σ^2 de $g(X)$ est définie, le théorème de la limite centrale (TCL) donne la convergence en loi suivante

$$\sqrt{n}\frac{I - \widehat{I}_n}{\sigma} \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0,1)$$

On peut donc obtenir un intervalle de confiance asymptotique pour $\mathbb{E}(g(X))$:

$$\mathbb{P}\Big(|I - \widehat{I}_n| \leq \epsilon I\Big) \simeq \mathbb{P}(|Z| \leq \frac{\sqrt{n}}{\sigma}\epsilon I),$$

avec $Z \sim \mathcal{N}(0,1)$. Pour $\mathbb{P}(|Z| \leq \frac{\sqrt{n}}{\sigma}\epsilon I) = 0.95$,

$$\frac{\sqrt{n}}{\sigma}\epsilon I \approx 2 \quad \text{c-à-d} \quad n \approx \frac{4\sigma^2}{\epsilon^2 I^2}$$

On obtient ainsi une approximation de

$$\mathbb{P}\Big(\Big[\widehat{I}_n - \frac{t\sigma}{\sqrt{n}}, \widehat{I}_n + \frac{t\sigma}{\sqrt{n}}\Big] \ni \mathbb{E}(g(X))\Big).$$

L'intervalle est d'autant plus petit (donc précis) que σ est petit, d'où l'intérêt de réduire la variance. Cependant, ce genre de simulation ordinaire se révèle impraticable pour estimer des probabilités de l'ordre de 10^{-9}. Pour en revenir à l'exemple précédent issu de la communication, il faudrait simuler un billion de paquets d'informations par paquet perdu. Donc pour estimer la probabilité d'un tel événement, la simulation d'au moins 100 billions de paquets serait nécessaire, ce qui prendrait des centaines de jours.

Pour éviter ces temps de calculs déraisonnables, on fait appel à des techniques d'accélération de la simulation, que nous présentons dans la section suivante.

2.2.2. Méthodes d'accélération. —

2.2.2.1. Echantillonnage préférentiel. — L'idée de base présentée dans [45] consiste à changer la dynamique de la simulation afin que l'événement rare se produise plus souvent. Ce qui est fait en changeant la loi sous-jacente du système étudié. Le but est alors de trouver un changement de probabilité qui, à précision de l'intervalle de confiance donnée, réduit le nombre de tirages, ou inversement. Pour cela, il faut avoir une bonne connaissance du système étudié et malgré tout, il se peut que l'échantillonnage préférentiel n'apporte pas d'amélioration, voire même entraîne une détérioration.

Plus précisément, soit X une variable aléatoire ayant pour densité f. Afin d'estimer

$$\mathbb{E}(g(X)) = \int g(x)f(x)dx,$$

on fait apparaître artificiellement une fonction $h > 0$ vérifiant $\int h(x)dx = 1$ et ainsi la quantité à estimer peut se réécrire sous la forme

$$\mathbb{E}(g(X)) = \int \frac{g(x)f(x)}{h(x)} h(x)dx = \mathbb{E}\left[\frac{g(Y)f(Y)}{h(Y)}\right],$$

Y étant une variable aléatoire ayant pour densité h. La quantité $L(x) := \frac{f(x)}{h(x)}$ est appelée *rapport de vraisemblance.*

Par la méthode de Monte Carlo, on obtient de la sorte une nouvelle approximation de $\mathbb{E}(g(X))$ par

$$\frac{1}{n}\sum_{i=1}^{n} \frac{g(Y_i)f(Y_i)}{h(Y_i)}.$$

Cette manipulation est intéressante si la variable aléatoire $Z = \frac{g(Y)f(Y)}{h(Y)}$ a une plus petite variance que $g(X)$ c-à-d si l'on réduit l'intervalle de confiance. Le but est donc de minimiser $\mathrm{var}(Z)$.

L'efficacité de cette méthode repose donc sur le bon changement de probabilité. La théorie des grandes déviations a permis de montrer des résultats sur certaines classes de systèmes relativement simples.

Remarque 2.2.1. — *Il est facile de calculer la variance de Z :*

$$var(Z) = \mathbb{E}(Z^2) - \mathbb{E}(Z)^2 = \int_{\mathbb{R}} \frac{g^2(y)f^2(y)}{h(y)}dy - \mathbb{E}(g(X))^2$$

On remarque que si $g > 0$, la fonction

$$h(x) = \frac{g(x)f(x)}{\mathbb{E}(g(X))}$$

annule la variance ! Mais ce résultat n'est pas exploitable puisque h dépend alors de $\mathbb{E}(g(X))$ qui est justement la quantité que l'on souhaite estimer ! Cependant, il permet de justifier la démarche suivante : dans la pratique, on choisit h aussi proche que possible de $|gf|$ puis on normalise afin d'obtenir une densité. Pour *éviter de calculer la constante de renormalisation, on peut utiliser l'estimateur suivant*

$$\widehat{g}_N = \frac{\sum_{i=1}^N g(y_i)f(y_i)/h(y_i)}{\sum_{i=1}^N f(y_i)/h(y_i)} = \sum_{i=1}^N g(y_i)\omega_i$$

où $y_1, \cdots, y_N$ *sont des variables aléatoires i.i.d. de densité commune* h *et les poids d'importance* $\omega_1, \cdots, \omega_N$ *donnés par*

$$\omega_i = \frac{f(y_i)/h(y_i)}{\sum_{i=1}^N f(y_i)/h(y_i)}.$$

Pour N *fixé, cet estimateur est biaisé mais asymptotiquement, sous de faibles conditions, la loi faible des grands nombres s'applique et assure que* $\widehat{g}_N$ *converge vers* $\mathbb{E}(g(X))$ *lorsque* $N \to \infty$. *Sous des conditions supplémentaires, le théorème central limite peut aussi être appliqué.*

Remarquons que $\widehat{g}_N$ *n'est rien d'autre que la fonction* $g(x)$ *intégrée par rapport à la mesure empirique*

$$\sum_{i=1}^N \omega_i \delta_{y_i}(dy)$$

où δ_a *est la mesure de Dirac en* a.

Dans le cadre de l'étude des événements rares, la quantité d'intérêt θ s'écrit

$$\theta = \mathbb{P}(X \in A) = \mathbb{E}(\mathbf{1}_A(X)) = \int_{-\infty}^{+\infty} \mathbf{1}_A(x) f(x) dx$$

et peut être estimée par la méthode de Monte Carlo à partir des échantillons $X_1, \cdots, X_N$ de densité commune f avec l'estimateur non biaisé suivant

$$\widehat{\theta}_N = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_A(X_i)$$

D'après la Section 2.2.1, la taille nécessaire d'échantillonnage pour garantir une erreur relative inférieure à ϵ avec une probabilité de $1-\alpha$ doit être

$$N = \left(\frac{q_{1-\alpha} \sigma(\widehat{\theta}_N)}{\epsilon \theta} \right)^2,$$

c-à-d proportionnelle à la probabilité de l'événement rare θ, où $q_{1-\alpha}$ est le quantile de la loi normale d'ordre $1-\alpha$.

On réécrit alors θ sous la forme :

$$\theta = \int_{-\infty}^{+\infty} \mathbf{1}_A(x) \frac{f(x)}{h(x)} h(x) dx = \mathbb{E}(\mathbf{1}_A(Y) L(Y))$$

Maintenant, en prenant des échantillons Y_i de $h(x)$, un estimateur sans biais de θ est donné par

$$\widehat{\theta}_N = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_A(Y_i) L(Y_i)$$

La méthode d'échantillonnage préférentiel est la méthode la plus utilisée dans l'étude des événements rares. Cependant, il existe d'autres techniques de réduction de la variance que nous présentons dans la suite.

Variables de contrôle. Le principe est analogue à la méthode précédente : on souhaite estimer $\mathbb{E}(f(X))$ que l'on réécrit en faisant apparaître artificiellement une fonction h :

$$\mathbb{E}(f(X)) = \mathbb{E}(f(X) - h(X)) + \mathbb{E}(h(X))$$

avec $\mathbb{E}(h(X))$ qui peut se calculer explicitement et $\mathrm{var}(f(X) - h(X))$ plus petite que $\mathrm{var}(f(X))$. On utilise alors une méthode de Monte Carlo pour évaluer $\mathbb{E}(f(X) - h(X))$ et le calcul direct pour $\mathbb{E}(h(X))$.

Exemple 2.2.1. — *Voir [45]. On souhaite calculer $\int_0^1 e^x dx$. On remarque que $\int_0^1 e^x dx = \mathbb{E}(f(X))$ avec $f(x) = e^x$ et que X suit une loi uniforme sur $[0,1]$. Au voisinage de 0, on a $e^x \sim 1 + x$ alors on écrit*

$$\int_0^1 e^x dx = \int_0^1 (e^x - 1 - x) dx + \frac{1}{2}$$

et il est facile de montrer que la variance par cette méthode diminue alors.

Variables antithétiques. Supposons que l'on cherche à calculer

$$I = \int_0^1 f(x) dx := \mathbb{E}(f(X))$$

avec X de loi uniforme sur $[0,1]$.

Comme la transformation $x \mapsto 1 - x$ laisse la mesure dx invariante, on a aussi

$$I = \frac{1}{2} \int_0^1 (f(x) + f(1-x)) dx$$

Et on utilise la méthode de Monte Carlo pour estimer I par

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}(f(U_i)+f(1-U_i))=\frac{1}{2n}\sum_{i=1}^{n}f(U_i)+f(1-U_i)$$

On peut montrer que si la fonction est continue et monotone, la qualité de l'approximation est meilleure que par rapport à la méthode de Monte Carlo directe avec $2n$ tirages.

On peut généraliser cette méthode en dimension supérieure et à d'autres transformations préservant la loi de la variable aléatoire.

Méthode de stratification. Supposons que l'on cherche à calculer

$$I=\mathbb{E}(g(X))=\int_{\mathbb{R}^d}g(x)f(x)dx$$

avec X est une v.a. de densité f.

On se donne une partition $(D_i, 1\leq i\leq m)$ de $\mathbb{R}^d$ et on décompose l'intégrale de la façon suivante

$$I=\sum_{i=1}^{m}\mathbb{E}(\mathbf{1}_{D_i}(X)g(X))=\sum_{i=1}^{m}\mathbb{E}(g(X)|X\in D_i)\mathbb{P}(X\in D_i)$$

Lorsqu'on connaît les $\mathbb{P}(X\in D_i)$ on utilise une méthode de Monte Carlo pour estimer les $\mathbb{E}(g(X)|X\in D_i)$ en répartissant les n tirages de façon optimale. Cette méthode accélère bien la simualtion car les espérances conditionnelles sont grandes.

Valeur moyenne ou conditionnement. Supposons que l'on cherche à calculer

$$\mathbb{E}(g(X,Y))=\int g(x,y)f(x,y)dxdy$$

où le couple (X,Y) a pour densité f. Posons

$$h(x)=\frac{1}{\int f(x,y)dx}\int g(x,y)f(x,y)dy$$

alors $\mathbb{E}(g(X,Y))=\mathbb{E}(h(X))$. Remarquons que

$$\mathbb{E}(g(X,Y)|X)=h(X)$$

Cette interprétation comme espérance conditionnelle permet de plus de prouver que la variance de $h(X)$ est inférieure à celle de $g(X,Y)$. Si l'on ne peut pas calculer directement $h(x)$, on utilise une méthode de Monte Carlo.

2.2.2.2. Une approche de grandes déviations. — Afin d'étudier les événements rares, on peut aussi faire appel à la théorie des grandes déviations, comme cela est fait dans [53] et [13] : lorsque l'on a un critère de grandes déviations, on utilise des méthodes d'échantillonage préférentiel et on démontre alors des résultats, en mêlant modélisation et approche analytique.

Soit $\mathbf{P} = \{P_k : k > 0\}$ une famille de mesures de probabilités sur un espace topologique $\mathcal{H}$. On suppose que $\mathbf{P}$ vérifie un principe de grandes déviations de fonction de taux $I : \mathcal{H} \to [0, \infty]$ c'est à dire que

- I est semi-continue inférieurement,
- pour tout O ouvert de $\mathcal{H}$, $\liminf_{k\to\infty} \frac{1}{k} \log(P_k(O)) \geq -I(O)$,
- pour tout F fermé de $\mathcal{H}$, $\limsup_{k\to\infty} \frac{1}{k} \log(P_k(F)) \leq -I(F)$,

où I(E)=$\inf_{x\in E} I(x)$.

Nous souhaitons estimer $P_k(E)$ pour un certain ensemble E. Pour cela, on utilise la technique d'échantillonnage préférentiel basée sur la méthode de Monte Carlo et exposée en section 2.2.2.1 : soit $X_k^{(1)}, \cdots, X_k^{(N_k)}$ un échantillon de N_k variables aléatoires indépendantes de loi commune Q_k. Alors l'estimateur de $P_k(E)$ correspondant est donné par

$$\widehat{P}_k(E) = \frac{1}{N_k} \sum_{l=1}^{N_k} \mathbf{1}_E(X_k^{(l)}) \frac{dP_k}{dQ_k}(X_k^{(l)}),$$

qui est bien défini et sans biais ssi $\mathbf{1}_E P_k \ll Q_k$.

Soit $N_k^c(Q_k) = \min\{N_k : \mathrm{var}_{Q_k}[\widehat{P}_k(E)] \leq cP_k(E)^2\}$ pour un certain $0 < c < \infty$. Alors $N_k^c(Q_k) = \lfloor v_k(E; Q_k)/(cP_k(E))^2 \rfloor$, où

$$v_k(E; Q_k) = \mathrm{var}_{Q_k}[\mathbf{1}_E(X_k) \frac{dP_k}{dQ_k}(X_k)] \geq 0.$$

Le contrôle de la variance de $\widehat{P}_k(E)$ donne un premier intervalle de confiance pour $P_k(E)$ par application de l'inégalité de Markov :

$$\begin{aligned}
\mathbb{P}\left(\frac{|\widehat{P}_k(E) - P_k(E)|}{\mathbb{P}(A)} \geq \alpha\right) &\leq \frac{1}{P_k(E)^2\alpha^2} \mathbb{E}\left(\left(\widehat{P}_k(E) - \mathbb{E}\left(\widehat{P}_k(E)\right)\right)^2\right) \\
&= \frac{\mathrm{var}_{Q_k}[\widehat{P}_k(E)]}{P_k(E)^2\alpha^2}
\end{aligned}$$

et par conséquent, le choix de $N_k^c(Q_k)$ garantit que la probabilité que l'erreur relative entre $P_k(E)$ et son estimateur dépasse α est bornée par c/α^2.

Il vient ainsi qu'une «bonne» distribution d'échantillonnage tendra à minimiser cette variance. Remarquons que, comme souligné en section 2.2.2.1, on peut annuler cette variance mais ce résultat n'est pas exploitable puisqu'alors $\mathbf{1}_E dP_k/dQ_k$ = constante et après normalisation, on obtient que la solution candidate annulant la variance est simplement la loi conditionnelle $Q_k(.) = \mathbf{1}_E P_k(.)/P_k(E)$ et qu'elle dépend de la quantité à estimer $P_k(E)$! Remarquons aussi que le coût total de la simulation est alors $N_k^c(Q_k)\times$«coût par échantillon» où le dernier terme correspond au coût de simulation de chaque échantillon $X_k^{(l)}$ auquel s'ajoute le coût engendré par l'évaluation du rapport de vraisemblance. La loi conditionnelle minimise seulement $N_k^c(Q_k)$ sans tenir compte de ce coût de simulation. Le rapport de vraisemblance devient $(dP_k/dQ_k)(X_k) = P_k(E)$ presque sûrement et l'estimateur est réduit à l'évaluation directe de $P_k(E)$ par la méthode de Monte Carlo classique. C'est pourquoi on obtient une variance nulle mais aussi un coût de simulation prohibitif (raison sans doute à l'origine de l'utilisation de la méthode d'échantillonnage préférentiel).

Il s'agit donc de trouver un compromis entre efficacité d'échantillonnage et complexité d'implémentation. Une stratégie largement utilisée est de contraindre les lois d'échantillonnage candidates à appartenir à une certaine famille. Par exemple, si P_k est une loi gaussienne, il est naturel de contraindre Q_k à être elle aussi gaussienne.

Contrairement au problème sans contrainte, la minimisation sous contrainte de $v_k(E; Q_k)$ peut s'avérer délicate. C'est pourquoi, on a souvent recours à des critères asymptotiques, particulièrement dans le cas des problèmes de grandes déviations. Puisque $\mathbf{P}$ vérifie un principe de grandes déviations, $P_k(E)$ est exponentiellement petit et $N_k^c(Q_k)$ sera exponentiellement grand pour $k \to \infty$ ($N_k^c(Q_k)$ est inversement proportionnel au carré de $P_k(E)$). Par conséquent, la vitesse de croissance exponentielle de $N_k^c(Q_k)$ est une caractérisation asymptotique naturelle de l'efficacité d'échantillonnage dans le contexte des problèmes de grandes déviations.

Nous avons déjà vu que Q_k est une loi d'échantillonnage *candidate* si

$$\mathbf{1}_E P_k \ll Q_k,$$

c'est une condition minimale pour que l'estimateur $\widehat{P}_k(E)$ soit bien défini. «Mesurons» maintenant ce taux de croissance exponentielle en étudiant

$$\limsup_{k\to\infty} \frac{1}{k} \log N_k^c(Q_k)$$

et définissons, pour un candidat $\mathbf{Q}$,

$$a(E, \mathbf{Q}) = \limsup_{k \to \infty} \frac{1}{k} \log N_k^c(Q_k)$$

Remarquons que $a(E, \mathbf{Q}) \geq 0$. Nous dirons qu'une suite candidate $\mathbf{Q}$ sera *efficace*, pour un ensemble E donné, si $a(E, \mathbf{Q}) = 0$.

On choisit $N_k^c(Q_k)$ et Q_k de la sorte afin de :
- minimiser le nombre de simulations, pour accélérer les calculs,
- diminuer la variance, comme expliqué précédemment.

Sadowsky établit une condition nécessaire générale d'efficacité asymptotique : de façon simplifiée, la suite candidate $\mathbb{Q} = (Q_k)$ doit être «localement proche» (en un sens à préciser et en des points réguliers) de la loi exponentiellement tiltée couramment utilisée en théorie des grandes déviations

$$P_k^\lambda(dx) = exp(k[< \lambda, x > -\Lambda_k(\lambda)])P_k(dx),$$

où $\Lambda_k(\lambda) = \frac{1}{k} \log \mathbb{E}_{P_k}[e^{k<\lambda,X_k>}]$. L'intérêt de ce résultat est qu'il s'applique à toute suite candidate $\mathbb{Q}$; nous ne contraignons pas ici la suite candidate à appartenir à une famille particulière. L'auteur présente aussi des conditions nécessaires et suffisantes pour l'optimalité des lois exponentiellement tiltées. Il apparaît que ces lois sont généralement non asymptotiquement optimales.

2.2.2.3. Echantillonnage préférentiel séquentiel. — Considérons maintenant une trajectoire aléatoire $x_{0:n} = (x_0, \cdots, x_n)$ où chaque x_i est un élément de l'espace d'état du système étudié. Nous souhaitons estimer l'espérance $\mathbb{E}(g(X))$ par rapport à la densité $f_n(x_{0:n})$. La méthode d'échantillonnage préférentiel présentée précédemment à la Remarque 2.2.1, estime cette espérance en générant les trajectoires $x_{0:n}^{(i)}$ selon $h_n(x_{0:n})$ et en calculant

$$\widehat{g}_N = \frac{\sum_{i=1}^N \omega_n^{(i)} g\left(x_{0:n}^{(i)}\right)}{\sum_{i=1}^N \omega_n^{(i)}}, \qquad \omega_n^{(i)} = \frac{f_n(x_{0:n}^{(i)})}{h_n(x_{0:n}^{(i)})}. \tag{3}$$

Ici, chaque tirage modifie la trajectoire tout entière. On peut alors imaginer un échantillonnage séquentiel consistant à actualiser la trajectoire sans modifier le passé, c-à-d que $(x_{0:n+1}^{(i)}) = (x_{0:n}^{(i)}, x_{n+1}^{(i)})$. On utilise alors l'identité suivante

$$f_n(x_{0:n}) = f_0(x_0)f_1(x_1|x_0)\cdots f_n(x_n|x_{0:n-1})$$

En pratique, on change la probabilité sous-jacente et on procède en générant X_0 selon la densité h_0 connue et proche de f_0 et récursivement, x_j est généré selon la densité conditionnelle $h_j(x_j|x_{0:j-1})$ de telle sorte que la densité de $x_{0:n}$ est donnée par

$$h_n(x_{0:n}) = h_0(x_0)h_1(x_1|x_0)\cdots h_n(x_n|x_{0:n-1}).$$

Les poids d'importance $\omega_n^i = f_n(x_{0:n}^{(i)})/h_n(x_{0:n}^{(i)})$ sont évalués récursivement puisque

$$\omega_{n+1}^i = \omega_n^i \frac{f_{n+1}(x_{n+1}^{(i)}|x_{0:n})}{h_{n+1}(x_{n+1}^{(i)}|x_{0:n})}.$$

Le problème rencontré ici est que lorsque, n croît, la distribution des poids d'importance $\omega_n^{(i)}$ devient de plus en plus étroite. En pratique, après quelques étapes seulement, seule une trajectoire a un poids d'importance non nul. Pour éviter cette dégénérescence, on introduit une étape de sélection supplémentaire. L'idée est d'éliminer les trajectoires ayant des poids d'importance faibles au profit des autres trajectoires. Formellement, on remplace l'estimateur pondéré (3) par

$$\widetilde{g}_N = \frac{1}{N}\sum_{i=1}^{N} m_n^{(i)} g\left(x_{0:n}^{(i)}\right)$$

où $m_n^{(i)}$ est le nombre d'«enfants» associé à la trajectoire $x_{0:n}^{(i)}$; c'est un entier naturel tel que $\sum_{i=1}^N m_n^{(i)} = N$. Les $m_n^{(i)}$ sont choisis de telle sorte que $\widetilde{g}_N$ est proche de $\widehat{g}_N$. Il existe différentes façons de procéder pour choisir les $m_n^{(i)}$: on peut par exemple générer selon une loi multinomiale de paramètres N et $\omega_n^{(i)},\quad i = 1, \cdots, N$. Voir [23] pour plus de détails.

2.2.2.4. Modèles multi-niveaux. — Dans la méthode d'échantillonnage préférentiel, les échantillons obtenus forment une trajectoire dans l'espace d'états, et la probabilité est modifiée pour que cette dernière ait une plus grande chance d'atteindre la cible : l'événement rare. Mais pour pouvoir appliquer l'échantillonnage préférentiel, il faut avoir une bonne connaissance du processus étudié.

Un autre moyen d'augmenter le nombre de visites à l'événement rare est d'utiliser des multi-trajectoires préférentielles. Cette technique s'appuie sur l'hypothèse qu'il existe des états intermédiaires identifiables visités par la trajectoire, plus souvent que l'événement rare lui-même et qui vont se comporter comme des passerelles vers l'événement rare. Contrairement à l'échantillonnage préférentiel classique, l'évolution du système suit la loi de probabilité originale. Nous considérons donc une suite décroissante de $M+1$ ensembles emboîtés B_i

menant à l'événement rare A :

$$A = B_{M+1} \subset B_M \subset \cdots \subset B_1.$$

Alors la probabilité d'intérêt $\mathbb{P}(A)$, donnée par (1), est simplement le produit des probabilités de transition $\mathbb{P}(B_i|B_{i-1})$. Comme expliqué en Introduction, ces probabilités conditionnelles ne sont généralement pas connues explicitement. Par contre, nous savons faire évoluer les particules du niveau B_i vers le niveau suivant B_{i+1}.

Le principe de l'algorithme est, dans un premier temps, de simuler simultanément plusieurs particules partant de B_0 ; au bout d'un certain temps, certaines d'entre elles ont «mal» évolué, les autres ayant par opposition «bien» évolué c-à-d ayant atteint le niveau supérieur B_1. Dans un deuxième temps, les «mauvaises» particules sont regénérées à la position de «bonnes» particules. On renouvelle cette procédure pour chacun des niveaux et ce jusqu'à ce que A soit atteint. De cette façon, les particules les plus «prometteuses» sont favorisées ; malheureusement cet algorithme est difficile à analyser directement en raison de l'intéraction introduite entre les particules. Des exemples de cette classe d'algorithmes peuvent être trouvés dans [3] avec le schéma «go with the winners» appliqué à la recherche de feuilles profondes dans un arbre et dans [18–20, 23] dans un cadre général.

Cependant, en pratique cet algorithme peut s'avérer délicat à appliquer. Par exemple, le cas de l'estimation de la probabilité d'un événemnt rare avec des systèmes dynamiques aléatoires est complexe, en raison de la difficulté à déterminer théoriquement les niveaux optimaux B_i. De plus, la probabilité d'atteindre B_i peut varier avec le point d'entrée du niveau B_{i-1}. Ce n'est pas le cas, par exemple, pour des modèles markoviens comme les diffusions.

Par exemple, dans le cadre d'une chaîne de Markov, on peut procéder comme suit : soit O l'état initial du système que nous supposerons récurrent. Notre but est d'estimer la probabilité (sous la mesure invariante) d'atteindre un ensemble rare A avant de revenir en O. On applique la méthode multi-trajectorielle qui consiste à regénérer les trajectoires qui atteignent O sur la position de particules ayant passé le niveau supérieur. Dans la pratique, on définit un temps maximal de simulation T et on évalue p=$\mathbb{P}(T_A < \min(T_O, T))$ où T_B désigne le temps d'atteinte de B. Il se peut aussi que le temps de retour en O à partir de B_i soit très long. On peut alors décider de tuer toute trajectoire qui sera descendue de d niveaux (en supposant donc qu'elle n'atteindra plus B_{i+1}). Mais on introduit par cette manipulation un biais.

Une variante est la méthode RESTART introduite par Manuel et José Villén-Altamirano dans [58, 60, 62]. La différence consiste à tuer chaque trajectoire partant du niveau B_i lorsqu'elle redescend en B_{i-1}, sauf la dernière qui est poursuivie jusqu'à ce qu'elle atteigne à nouveau B_i où elle est scindée une nouvelle fois. Pour parer l'effet d'un arrêt prématuré des premières trajectoires, le poids de la dernière est R_i si elle touche un niveau inférieur avant d'atteindre un niveau supérieur au i-ème. On estime donc directement $\mathbb{P}(A)$. Un outil basé sur le modèle RESTART a été développé par Kelling pour l'étude des réseaux stochastiques de Petri [41]. Voir Section 2.3 pour plus de détails sur le modèle RESTART.

D'autres variantes et extensions ont été proposées : citons par exemple l'extension fournissant une erreur relative bornée appelée méthode LRE développée par Schreiber, Görg et Fuß dans [35] et enrichie avec d'autres méthodes dans [33] et [34]. Une autre extension a été proposée par Below dans [9] et [10]. Plusieurs extensions de RESTART et des modèles multi-niveaux ont été étudiées par Garvels, Kroese et van Ommeren dans [27], [28] et [29]. Enfin, une autre interprétation des modèles multi-niveaux est présentée par Harazti et Townsend dans [36] appelée DPR (Direct Probability Redistribution).

Nous présentons de façon plus détaillée, dans la suite de ce chapitre, le modèle RESTART et les algorithmes particulaires et rappelons les principaux résultats établis pour ces deux modèles.

2.3. Le modèle RESTART

On cherche à estimer la probabilité d'un événement rare A correspondant par exemple au franchissement d'un certain niveau L par un processus $X(t)$. Pour cela, on introduit M ensembles imbriqués C_i tels que $A \subset C_M \subset \cdots \subset C_1$. Ces ensembles sont définis par une fonction Φ de l'espace d'état dans $\mathbb{R}$ qui associe à chaque état de l'ensemble C_i un seuil T_i tel que $T_1 \leq T_2 \leq \cdots \leq T_M \leq L$. Dans une simulation classique, on génère plusieurs trajectoires telles que représentées en figure 1.

Dans la suite, nous désignerons par événements B_i et D_i le franchissement du seuil T_i respectivement par le bas et par le haut.

La méthode RESTART (<u>RE</u>petitive <u>S</u>imulation <u>T</u>rials <u>A</u>fter <u>R</u>eaching <u>T</u>hresholds) illustrée à la figure 2 est la suivante :

- On génère une trajectoire principale (trajectoire en gras dans la figure 2) comme dans une simulation classique. Le critère de fin de simulation

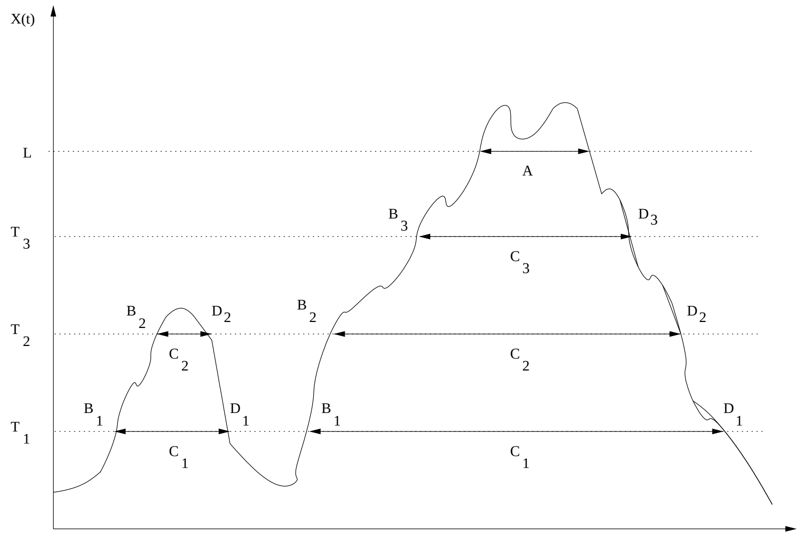


FIG. 1. Simulation classique

pourra être soit une durée définie, soit l'occurrence d'un certain événement. Ce critère sera appelé dans la suite *critère principal d'arrêt*.

– A chaque instant B_1, l'état du système est sauvegardé et la simulation principale est interrompue. On génère alors $R_1 - 1$ trajectoires partant de l'état X_{B_1} et se terminant à l'instant D_1.
– Après $R_1 - 1$ simulations secondaires, la simulation principale redémarre à partir de l'état sauvegardé au précédent instant B_1 et continue jusqu'au prochain événement B_1. Si une trajectoire secondaire atteint le critère principal d'arrêt avant l'occurrence d'un événement D_1, elle sera interrompue.
– Si un événement B_2 se produit lors d'une simulation secondaire ou de la simulation principale dans l'intervalle $[B_1, D_1)$, on génère alors R_2 retirages en suivant la même procédure que celle précédemment décrite. Notons que la trajectoire principale peut générer plusieurs événements B_2 dans l'intervalle de temps $[B_1, D_1)$.

De façon plus générale, nous avons l'algorithme suivant

– On génère une trajectoire principale (appelée également trajectoire de niveau 0) avec pour critère d'arrêt un critère d'arrêt principal. On adoptera la convention que l'événement D_0 est vide . On pose $i = 1$.
– Si un événement B_i se produit lors de la simulation d'une trajectoire secondaire de niveau $i - 1$, on interrompt la simulation, on mémorise l'état

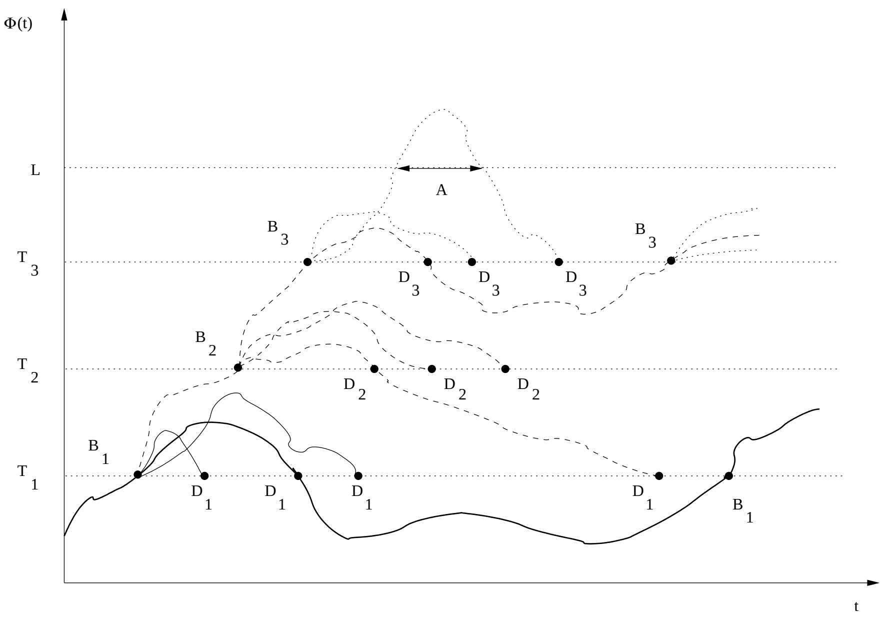


FIG. 2. Simulation RESTART

du système lors de l'occurrence de cet événement. Ensuite, $R_i - 1$ nouvelles trajectoires sont générées avec comme critère d'arrêt soit le critère principal d'arrêt soit l'occurrence d'un événement D_i.

– Après les $R_i - 1$ simulations secondaires de niveau i, la trajectoire de niveau $i-1$ redémarre à partir de l'état sauvegardé. La trajectoire peut alors :
  • soit atteindre l'ensemble C_i, et dans ce cas on procède à nouveau comme dans l'étape précédente,
  • soit générer un événement D_{i-1}. Dans ce cas, la trajectoire sera tuée si elle a été générée lors de l'occurrence d'un événement $i-1$, et continuera sinon,
  • soit vérifier le critère d'arrêt, et dans ce cas la trajectoire sera tuée.
– $i \leftarrow i+1$ et retourner à l'étape 2.

2.3.1. Notations et description récurrente. — Soit N_1^0 le nombre d'occurrences de l'événement B_1 générées par la trajectoire principale. Introduisons les instants successifs, définis à partir de la trajectoire principale, $\tau_i^{(n)}, n =$

$1, \cdots, N_1^0$ d'entrée dans C_i et $\sigma_i^{(n)}$ de sortie C_i :

$$\tau_i^{(n)} = \inf\{t \geq \sigma_i^{(n-1)} : X_t \in C_i\} = \inf\{t \geq \sigma_i^{(n-1)} : \Phi(X_t) = T_i\},$$
$$\sigma_i^{(n)} = \inf\{t \geq \tau_i^{(n)} : X_t \notin C_i\} = \inf\{t \geq \tau_i^{(n)} : \Phi(X_t) = T_{i-1}\}$$

avec la convention

$$\sigma_0^{(n)} = 0 \quad \text{pour tout } n \geq 0.$$

Dans chaque intervalle $[\tau_1^{(n)}, \sigma_1^{(n)})$, $R_1 - 1$ trajectoires secondaires sont générées de façon indépendante et avec même état initial $X_{\tau_1^{(n)}}$. Nous désignerons ces trajectoires secondaires et leurs ramifications successives par $X_t^{(j,n)}, j = 2, \cdots, R_1$. De même $X_t^{(1,n)}$ désignera la portion de la trajectoire principale et ses ramifications successives. Notons que chaque trajectoire $X_t^{(j,n)}$ démarre à l'instant $\tau_1^{(n)}$ et est tuée à l'instant aléatoire correspondant au franchissement par le haut du seuil T_1.

Si dans une simulation avec M niveaux, on supprime toutes les trajectoires $X_t^{(j,n)}$ pour $j = 2, \cdots, R_1$, on obtient une simulation à $M-1$ niveaux de seuils respectifs $T_2, \cdots, T_M$.
Introduisons les variables aléatoires suivantes :

$$\alpha_j^l = \int_{\tau_1^{(l)}}^{\infty} 1_{\{X_t^{(j,l)} \in A\}} dt, \quad j = 2, \cdots, R_1$$
$$\alpha_j = \sum_{k=1}^{N_1^0} \alpha_j^k$$

Ainsi α_j^l est la durée totale passée dans A par la j-ième trajectoire secondaire (et ses ramifications) générée après l'occurrence du l-ième événement B_1. Par conséquent α_j sera la durée totale passée dans A par les trajectoires secondaires d'indice j (y compris leurs ramifications). Notons que la trajectoire principale et les trajectoires secondaires partent toutes du même état et sont générées de façon indépendante, d'où pour un l fixé et $j = 1, \cdots, R_1$, les variables aléatoires $\alpha_j^{(l)}$ sont i.i.d., conditionnellement à l'état $X_{\tau_1^{(l)}}$. De même, les α_j sont i.i.d. conditionnellement à $\chi_1 = (N_1^0, X_{\tau_1^{(1)}}, \cdots, X_{\tau_1^{(N_1^0)}})$.

Désignons par N_A^m la durée totale dans A pour une simulation à m niveaux. Nous avons les relations suivantes :

$$N_A^{M-1} = \alpha_1, \qquad N_A^M = N_A = N_A^{M-1} + \sum_{j=2}^{R_1} \alpha_j.$$

On en déduit que $\mathbb{E}(N_A^M) = R_1\mathbb{E}(N_A^{M-1})$, et par conséquent l'estimateur $\widehat{P}^M$ de la probabilité de l'événement A pour une simulation à M niveaux

$$\widehat{P}^M = \frac{N_A^M}{Nr_M} \quad \text{avec} \quad r_0 = 1 \quad \text{et} \quad r_i = \prod_{i=1}^{i} R_j \quad \text{pour} \quad i = 1 \cdots M,$$

est non biaisé, car

$$\mathbb{E}(\widehat{P}^M) = \frac{\mathbb{E}(N_A^M)}{N\prod_{i=1}^M R_i} = \frac{\mathbb{E}(N_A^{M-1})}{N\prod_{i=2}^M R_i} = \mathbb{E}(\widehat{P}^{M-1}) = \mathbb{E}(\widehat{P}^0)$$

ce dernier étant non biaisé (estimateur pour une simulation conventionnelle).

Plus généralement, on définit les variables $Z_{i,j}^l$ comme la durée totale passée dans A par la j-ième trajectoire secondaire (et ses ramifications) générée après l'occurrence du l-ième événement B_i. Ainsi, $\alpha_j^l = Z_{1,j}^l$.

Remarque 2.3.1. — *De la même façon que la simulation simple, la méthode RESTART permet d'estimer la probabilité $\mathbb{P}(A)$ de l'événement rare A, quelle que soit la définition de $\mathbb{P}(A)$. En effet, on peut définir $\mathbb{P}(A)$ comme*

- la probabilité d'être en A (c'est-à-dire d'être dans un état de l'ensemble A) à un instant aléatoire donné ou en un instant d'occurrence de certains événements appelés instants de référence.

- ou encore la probabilité d'atteindre l'ensemble A.

Suivant les applications, on choisit l'une ou l'autre de ces définitions. Par exemple, dans l'étude d'une file d'attente, si l'événement rare est que la queue ait une longueur maximale (capacité de la queue), on ne s'interesse pas à la probabilité que la queue soit maximale en un instant aléatoire mais plutôt à celle qu'elle soit maximale en un instant d'arrivée.

Pour plus de simplicité, la notation $\mathbb{P}(A)$ fera, dans cette partie, référence à la première définition. De même pour les probabilités des autres événements.

Autres notations

- $P_{h|i} = \frac{\mathbb{P}(C_h)}{\mathbb{P}(C_h)}$, $P_{i+1|i} = P_{i+1}$ pour $i = 0 \cdots M$ et $P_{A|i} = P_{M+1|i}$.
- N_i^0 nombre d'événements B_i dans le tirage principal et N durée de la simulation (nombre d'instants de référence).
- a_i durée moyenne en C_i dans un tirage $[B_i, D_i)$.
- $P_{A|X_i}^*$ durée moyenne en A dans un tirage $[B_i, D_i)$ sachant que le système en B_i est X_i et $P_{A|i}^*$ son espérance. Remarquons aussi que $P_{A|i}^* = a_i P_{A|i}$.

Hypothèses

Supposons que le processus est régénératif, ce qui signifie, grossièrement, qu'à chaque fois qu'un phénomène particulier se produit (ici passage d'un seuil) le futur du processus à partir de cet instant d'occurrence est une réplique du futur du processus à partir de l'instant 0. Les temps d'occurrence sont appelés temps régénératifs.

2.3.2. Etude de la variance et optimisation. —

2.3.2.1. Calcul de la variance. — En utilisant les notations introduites en section 2.3.1, la variance de $\widehat{P}^M$ s'exprime par (cf [61]) :

Proposition 2.3.1. —

$$var(\widehat{P}^M) = \frac{K_A \mathbb{P}(A)}{N}\Big[\frac{1}{r_M} + \sum_{i=1}^{M} \frac{s_i P_{A|i}(R_i - 1)}{r_i}\Big], \tag{4}$$

avec

$$K_A = \frac{Var(N_A^0)}{\mathbb{E}[N_A^0]},$$

$$s_i = \frac{1}{K_A P_{A/i}} \frac{Var\big(\mathbb{E}[N_A^0|\chi_i]\big)}{\mathbb{E}[N_A^0]} \quad où \quad \chi_i = (N_i^0, X_{\tau_i^{(1)}}, \cdots, X_{\tau_i^{(N_i^0)}}).$$

Pour la preuve voir en Annexe 1. Le facteur K_A traduit l'autocorrélation du processus.

Remarque 2.3.2. — *Cette valeur est à comparer à la variance de l'estimateur $\widehat{P}$ sans RESTART (0 seuil, simulation simple sans retirage) qui est donnée par*

$$var(\widehat{P}) = \frac{K_A \mathbb{P}(A)}{N}$$

Nous souhaitons dans cette section optimiser l'algorithme. Pour ce faire, nous minimisons le gain de la simulation : le rapport entre les produits coût× variance dans la simulation simple et dans la simulaiton RESTART. Il s'agit donc maintenant de définir la notion de coût, d'effort.

2.3.2.2. Calcul du coût. — Nous définirons le coût de la simulation par le temps nécessaire pour réaliser la simulation, en prenant comme unité de temps le temps moyen entre chaque événement de référence dans la simulation simple. Avec cette définition le coût de la simulation simple est N.

Dans le cadre de la simulation RESTART, le coût est sensiblement plus grand. En effet,
- pour chaque événement, il y a une augmentation du coût, principalement due à la nécessité de calculer la fonction d'importance Φ de cet événement et de comparer cette valeur à celles des différents niveaux.
- pour chaque retirage, il y a une augmentation du coût, principalement due à la restauration de l'événement en B_i.

Pour tenir compte de ces coûts supplémentaires, le coût moyen d'un événement de référence dans le cadre RESTART est multiplié par :
- un facteur $y_e > 1$ dans tous les cas,
- un facteur supplémentaire $y_{ri} > 1$ si l'événement de référence se produit à un retirage du niveau i.

Avec ces définitions, le coût moyen d'un événement de référence est
- $y_0 = y_e$ dans le tirage principal,
- $y_i = y_e y_{ri}$ dans le retirage du niveau i.

Or, le nombre moyen d'événements de référence dans les retirages du niveau i est $NP_{i|0}r_{i-1}(R_i - 1)$.

Proposition 2.3.2. —

$$C = N\Big[y_0 + \sum_{i=1}^{M} y_i P_{i|0} r_{i-1}(R_i - 1)\Big] \tag{5}$$

2.3.2.3. Calcul du gain. — Comme expliqué précédemment, nous définissons le gain par

$$G = \frac{C_s V_s}{C_r V_r} = \frac{\text{coût de la simulation simple.variance de } \widehat{P}^0}{\text{coût de la simulation avec RESTART.variance de } \widehat{P}^M}$$

et on obtient :

Proposition 2.3.3. — *En utilisant les notations introduites en section 2.3.1,*

$$G = \frac{1}{P[\sum_{i=0}^{M} \frac{s_{i+1}u_i(1-P_{i+1})}{P_{i+1|0}r_i}][\sum_{i=0}^{M} y_i v_{i+1} P_{i|0} r_i (1 - P_{i+1})]}. \tag{6}$$

avec

$$\begin{cases} u_i & = \frac{1-(s_i/s_{i+1})P_{i+1|i}}{1-P_{i+1|i}} \\ v_{i+1} & = \frac{1-(y_{i+1}/y_i)P_{i+1|i}}{1-P_{i+1|i}} \\ s_0 & = 0, s_{M+1} = 1 \quad et \quad y_{M+1} = 0. \end{cases}$$

Pour la preuve voir en Annexe 1.

2.3.2.4. Optimisation. — L'optimisation de l'algorithme se fait en minimisant le gain en 3 étapes :

- on optimise le gain en les r_i à P_i fixés c-à-d aux niveaux fixés,
- on en déduit les valeurs optimales des P_i,
- on en déduit le nombre optimal de niveaux.

- Nombre optimal de retirage
  Pour calculer les valeurs optimales des r_i, on annule le gradient du gain.

  Soit $G^* = \frac{1}{GP}$. On a, pour $i = 1 \cdots M$,

$$\frac{\partial G^*}{\partial r_i} = [-\frac{s_{i+1}u_i(1-P_{i+1|i})}{P_{i+1|0}r_i^2}][\sum_{j=0}^{M} y_j v_{j+1} P_{j|0} r_j (1 - Pj + 1|j)]$$
$$+ [\sum_{j=0}^{M} \frac{s_{j+1}u_j(1-P_{j+1|j})}{P_{j+1|0}r_j}][y_i v_{i+1} P_{i|0}(1 - Pi + 1|i)].$$

  L'annulation de ces dérivées conduit à

$$r_i = \frac{1}{\sqrt{P_{i+1|1}P_{i|0}}}\sqrt{\frac{s_{i+1}y_0}{s_1 y_i}}\sqrt{\frac{u_i v_1}{u_0 v_{i+1}}} \qquad \text{pour } 1 \leq i \leq M$$

  Ainsi,

$$R_i = \frac{r_i}{r_{i-1}} = \sqrt{\frac{Pi|1P_{i-1|0}}{Pi+1|1P_{i|0}}}\sqrt{\frac{s_{i+1}y_{i-1}}{s_i y_i}}\sqrt{\frac{u_i v_i}{u_{i-1}v_{i+1}}},$$

  D'où,

$$R_i = \sqrt{\frac{P_{i-1|1}}{Pi+1|1}}\sqrt{\frac{s_{i+1}y_{i-1}}{s_i y_i}}\sqrt{\frac{u_i v_i}{u_{i-1}v_{i+1}}} \qquad 1 \leq i \leq M$$

  On note que $R_1 = r_1$ vérifie bien la relation précédente.

Remarque 2.3.3. — *Malheureusement, les valeurs de R_i ainsi obtenues ne sont pas toujours réelles et supérieures ou égales à 1.*

(*) Condition pour avoir $R_i \in \mathbb{R}$: remarquons que si pour tout i,

$$P_{i+1} \leq \frac{s_{i+1}}{s_i} \quad et \quad P_i \leq \frac{y_{i-1}}{y_i} \tag{7}$$

alors on aura $\forall i, \quad R_i \in \mathbb{R}$.

(**) Condition pour avoir $R_i \geq 1$: on remarque que

$$\left(\frac{s_i}{s_{i+1}} P_{i+1}\right)\left(\frac{y_i}{y_{i-1}} P_i\right) \leq \frac{u_i v_i}{u_{i-1} v_{i+1}} \Leftrightarrow 1 \leq R_i \tag{8}$$

- Position optimale des niveaux
  Calculs peu praticables de part la complexité des formules.

- Nombre optimal de niveaux
  Etant donné un ensemble de niveaux satisfaisant (7) et (8) alors les R_i qui maximisent le gain sont strictement supérieurs à 1 en général. Ce qui signifie qu'étant donné n'importe quel ensemble de valeurs de Φ telles que (7) et (8) sont vérifiées alors le gain est maximal lorsque toutes sont prises comme les valeurs d'un niveau. Ceci est encore valable lorsque l'on considère l'ensemble de toutes les valeurs de Φ vérifiant (7) et (8). Donc la solution serait de mettre le plus de niveaux possible.

2.3.2.5. Analyse des paramètres. — Tout d'abord, on obtient les expressions de var$(\mathbb{E}(N_A^0|\chi_1))$ et K_A :

Proposition 2.3.4. — *On a*

$$var(\mathbb{E}(N_A^0|\chi_1)) = \mathbb{E}(N_i^0)(P_{A|i})^2[K_i^{'} + \frac{var(P^*_{A|X_i})}{(P^*_{A|i})^2}\gamma_i] \tag{9}$$

et

$$K_A = a_j P_{A|j}\left(K_j^{'} + \frac{var(P^*_{A|X_j})}{(P^*_{A|j})^2}\gamma_i + \frac{\mathbb{E}(\ var(Z_{i,1}|X_j))}{(P^*_{A|j})^2}\right) \tag{10}$$

Comme nous l'avons déjà dit, K_A traduit l'autocorrélation du processus. S'il est décorrélé, $K_A = 1 - \mathbb{P}(A)$ et plus généralement on peut montrer que

$$K_A = 1 - P + \sum_{1\leq j\leq N,\ j\neq i} [\mathbb{P}(Y_j = 1|Y_i = 1) - P].$$

Pour les preuves de ces résultats voir en Annexe 1.

Ensuite, comme $\mathbb{E}(N_A^0) = P^*_{A|i}\mathbb{E}(N_i^0)$ et $P^*_{A|i} = a_i P_{A|i}$, on obtient l'expression de s_i :

Proposition 2.3.5. —

$$s_i = \frac{a_i}{K_A}\left[K_i^{'} + \frac{var(P^*_{A|X_i})}{(P^*_{A|i})^2}\gamma_i\right] \tag{11}$$

Analysons cette formule :
- Le facteur $K_i^{'} = \frac{\mathrm{var}(N_i^0)}{\mathbb{E}(N_i^0)}$ mesure l'autocorrélation du processus en les événements B_i dans le tirage principal. Si le processus est décorrelé, $K_i^{'} = 1 - \frac{P_{i|0}}{a_i}$. En fait, il se calcule de la même façon que K_A, à la différence que N_i^0 n'est pas la durée passée en C_i par le tirage principal mais le nombre d'événements B_i dans ce même tirage. Il faut donc remplacer A par B_i, $\mathbb{P}(A)$ par $\frac{P_{i|0}}{a_i}$.
On obtient ainsi, $K_i^{'} = 1 - \frac{P_{i|0}}{a_i} + \sum_{1\leq j\leq N,\ j\neq k}[\mathbb{P}(Y_j = 1|Y_k = 1) - \frac{P_{i|0}}{a_i}]$.

- Le facteur γ_i mesure la dépendance entre les états du système au niveau i. En général, il y a dépendance entre les états du système pour des événements B_i proches mais cette dépendance devient négligeable pour des événements B_i éloignés. Pour illustrer ceci, on peut penser par exemple aux tremblements de Terre et aux répliques qui suivent la première secousse. Si les X_i sont indépendants alors $\gamma_i = 1$.

- $\frac{\mathrm{var}(P^*_{A|X_i})}{(P^*_{A|i})^2}$ dépend grandement de la fonction d'importance choisie.

Enfin, en utilisant l'expression de K_A, s_i devient

Proposition 2.3.6. —

$$s_i = \frac{a_i}{a_j}\frac{1}{P_{A|j}}\frac{K_i^{'} + \frac{var(P^*_{A|X_i})}{(P^*_{A|i})^2}\gamma_i}{K_j^{'} + \frac{var(P^*_{A|X_j})}{(P^*_{A|j})^2}\gamma_j + \frac{\mathbb{E}(\ var(Z_{i,1}|X_j))}{(P^*_{A|j})^2}} \tag{12}$$

On remarque que le dénominateur fait intervenir l'indice j que l'on choisit de façon quelconque, il est indépendant de l'indice i en lequel est calculé s. De plus, ceci ne s'applique pas seulement à n'importe quel niveau T_j de Φ déterminant un ensemble C_j mais à tout niveau T définissant un ensemble C contenant A. Cette formule est en particulier vraie pour le niveau T_L qui définit A. Alors $P_{A|L} = 1$ et $P^*_{A|L} = \mathbb{E}(Z_L) = a_L$. Si le dernier niveau T_M est égal à T_L, alors $s_M \leq 1$. De même, pour des niveaux proches de T_L, $s_i \leq 1$.

2.3.3. Conclusion. — L'étude analytique du modèle RESTART n'a pu être que partielle : certains résultats nécessitent des hypothèses supplémentaires pour être appliqués dans la pratique et la complexité des formules ne nous a pas permis d'optimiser de manière précise la méthode ni de vérifier que cette méthode permet bien une accélération de la simulation c-à-d une amélioration par rapport à la méthode simple sans retirage.

2.4. Les algorithmes particulaires

Nous présentons dans cette section les systèmes particulaires dans le cadre précis où l'on souhaite estimer la probabilité d'un événement A. Voir, par exemple, [18, 20, 23] pour la théorie générale des systèmes particulaires.

2.4.1. Cas particulier. — Etant donnée une variable aléatoire X prenant ses valeurs dans $(F, \mathcal{F})$ et une fonction réelle V définie sur F, on cherche à estimer par simulation

$$\mathbb{P}(V(X) \geq a)$$

pour un a tel que cette probabilité soit très petite. Désignons par μ la loi de X.

Exemple : 1) Soit X un vecteur aléatoire et $V(X)$ sa norme. On cherche à estimer la probabilité que la norme de X dépasse une certaine valeur a.
2) Soit X une matrice aléatoire et $V(X)$ sa plus grande valeur propre. On cherche la probabilité que la plus grande valeur propre de X dépasse une certaine valeur a.

Une première alternative serait de simuler un grand nombre N de variables aléatoires i.i.d. $X_1, .., X_n$ de loi μ et d'approcher $\mathbb{P}(V(X) \geq a)$ par

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{V(X_i)\geq a}$$

Ceci est légitime et fournit une bonne approximation grâce à la loi forte des grands nombres lorsque $N \to +\infty$.

Lorsqu'on ne connaît pas la loi μ, une seconde alternative consiste à utiliser un algorithme de Markov Chain Monte Carlo (MCMC) : on adapte l'algorithme d'acceptation-rejet présenté en Annexe 3 à la loi μ définie sur un espace F. Ici on veut générer un vecteur aléatoire $X \in F$ de loi μ de densité p par rapport à la mesure de Lebesgue. L'algorithme d'acceptation-rejet génère une chaîne de Markov $X_0, X_1, .., X_n, ..$ à valeurs dans F de loi réversible μ. On se donne une chaîne de Markov facile à simuler de noyau $q(x, y)$, qui nous servira à proposer une nouvelle réalisation. Cette chaîne devra être irréductible et apériodique.
Soit X_k la dernière valeur retenue et

$$h(x, y) = \frac{p(y)q(y, x)}{p(x)q(x, y)} \wedge 1$$

L'algorithme est le suivant :
(1) on génère $Y_k \sim q(X_k, .)$,
(2) on accepte Y_k avec la probabilité $h(X_k, Y_k)$ et $X_{k+1} = Y_k$, sinon on le rejette et $X_{k+1} = X_k$. Le noyau $K(x, dy)$ de la chaîne X correspondant s'exprime comme la somme de deux termes :

$$K(x, dy) = h(x, y)q(x, y)dy + (1 - \int h(x, z)q(x, z)dz)\delta_x(dy)$$

le premier terme correspond à l'acceptation et le deuxième au rejet.

Ainsi on est en mesure d'approcher μ : comme μ est invariant pour K et en supposant l'apériodicité et l'irréductibilité, on sait que pour $n \to +\infty$, la loi de X_n converge vers μ et que

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{V(X_k)\geq a\}} \to_{n\to+\infty} \mu(V \leq a) = \mathbb{P}(V(X) \geq a)$$

Le problème de ces deux modèles est que lorsque la probabilité recherchée est petite, on a un temps de simulation très long. On va donc chercher à accélérer celle-ci.

Principe

On se donne des seuils $a_0 < a_1 < \cdots < a_{M+1} = a$ et les mesures suivantes :

$$\mu_0 = \mu \quad \text{et} \quad \mu_i(.) = \frac{\mu(\mathbf{1}_{\{V(X)\geq a_{i-1}\}})}{\mu(\mathbf{1}_{\{V(X)\geq a_{i-1}\}})}.$$

On a alors la décomposition suivante :

$$\mathbb{P}(V(X) \geq a) = \mu(V \geq a) = \frac{\mu(V \geq a_{M+1})}{\mu(V \geq a_M)} \frac{\mu(V \geq a_M)}{\mu(V \geq a_{M-1})} \cdots \frac{\mu(V \geq a_1)}{\mu(V \geq a_0)} \mu(V \geq a_0)$$

On remarque que

$$\frac{\mu(V \geq a_i)}{\mu(V \geq a_{i-1})} = \mathbb{P}(V(X) \geq a_i | V(X) \geq a_{i-1})$$

Finalement,

$$\mathbb{P}(V(X) \geq a) = \prod_{p=0}^{M+1} \mu_p(V \geq a_p)$$

Les quantités $(\mu_p(V \geq a_p))_{p=0..M+1}$ sont plus faciles à estimer puisque ce sont les probabilités d'événements «moins rares» que $\{V(X) \geq a\}$.

On approchera les μ_i par des mesures empiriques μ_i^N, N étant le nombre de particules utilisées. Pour obtenir ces mesures empiriques, on utilisera l'algorithme précédent d'acceptation-rejet.

On se donne un noyau de proposition M tel que la mesure μ soit réversible pour M c-à-d tel que pour tous f et g

$$\mu(fM(g)) = \mu(M(f)g)$$

où $\mu(f) = \int f(x)\mu(dx)$ et $M(f)(x) = \int f(y)M(x, dy)$.

Ensuite pour chaque i, on construit les noyaux M_i admettant μ_i pour mesure invariante. On pose

$$M_i(x, dy) = M(x, dy)\mathbf{1}_{V(y)\geq a_{i-1}} + (1 - M(x, V \geq a_{i-1}))\delta_x(dy)$$

qui accepte les y proposés par le noyau M lorsque $V(y) \geq a_{i-1}$ et conserve x sinon. On vérifie aisément que μ_i est réversible pour le noyau M_i. En effet,

$$M_i(f)(x) = M(f\mathbf{1}_{\{V\geq a_{i-1}\}}) + M(x, V < a_{i-1})f(x)$$

et

$$\begin{aligned}
\mu_i(M_i(f)g) &= \frac{\mu(M_i(f)g\mathbf{1}_{\{V\geq a_{i-1}\}})}{\mu(V\geq a_{i-1})} \\
&\propto \mu(M(f\mathbf{1}_{\{V\geq a_{i-1}\}})g\mathbf{1}_{\{V\geq a_{i-1}\}}) + \mu(fM(.,\mathbf{1}_{\{V<a_{i-1}\}})g\mathbf{1}_{\{V\geq a_{i-1}\}}) \\
&\propto \mu(M(g\mathbf{1}_{\{V\geq a_{i-1}\}})f\mathbf{1}_{\{V\geq a_{i-1}\}}) + \mu(fM(.,\mathbf{1}_{\{V<a_{i-1}\}})g\mathbf{1}_{\{V\geq a_{i-1}\}}) \\
&= \mu_i(M_i(g)f)
\end{aligned}$$

par réversibilité de μ pour M.

On a donc un noyau réversible pour chacune des mesures μ_i. Ainsi en déroulant l'algorithme d'acceptation-rejet un nombre n suffisant de fois et par la formule produit, on obtient une estimation de notre probabilité. Cela revient à estimer les termes du produit de façon indépendante.

Un moyen d'accélérer à nouveau l'algorithme est d'utiliser un algorithme particulaire. Ici on estimera la probabilité d'un seul coup. Le principe est le suivant : partant de x_{i-1}, on fait évoluer la chaîne régie par M_i un nombre T assez grand de fois (suffisamment grand pour atteindre le régime stationnaire), c'est l'étape de mutation. On ne conserve ensuite que les particules ayant franchi le seuil a_i, les autres étant tuées puis redistribuées sur les particules vivantes : c'est l'étape de sélection.

2.4.2. Formalisme des algorithmes particulaires. — Soient $A = \{x : V(x) \geq a\}$ et pour $i = 1 \cdots M+1$, $B_i = \{x : V(x) \geq a_i\}$. Ainsi

$$A = B_{M+1} \subset B_M \cdots \subset B_1 \subset B_0$$

ainsi on a $a = a_{M+1}$.

Concrètement, on veut simuler des v.a. de loi μ_i et regarder si elles arrivent en B_{i+1}. On va donc, pour chaque i, dérouler la chaîne de Markov décrite dans l'algorithme MCMC et régie par M_i pendant un nombre de pas T_{μ_i} suffisant pour atteindre le régime stationnaire. On regarde ensuite si au bout de ces T_{μ_i} étapes, la particule a atteint le niveau B_{i+1}. On a donc défini des temps T_{μ_i}, nombres d'étapes nécessaires à la chaîne régie par M_i pour atteindre le régime stationnaire (ils dépendent chacun de T_μ). Notons $T = sup_{i=1..M+1}T_{\mu_i}$.

Pour capturer le comportement de la chaîne entre les différents niveaux, on introduit la suite stochastique discrète χ_n représentant les «chemins» (non continus) de X entre chaque étape. Ici les chemins ont la même longueur déterministe T. χ_n est une chaîne de Markov à valeurs dans l'ensemble des excursions $E = \cup_{p\geq 0}(\{p\} \times F^T)$ et de matrice de transition $\mathcal{M}_n$. Alors on a

$\chi_0 = (T_0, X_{T_0})$ et $T_0 = 0$. Si la particule est arrivée en B_{n-1}, on fait évoluer la chaîne pendant un temps T. Au bout de ce temps on regarde si la particule est en B_n. Si oui, $T_n = nT + 1$, $X_{T_n} \in B_n$ et $\chi_n = (T_n, X_{[T_{n-1}+1, T_n]}) = (T_n, X_{T_{n-1}+1}, \cdots, X_{T_n})$. Dans le cas contraire, $T_n = T_{n-1}$, $X_{T_n} \notin B_n$ et $\chi_n = X_{T_{n-1}}$. A l'intérieur de χ_n, X évolue selon M_n.

Pour vérifier si les chemins ont réussi à atteindre le niveau B_n, on introduit les fonctions g_n définies pour chaque chemin

$$x_{k:l} = (l, (x_j)_{j=k\cdots l})$$

par

$$g_n(x_{k:l}) = \mathbf{1}_{\{x_l \in B_n\}}$$

Ainsi le succès se traduit par l'événement

$$(g_1(\chi_1) = 1, \cdots, g_{M+1}(\chi_{M+1}) = 1)$$

et

$$\mathbb{P}(V(X) \geq a) = \mathbb{E}(\prod_{p=1}^{M+1} g_p(\chi_p))$$

On rappelle que la mesure de probabilité $\mu_n \in \mathcal{P}(E)$ sur l'espace E est définie pour toute fonction bornée f à valeurs dans E par

$$\mu_n(f) = \int f(x)\mu_n(dx) = \frac{\mu(\mathbf{1}_{\{V(X)\geq a_{n-1}\}}\cdot)}{\mu(\mathbf{1}_{\{V(X)\geq a_{n-1}\}})} := \frac{\gamma_n(f)}{\gamma_n(1)}$$

avec

$$\gamma_n(f) = \mathbb{E}(f(\chi_n)\prod_{p=1}^{n-1} g_p(\chi_p))$$

On appelle μ_n (resp. γ_n) la **mesure normalisée** (resp. **non-normalisée**) **de prédiction**. On définit le modèle de correction de la façon suivante :

$$\widehat{\mu}_n(f) := \frac{\widehat{\gamma}_n(f)}{\widehat{\gamma}_n(1)}$$

avec

$$\widehat{\gamma}_n(f) = \mathbb{E}(f(\chi_n)\prod_{p=1}^{n} g_p(\chi_p))$$

On appelle $\widehat{\mu}_n$ (resp. $\widehat{\gamma}_n$) la **mesure normalisée** (resp. **non-normalisée**) **de correction**.

Proposition 2.4.1. — *On a les relations suivantes :*

- *Avec ces notations,*

$$\begin{aligned} \mu_n(f) &= \mathbb{E}(f(\chi_n)|T_{n-1} = (n-1)T_{n-2} + 1) \\ \widehat{\mu}_n(f) &= \mathbb{E}(f(\chi_n)|T_n = nT_{n-1} + 1) \end{aligned}$$

- *Lien entre* γ_n *et* $\widehat{\gamma}_n$: $\gamma_n(f_n g_n) = \widehat{\gamma}_n(f_n)$
- *Lien entre* μ_n *et* $\widehat{\mu}_n$: $\widehat{\mu}_n(f_n) = \frac{\gamma_n(f_n g_n)}{\gamma_n(g_n)} = \frac{\mu_n(f_n g_n)}{\mu_n(g_n)}$
- *Lien entre* γ_n *et* μ_n : $\gamma_n(f_n) = \mu_n(f_n)\prod_{p=1}^{n-1}\mu_p(g_p)$
- *Lien entre* $\widehat{\gamma}_n$ *et* $\widehat{\mu}_n$: $\widehat{\gamma}_n(f_n) = \widehat{\mu}_n(f_n)\prod_{p=1}^{n-1}\mu_p(g_p)$

Le lecteur est renvoyé à l'Annexe 2 pour les preuves de cette section.

De manière évidente, la quantité recherchée $\mathbb{P}(V(X) \geq a)$ n'est autre que

$$\gamma_{M+1}(g_{M+1}) = \gamma_{M+2}(1)$$

Ainsi on a reformulé notre problème avec le formalisme de Feynman-Kac et on va donc pouvoir utiliser toute la théorie sous-jacente. Tout d'abord, définissons $\mathcal{P}(E)$ et $\mathcal{P}_n(E)$ par

$$\mathcal{P}(E) := \{\text{mesures de probabilités sur E}\}$$

et

$$\mathcal{P}_n(E) := \{\eta \in \mathcal{P}(E) : \eta(g_n) > 0\}$$

Théorème 2.4.1. — *L'évolution de* μ_n *est régie par le système dynamique non-linéaire suivant :*

$$\mu_{n+1} = \Phi_{n+1}(\mu_n) \tag{13}$$

avec $\Phi : \mathcal{P}_n(E) \to \mathcal{P}(E)$

$$\Phi_{n+1}(\mu) = \Psi_n(\mu)\mathcal{M}_{n+1}$$

où

- $\Psi_n : \mathcal{P}_n(E) \to \mathcal{P}_n(E)$ *et*

$$\Psi_n(\mu)(dx_n) = \frac{g_n(x_n)}{\mu(g_n)}\mu(dx_n)$$

  Ψ_n *est la* ***transformation de Gibbs-Boltzman***
- $\mathcal{M}_{n+1}$ *représente le noyau de transition de la chaîne de Markov* χ_{n+1}*, ie*

$$\mathcal{M}_{n+1}(u, dv) = \mathbb{P}(\chi_{n+1} \in dv|\chi_n = u)$$

La récursion (13) se traduit par deux transitions séparées :

$$\mu_n \underset{\text{sélection}}{\longrightarrow} \widehat{\mu}_n = \Psi(\mu_n) \underset{\text{mutation}}{\longrightarrow} \mu_{n+1} \tag{14}$$

Comme les g_n sont des indicatrices, on écrit $\Psi_n(\mu) = \mu\mathcal{S}_{n,\mu}$ en termes de transitions markoviennes de sélection

$$\mathcal{S}_{n,\mu}(u, dv) = \mathbf{1}_{g_n^{-1}(1)}(u)\delta_u(dv) + [1 - \mathbf{1}_{g_n^{-1}(1)}(u)]\Psi_n(\mu)(dv)$$

L'interprétation de ces modèles discrets de Feynman-Kac à l'aide de particules se fait naturellement. Le modèle consiste en N «particules-excursions» :

$$\begin{aligned} \zeta_n^i &= (T_n^i, X^i_{[T^i_{n-1}, T^i_n]}) \in E \cup \{\Delta\} \\ \widehat{\zeta}_n^i &= (\widehat{T}_n^i, \widehat{X}^i_{[\widehat{T}^i_{n-1}, \widehat{T}^i_n]}) \in E \cup \{\Delta\} \end{aligned}$$

Au temps $n = 0$, le système initial se compose de N particules indépendantes et identiquement distribuées $\zeta_0^i = (T_0^i, X_0^i)$ de loi $\mu_0 = \delta_0 * \mu$. Et comme $g_0(u_0) = 1$, il n'y a pas de transition de correction à la première étape.

Puisqu'on ne sait pas résoudre explicitement le système dynamique (14), on va chercher à approximer le modèle : le système génétique N-particulaire est basé sur l'idée naturelle d'approximer $(\mu_n)_{n\geq 1}$ par une suite de mesures empiriques

$$\mu_n \approx \mu_n^N = \frac{1}{N}\sum_{i=1}^N \delta_{\zeta_n^i}$$

associées au système des N particules interactives $\zeta_n = (\zeta_n^1, \cdots, \zeta_n^N)$. En d'autres termes, on approxime le schéma de double transition (14) du système en un schéma en deux temps

$$\mu_n^N \underset{\text{sélection}}{\longrightarrow} \widehat{\mu}_n^N = \frac{1}{N}\sum_{i=1}^N \delta_{\widehat{\zeta}_n^i} \underset{\text{mutation}}{\longrightarrow} \mu_{n+1}^N \tag{15}$$

où les $\widehat{\zeta}_n^i$ sont N variables aléatoires indépendantes de même loi

$$\widehat{\mu}_n^N = \Psi_n(\mu_n^N) = \Psi_n(\frac{1}{N}\sum_{i=1}^N \delta_{\zeta_n^i}) = \sum_{i=1}^N \frac{g_n(\zeta_n^i)}{\sum_{j=1}^N g_n(\zeta_n^j)}\delta_{\zeta_n^i}$$

On remarquera que $\widehat{\mu}_n^N$ est la mesure d'occupation des excursions ayant atteint le n-ième niveau.

En utilisant les mesures approchées μ_n^N, on approxime γ_n par la mesure γ_n^N :

$$\gamma_n \approx \gamma_n^N(f) = \mu_n^N(f) \prod_{p=1}^{n-1} \mu_p^N(g_p)$$

Ainsi on a transformé l'espérance d'un produit en le produit d'espérances.

Par conséquent, une estimée de $\mathbb{P}(V(X) \geq a)$ est donnée par

$$\mathbb{P}(V(X) \geq a) \approx \gamma_{M+1}^N(g_{M+1}) = \gamma_{M+2}^N(1) = \mu_{M+2}^N(1) \prod_{p=1}^{M+1} \mu_p^N(g_p)$$

Comme les potentiels g_n sont dégénérés, on définit un point cimetière Δ tel que $\zeta_n = \Delta$ signifie qu'aucune des particules ζ_n^i n'a réussi à atteindre le niveau B_{n+1} et on étudiera le système pour les $n < \tau^N$ où

$$\tau^N := \inf\{n \geq 0 : X^i_{T^i_n} \notin F, \quad 1 \leq i \leq N\}$$

Concrètement, tant que $n < \tau^N$, l'algorithme génétique procède comme suit :

Etape 1 : Initialisation, n=0
Pour i=1..N, on génère $\zeta_0^i \sim \mu_0$ et on pose $\widehat{\zeta}_0^i = \zeta_0^i$.

Etape 2 : Mutation/Prédiction $\widehat{\zeta}_n^i \to \zeta_{n+1}^i$
• Si $\widehat{\zeta}_n = \Delta$, alors $\zeta_{n+1} = \Delta$: l'algorithme est stoppé.
• Sinon, pour i=1..N, on génère une nouvelle excursion depuis $X^i_{T_n}$, selon la transition markovienne de la chaîne de Markov χ_{n+1}. Cela revient à faire évoluer X en partant de $X^i_{T_n}$ en T étapes ou pas selon le noyau markovien M_{n+1}.

Etape 3 : Sélection/Updating $\zeta_{n+1}^i \to \widehat{\zeta}_{n+1}^i$
• On rééchantillonne les N particules ζ_{n+1}^i selon la mesure empirique :

$$\Psi_{n+1}\left(\frac{1}{N}\sum_{i=1}^N \delta_{\zeta_{n+1}^i}\right) = \sum_{i=1}^N \frac{g_n(\zeta_{n+1}^i)}{\sum_{j=1}^N g_{n+1}(\zeta_{n+1}^j)} \delta_{\zeta_{n+1}^i} = \frac{1}{|I_{n+1}^N|} \sum_{i \in I_{n+1}^N} \delta_{\zeta_{n+1}^i}$$

où I_{n+1}^N est l'ensemble des indices des particules ayant atteint le niveau $n+1$ c-à-d

$$I_{n+1}^N = \{1 \leq i \leq N : X^i_{T^i_{n+1}} \in B_{n+1}\}$$

On travaille ainsi à taille de population fixée et égale à N.
• on pose $n \leftarrow n+1$ et on revient à l'étape 2.

2.4.3. Etude de l'estimateur. — Il est possible de borner la probabilité d'extinction de l'algorithme comme le montre le théorème suivant :

Théorème 2.4.2. — $\forall N \geq 1,\ n \geq 0,$

$$\mathbb{P}(\tau^N \geq n) \leq a(n)\exp\{-Nb(n)\}$$

avec des constantes $a(n)$ et $b(n)$ ne dépendant que de n.

Rappelons que l'estimateur de $\mathbb{P}(V(X) \geq a)$ est

$$\gamma^N_{M+2}(1)\mathbf{1}_{\tau^N \geq M+2}$$

et c'est un estimateur sans biais :

Théorème 2.4.3. — $\forall N \geq 1,$

$$\mathbb{E}(\gamma^N_{M+2}(1)\mathbf{1}_{\tau^N \geq M+2}) = \gamma_{M+2}(1)$$

Enonçons maintenant le théorème central limite :

Théorème 2.4.4. —

$$\sqrt{N}(\mathbf{1}_{\tau^N \geq M+2}\gamma^N_{M+2}(1) - \mathbb{P}(V(X) \geq a)) \underset{\mathcal{L}, N \to +\infty}{\longrightarrow} L_{M+2}$$

où L_{M+2} est un processus gaussien centré de variance

$$v^2_{M+1} = \sum_{q=1}^{M+1} \gamma_q(1)^2 \mu_{q-1}\{K_{q,\mu_{q-1}}[Q_{q,M+2}(1) - K_{q,\mu_{q-1}}(Q_{q,M+2}(1))]^2\}$$

Pour la définition des opérateurs $K_{q,\mu_{q-1}}$ et $Q_{q,M+2}$, le lecteur est renvoyé à l'Annexe 2.

2.4.4. Etude de la variance asymptotique. — Dans le cas où les fonctions de sélection g_k sont des indicatrices et plus particulièrement dans notre cas, la variance asymptotique devient simplement

Proposition 2.4.2. —

$$v^2_{M+1} = \mathbb{P}(V(X) \geq a)^2 \sum_{i=0}^{M} \left(\frac{1}{P_{i+1}} - 1\right) \tag{16}$$

et si les probabilités de transition sont toutes égales à P_0,

Corollaire 2.4.1. —

$$v_{M+1}^2 = \mathbb{P}(V(X) \geq a)^2 (M+1)\left(\frac{1}{P_0} - 1\right) \tag{17}$$

Comparaison avec différents modèles

1. On considère la méthode naïve basée sur N copies indépendantes X^i de X. L'estimateur sans biais correspondant est simplement donné par $N^{-1}\sum_{i=1}^N \mathbf{1}_{B_n}(X^i)$. Par le théorème central limite classique, la suite

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N [\mathbf{1}_{B_{M+1}}(X^i) - \mathbb{P}(V(X) \geq a)]$$

converge en loi quand $N \to +\infty$ vers une v.a. $\overline{L}_{M+2}$ gaussienne centrée et de variance

$$\overline{v}_{M+1}^2 = \mathbb{P}(V(X) \geq a)(1 - \mathbb{P}(V(X) \geq a)) = P_0^{M+1}\left(1 - P_0^{M+1}\right)$$

puisque $\mathbb{P}(V(X) \geq a) = P_0^{M+1}$. On obtient alors,

$$\frac{\overline{v}_{M+1}^2}{v_{M+1}^2} = \frac{1 - P_0^{M+1}}{P_0^{M+1}(M+1)\left(\frac{1}{P_0} - 1\right)} \underset{M \to +\infty}{\longrightarrow} +\infty$$

2. On considère le modèle de branchement avec duplication des trajectoires, déjà présenté au Chapitre 1 et étudié dans cette thèse dans les chapitres suivants. L'estimateur sans biais correspondant est simplement donné par $N^{-1}\sum_{i=1}^N \widetilde{Q}_i$ avec

$$\widetilde{Q}_i := \frac{N_A^{(i)}}{R^M}$$

où $N_A^{(i)}$ représente le nombre de particules ayant atteint l'événement A issues de la particule initiale numéro i et R le nombre de splitting.

Le coût moyen d'une telle simulation est donné par

$$C = N(M+1)$$

de même que pour les algorithmes particulaires. Par le théorème central limite classique, comme les $\widetilde{Q}_i$ sont des variables indépendantes, la suite

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N [\widetilde{Q}_i - \mathbb{P}(V(X) \geq a)]$$

converge en loi quand $N \to +\infty$ vers une v.a. $\widetilde{L}_{M+2}$ gaussienne centrée et de variance

$$\widetilde{v}^2_{M+1} = \mathbb{P}(V(X) \geq a)^2 (M+1) \left(\frac{1}{P_0} - 1 \right)$$

On obtient alors,

$$\frac{\widetilde{v}^2_{M+1}}{v^2_{M+1}} = 1$$

3. On considère maintenant le modèle multi-niveaux adaptatif introduit par Frédéric Cérou et Arnaud Guyader dans [19] et présenté au chapitre 5 en section 5.4. On génère N copies indépendantes du processus étudié et on place le premier niveau de telle façon qu'une proportion $p = k/N$ de particules arrive à ce niveau, particules que l'on conserve. On regénère à partir de ce niveau $(N-k)$ nouvelles particules, de façon à avoir une taille de population constante. On répète ce processus jusqu'à dépasser le niveau étudié. L'estimateur $\widehat{P}_N$ sans biais correspondant est donné par rp^m où r est la proportion de particules ayant atteint l'événement rare partant de l'avant-dernier niveau. Ecrivons $\mathbb{P}(V(X) \geq a) = \rho p^m$. On montre alors que la suite

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\widehat{P}_N - \mathbb{P}(V(X) \geq a) \right]$$

converge en loi quand $N \to +\infty$ vers une v.a. $\widehat{L}_m$ gaussienne centrée et de variance

$$\widehat{v}^2_m = \mathbb{P}(V(X) \geq a)^2 \left[m(\frac{1}{p} - 1) + (\frac{1}{\rho} - 1) \right]$$

Puisque $\rho p^m = P_0^{M+1}$,

$$\frac{m}{M+1} = \frac{\log P_0 - \log \rho / M + 1}{\log p}$$

On obtient alors que, lorsque $a \to +\infty$, $M+1$ et m tendent vers l'infini et que par conséquent

$$\frac{\widehat{v}^2_m}{v^2_{M+1}} \underset{a \to +\infty}{\longrightarrow} \frac{\log P_0}{\log p} \frac{\frac{1}{p} - 1}{\frac{1}{P_0} - 1}$$

On remarque que si la proportion p est choisie de façon optimale, on obtient bien l'équivalence des variances asymptotiques.

2.4.5. Conclusion. — L'étude analytique des algorithmes particulaires a pu être menée grâce l'approche de Feynman-Kac développée dans [20], outil mathématique particulièrement bien adapté pour appréhender ce type de modèle. Cependant, seuls des résultats asymptotiques ont pu être obtenus, comme un résultat de loi forte des grands nombres, un théorème central limite. Il est aussi possible d'obtenir des principes de grandes déviations. Par contre, déterminer des intervalles de confiance est bien plus délicat. Toute la difficulté de l'étude de ce modèle provient directement de la dépendance entre les particules. D'où la nécessité d'introduire un modèle simplifié sans interaction entre les particules. Nous sommes guidés par le fait que les particules deviennent finalement «indépendantes» (propagation du chaos), voir [20]. Nous introduisons dans le chapitre suivant ce modèle simplifié et sans intéraction entre les particules afin d'établir des résultats précis, comme des bornes de Chernoff de l'estimateur, et d'avoir de meilleures connaissance et compréhension des algorithmes à trajectoires préférentielles généraux.

CHAPITRE 3

MODÈLE DE BRANCHEMENT AVEC DUPLICATION DES TRAJECTOIRES

Nous nous proposons, dans ce chapitre, d'introduire le modèle de branchement avec duplication des trajectoires, de présenter les premiers résultats concernant ce modèle ainsi que l'algorithme optimal obtenu par minimisation de la variance de l'estimateur sous une contrainte de coût. Les résultats de ce chapitre sont l'objet de l'article [43] publié dans la revue Probability in the Engineering and Informational Sciences.

3.1. Présentation du modèle

Le but est d'estimer la probabilité $\mathbb{P}(A)$ d'un événement rare A correspondant par exemple au dépassement d'un certain niveau L par un processus $X(t)$. Contrairement aux algorithmes d'échantillonnage préférentiel, dans les algorithmes multi-trajectoires préférentiels, le système évolue selon la mesure de probabilité initiale. Cette technique s'appuie sur l'hypothèse qu'il existe des états intermédiaires identifiables visités par le système plus souvent que l'événement rare lui-même : définissons une suite de $M+1$ ensembles B_i emboîtés

$$B_{M+1} \subset B_M \subset \cdots \subset B_2 \subset B_1$$

où B_{M+1} correspond à A. On obtient ainsi une partition de l'espace d'état en régions $B_i - B_{i+1}$ appelées *régions d'importance*. En général, ces ensembles sont définis grâce à une fonction Φ appelée *fonction d'importance* définie sur l'espace d'état et à valeurs dans $\mathbb{R}$ telle que pour tout i, $B_i = \{\Phi \leq T_i\}$ pour une certaine valeur T_i appelée *niveau* avec $T_1 \leq T_2 \leq .. \leq T_M \leq L$.

Précisons que nous ne considèrerons, dans cette thèse, que des modèles unidimensionnels comme introduits par Garvels [26] et dans la méthode RESTART exposée au chapitre 2.

Dans ce cas, nous avons la formule produit

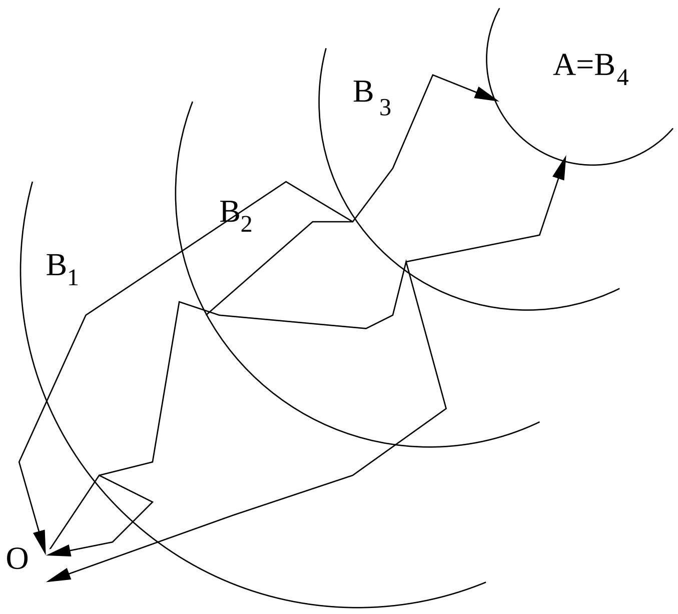


FIG. 1. Modèle de branchement avec duplication des trajectoires

$$\mathbb{P}(A) = \mathbb{P}(A|B_M)\mathbb{P}(B_M|B_{M-1})\cdots\mathbb{P}(B_2|B_1)\mathbb{P}(B_1). \tag{18}$$

La probabilité d'intérêt $\mathbb{P}(A)$ s'écrit ainsi comme le produit de $M+1$ quantités (probabilités conditionnelles) «moins rares» et donc plus faciles à estimer et avec plus de précision que $\mathbb{P}(A)$ elle-même, pour un coût de simulation donné.

Dans ce modèle, appelé *modèle de branchement avec duplication des trajectoires*, une apparition plus fréquente de A est réalisée en dupliquant le processus en R_i sous-processus dès qu'il entre dans une région B_i où sa chance d'atteindre l'événement rare est plus grande. On privilégie ainsi les trajectoires favorables. Un principe implicite nécessaire est que le niveau B_{i+1} ne peut être atteint depuis B_{i-1} sans que ne le soit aussi B_i, ce qui est bien le cas puisque les ensembles sont emboîtés.

Plus précisément, dans ce modèle et contrairement aux algorithmes particulaires, nous ne tenons pas compte de l'évolution du processus entre les niveaux mais seulement du franchissement des niveaux par ce dernier. L'idée de base consiste alors à faire un tirage de Bernoulli et vérifier si B_1 est atteint ou non. Si c'est le cas, on scinde le tirage en R_1 tirages de Bernoulli $\{0,1\}$ et on vérifie si B_2 est atteint ou non, et ainsi de suite. Plus précisément, chaque fois que l'événement B_i se produit, on simule R_i tirages et on renouvelle cette technique chaque fois que l'événement B_{i+1} se produit. Si aucun seuil supérieur, ni A, n'est atteint, on arrête le tirage concerné. En itérant de façon indépendante N fois cette procédure, on aura considéré $NR_1 \cdots R_M$ tirages, en tenant compte du fait que par exemple, si un niveau B_i à la i-ième étape n'a pas été atteint, les $R_i \cdots R_M$ tirages possibles ont échoué.

Les paramètres inconnus sont le nombre N de particules envoyées au départ, les nombres $R_1, .., R_M$ de retirage, les probabilités $P_1, .., P_{M+1}$ de transition et le nombre M de niveaux.

3.2. Estimateur de $\mathbb{P}(A)$

Un estimateur sans biais de $\mathbb{P}(A)$ est donné par

$$\widehat{P} = \frac{N_A}{N \prod_{i=1}^{M} R_i}, \tag{19}$$

où N_A est le nombre total de tirages ayant atteint A.

L'estimateur $\widehat{P}$ s'écrit aussi sous la forme

$$\widehat{P} = \frac{1}{NR_1..R_M} \sum_{i_0=1}^{N} \sum_{i_1=1}^{R_1} .. \sum_{i_M=1}^{R_M} \mathbf{1}_{i_0} \mathbf{1}_{i_0 i_1} .. \mathbf{1}_{i_0 i_1 .. i_M} \tag{20}$$

où $\mathbf{1}_{i_0 i_1 .. i_j}$ représente le résultat du j-ième tirage.

En fait, cet algorithme peut être représenté par N processus de branchement indépendants. Notons $P_i = \mathbb{P}(B_i | B_{i-1})$, $i = 2, \cdots, M+1$ et $P_1 = \mathbb{P}(B_1)$ alors l'identité (18) devient

$$\mathbb{P}(A) = P_1 \cdots P_{M+1}. \tag{21}$$

Considérons maintenant N processus de branchement indépendants $(Z_n^{(i)})_{n \geq 0}$, $i = 1, \cdots, N$ où pour chaque i, $Z_n^{(i)}$ est le nombre de particules issues de la

i-ième particule ($Z_0^{(i)}$=1) qui ont atteint le niveau B_n. Alors

$$\widehat{P} := \frac{1}{N}\sum_{i=1}^{N} \widetilde{Q}_i, \quad \text{où} \quad \widetilde{Q}_i = \frac{Z_{M+1}^{(i)}}{R_1 \cdots R_M}. \tag{22}$$

Pour alléger les notations on considère $N = 1$ dans la suite, et ainsi on considère le processus $(Z_n)_{n\geq 0}$ avec $Z_0 = 1$. On a donc

$$Z_{n+1} = \sum_{j=1}^{Z_n} X_n^{(j)} \tag{23}$$

où pour chaque n, les variables aléatoires $(X_n^{(j)})_{j\geq 1}$ sont i.i.d. de loi binomiale de paramètres (R_n, P_{n+1}) pour $n \geq 1$ et de loi de Bernoulli de paramètre P_1 pour $n = 0$.

Notons que l'on peut réécrire Z_{i+1} sous la forme

$$Z_{i+1} = \sum_{j=1}^{Z_i}\sum_{k=1}^{R_i} Y_i^{(j,k)}$$

où les $Y_i^{(j,k)}$ sont des lois de Bernoulli de paramètre P_{i+1}.

Introduisons les quantités suivantes

$$m_0 = P_1, \qquad m_n = R_n P_{n+1}, \quad n = 1, \cdots, M+1,$$

nombres moyens de particules réussissant à atteindre le niveau supérieur, ainsi que

$$r_0 = 1, \qquad r_n = R_1 \cdots R_n, \quad n = 1, \cdots, M,$$

et

$$p_0 = 1, \qquad p_n = P_1 \cdots P_n, \quad n = 1, \cdots, M+1,$$

Alors

$$\widetilde{Q} = \frac{Z_{M+1}}{R_1 \cdots R_M} = \mathbb{P}(A)\frac{Z_{M+1}}{m_0 \cdots m_M} \tag{24}$$

Selon (23), on a $\mathbb{E}(Z_{n+1}|Z_n) = m_n Z_n$, et en itérant $\mathbb{E}(Z_{n+1}) = \prod_{i=0}^{n} m_i$. En introduisant la variable aléatoire $W_n = Z_n/\mathbb{E}(Z_n)$ on obtient une nouvelle expression donnée par

$$\widetilde{Q} = \mathbb{P}(A) W_{M+1} \tag{25}$$

et on déduit aisément que l'estimateur $\widehat{P}$ est sans biais.

Remarque 3.2.1. — *Définissons maintenant la filtration $\mathcal{F}$ par la famille des $\mathcal{F}_i$ définis par $\mathcal{F}_i = \sigma(Z_j, \quad j = 1 \cdots i)$. Il est alors immédiat que $(W_n)_n$ est une martingale pour la filtration $\mathcal{F}$. De plus,*

$$\begin{aligned}
\mathbb{E}(Z_{i+1}|\mathcal{F}_i) &= \mathbb{E}(Z_{i+1}|Z_i) = \mathbb{E}(\sum_{j=1}^{Z_i}\sum_{k=1}^{R_i} Y_i^{(j,k)}|Z_i) \\
&= Z_i \sum_{k=1}^{R_i} \mathbb{E}(Y_i^{(j,k)}|Z_i) = Z_i R_i \mathbb{E}(Y_i^{(j,k)}) = Z_i R_i P_{i+1}
\end{aligned}$$

D'où l'on déduit que

- *$(Z_i)_i$ est une $\mathcal{F}$-martingale si $R_i P_{i+1} = 1$ pour tout $i = 1 \cdots M$,*
- *$(Z_i)_i$ est une $\mathcal{F}$-sur-martingale si $R_i P_{i+1} < 1$ pour tout $i = 1 \cdots M$,*
- *$(Z_i)_i$ est une $\mathcal{F}$-sous-martingale si $R_i P_{i+1} > 1$ pour tout $i = 1 \cdots M$.*

La relation $R_i P_{i+1} = 1$ joue donc un rôle déterminant. Lorsqu'elle est vérifiée, on a

$$\mathbb{E}(Z_i) = NP_1$$

qui signifie que le nombre moyen de succès à chaque niveau est constant.

Une autre relation intéressante est la suivante :

$$R_i P_i = 1$$

Dans ce cas, $\mathbb{E}(Z_i) = NP_i$; ce qui traduit le fait que le nombre moyen $R_i Z_i$ d'échantillons générés au niveau i est constant et égal à N. On travaille donc, dans ce cas-là, à taille moyenne de population constante à chaque niveau. La variable aléatoire $\widetilde{Z}_i = R_i Z_i$, qui représente le nombre (aléatoire) de particules générées au niveau i, est alors une martingale.

Rappelons que, dans le modèle des systèmes particulaires présenté en section 2.4, la taille de la population à chaque niveau est aussi constante.

3.3. Le modèle de Galton-Watson

Nous venons de voir que nous pouvons représenter l'algorithme de branchement avec duplication des trajectoires comme N processus de branchement de Galton-Watson, les seuils représentant les différentes générations. Nous rappelons dans cette partie quelques propriétés élémentaires des processus de Galton-Watson. Pour plus de détails voir, par exemple, Harris [37], Lyons [50] et Athreya et Ney [6].

Le modèle de Galton-Watson a été introduit par Galton et Watson dans les années 1870 afin d'étudier l'extinction des familles dans une population. On considère ainsi une population telle que les individus de l'ensemble initial ont des descendants qui constituent la 1ère génération, les descendants de ceux-ci forment la 2ième génération...et ainsi de suite. Ce processus est de nature aléatoire.

Ainsi défini, si Z_n représente le nombre d'individus à la n-ième génération et $Z_0 = 1$, alors

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_i^n,$$

où les X_i^n sont des v.a. i.i.d. qui représentent le nombre de descendants du i-ième individu à la n-ième génération.

Nous ferons les hypothèses suivantes :

- (Z_n) est une chaîne de Markov, dont les transitions sont indépendantes du temps.
- Le nombre de descendants d'un père est indépendant du nombre d'individus présents.

Définition 3.3.1. — *Si $\mathbb{P}(Z_1 = k) = p_k$ avec $\sum_k p_k = 1$, alors on définit la* fonction génératrice *de Z_1 par*

$$f(s) = \mathbb{E}(s^{Z_1}) = \sum_{k=0}^{\infty} p_k s^k \quad \textit{pour} \quad |s| \leq 1.$$

Les itérées de la fonction f sont alors

$$\begin{cases} f_0(s) = s \\ f_1(s) = f(s) \\ f_{n+1}(s) = f(f_n(s)). \end{cases}$$

Dans ce cas, $f_{n+m}(s) = f_m(f_n(s))$ et $f_{n+1}(s) = f_n(f(s))$.

Un résultat remarquable est le suivant :

Théorème 3.3.1. — *La fonction génératrice de Z_n est la n-ième itérée de f, la fonction génératrice de Z_1.*

On en déduit par récurrence les moments de Z_n, notamment l'espérance et la variance :

Lemme 3.3.1. — *Soit $m = \mathbb{E}(Z_1)$. Alors l'espérance de Z_n, $\mathbb{E}(Z_n)$ est m^n et sa variance est donnée par*

$$var(Z_n) = \begin{cases} \frac{m^n(m^n-1)}{m^2-m} var(Z_1) & si \;\; m \neq 1 \\ n var(Z_1) & si \;\; m = 1. \end{cases}$$

Si les moments d'ordres supérieurs de Z_1 existent alors il en est de même pour ceux de Z_n et ils sont obtenus de manière similaire, par récurrence.

L'introduction de la fonction génératrice permet surtout d'étudier la probabilité d'extinction du processus, c.-à-d. la probabilité de l'événement $\{Z_n \underset{n\to\infty}{\to} 0\}$. Cette probabilité d'extinction est alors donnée par

$$q := \mathbb{P}(Z_n \underset{n\to\infty}{\to} 0) = \mathbb{P}(Z_n = 0 \text{ pour un } n \in \mathbb{N}) = \lim_{n\to\infty} f_n(0).$$

On a le théorème suivant :

Théorème 3.3.2. — *Si $m \leq 1$, alors la probabilité d'extinction q vaut 1. Si $m > 1$, alors q est l'unique solution positive et inférieure à 1 de l'équation*

$$s = f(s). \tag{26}$$

La suite (Z_n) tend soit vers l'infini soit vers zéro ; comme le montre le résultat suivant, elle ne peut rester strictement positive et bornée, même lorsque $m = 1$.

Théorème 3.3.3. — *Quel que soit m fini, $\lim_{n\to\infty} \mathbb{P}(Z_n = k) = 0$, $k = 1, 2, \cdots$ De plus, $Z_n \underset{n\to\infty}{\to} \infty$ avec la probabilité $1 - q$ et $Z_n \underset{n\to\infty}{\to} 0$ avec la probabilité q.*

Nous dirons que le processus est *critique* si $m = 1$, *surcritique* si $m > 1$ et *sous-critique* si $m < 1$. Etudions le comportement asymptotique de Z_n.

***Théorème 3.3.4* (Cas surcritique)**. — *Notons $W_n = Z_n/m^n$ pour $n = 0, 1, \cdots$. Si $m > 1$ et $\mathbb{E}(Z_1^2) < \infty$, les variables aléatoires W_n convergent presque sûrement et en moyenne quadratique vers une variable aléatoire W telle que*

$$\mathbb{E}(W) = 1 \quad et \quad var(W) = \frac{var(Z_1)}{m^2 - m} > 0$$

***Théorème 3.3.5* (Cas sous-critique).** — *Si $m < 1$ et $\mathbb{E}(Z_1^2) < \infty$, pour tout $j = 1, 2, \cdots$, on a*

$$\lim_{n\to\infty} \mathbb{P}(Z_n = j | Z_n \neq 0) = b_j$$

existe et $\sum_{j=1}^{\infty} b_j = 1$. De plus, la fonction génératrice $g(s) = \sum_j b_j s^j$ vérifie l'équation fonctionnelle

$$gof(s) = mg(s) + 1 - m, \quad |s| \leq 1$$

Notons que dans le cas limite $m = 1$ du théorème 3.3.2 (cas critique), la population s'éteint avec probabilité 1 lorsque n tend vers l'infini. Nous avons en fait un résultat asymptotique plus précis :

***Lemme 3.3.2*.** — *Soit $m = 1$ et $f^{(3)}(1) < \infty$. Soit S l'ensemble des points étant : soit intérieurs au cercle unité, soit sur l'arc $-\theta_0 \leq arg(s) \leq \theta_0$ du cercle unité, point $s = 1$ exclu, pour un certain θ_0. On a alors*

$$\frac{1}{1 - f_n(s)} = \frac{1}{1-s} + \frac{nf''(1)}{2} + O(\log n), \quad s \in S, \quad n \to \infty.$$

En prenant $s = 0$ dans le lemme précédent, on en déduit la probabilité d'extinction quand n est grand :

$$\mathbb{P}(Z_n > 0) \sim \frac{2}{nf''(1)}.$$

Enfin, on obtient le théorème suivant décrivant la distribution de Z_n quand n est grand :

***Théorème 3.3.6* (Cas critique).** — *Soit $m = 1$ et $f^{(3)}(1) < \infty$. On a alors*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{2Z_n}{nf''(1)} > u | Z_n \neq 0\right) = e^{-u}, u \geq 0.$$

Ce résultat traduit bien le caractère d'événement rare lorsque le nombre de générations augmente.

3.4. Optimisation de l'algorithme

3.4.1. Etude de la variance. — Comme fait dans [60], on déduit l'expression de la variance de l'estimateur $\widehat{P}$ par récurrence.

Proposition 3.4.1. — *En utilisant les notations introduites en Section 3.2, il vient*

$$var(\widehat{P}) = \frac{\mathbb{P}(A)^2}{N}\Big[\sum_{i=0}^{M} \frac{1}{r_i}\left(\frac{1}{p_{i+1}} - \frac{1}{p_i}\right)\Big] \tag{27}$$

Avec les notations précédentes, on peut réécrire la variance sous la forme

$$\mathrm{var}(\widehat{P}) = \frac{\mathbb{P}(A)^2}{N}\sum_{i=0}^{M}\frac{1-P_{i+1}}{m_0\cdots m_i} \tag{28}$$

Démonstration. — Tout d'abord, $\mathrm{var}(\widehat{P}) = \frac{\mathbb{P}(A)^2}{N}\,\mathrm{var}(W_{M+1})$. Calculer $\mathrm{var}(\widehat{P})$ revient donc à calculer $\mathrm{var}(W_{M+1})$.

Notons que pour toute variable aléatoire X et $\mathcal{F}$ filtration,

$$\mathrm{var}(X) = \mathrm{var}(\mathbb{E}(X|\mathcal{F})) + \mathbb{E}(\mathrm{var}(X|\mathcal{F})) \tag{29}$$

En appliquant (66) à $X = W_{M+1} = \frac{Z_{M+1}}{P_1\prod_{i=1}^{M} R_i P_{i+1}}$ et $\mathcal{F} = \sigma(Z_M)$ et puisque W_M est une martingale, on obtient

$$\begin{aligned}
\mathrm{var}(W_{M+1}) &= \mathrm{var}(W_M) + \frac{1}{[P_1\prod_{i=1}^{M} R_iP_{i+1}]^2}\mathbb{E}(\mathrm{var}(Z_{M+1}|\sigma(Z_M)))\\
&= \mathrm{var}(W_M) + \frac{1}{[P_1\prod_{i=1}^{M} R_iP_{i+1}]^2}\mathbb{E}(Z_M\,\mathrm{var}(X_M|\sigma(Z_M)))\\
&= \mathrm{var}(W_M) + \frac{R_MP_{M+1}(1-P_{M+1})}{[P_1\prod_{i=1}^{M} R_iP_{i+1}]^2}\mathbb{E}(Z_M)\\
&\qquad\qquad\text{puisque } X_M^i \sim Bin(R_M, P_{M+1})\\
&= \mathrm{var}(W_M) + \frac{1-P_{M+1}}{P_1\prod_{i=1}^{M} R_iP_{i+1}}\mathbb{E}(W_M)
\end{aligned}$$

Par une récurrence descendante,

$$\mathrm{var}(W_{M+1}) = \mathrm{var}(W_0) + \sum_{k=0}^{M}\frac{1-P_{k+1}}{P_1\prod_{i=1}^{k} R_iP_{i+1}}\mathbb{E}(W_k)$$

Il reste donc à calculer l'espérance de W_k mais comme W_k est une martingale d'après la Remarque 3.2.1,

$$\mathbb{E}(W_M) = \mathbb{E}(W_{M-1}) = \cdots = \mathbb{E}(W_0) = 1$$

D'où on déduit le résultat. □

Remarque 3.4.1. — *Le principe de la récurrence a une interprétation concrète : si dans une simulation avec n seuils, les retirages générés au premier niveau ne sont pas pris en compte sauf un que l'on appellera* trajectoire principale, *on obtient une simulation avec $n-1$ seuils.*

Remarque 3.4.2. — *On déduit de cette preuve, pour le cas où $R_k = R$ et $P_k = P_0$ pour tout k, que*

$$var(W_{M+1}) \quad = \quad \begin{cases} \left(\frac{1}{P_0} - 1\right) \frac{m^{M+1}-1}{(m-1)m^M} =: \sigma^2_{M+1} \quad pour \quad m \neq 1 \\ \left(\frac{1}{P_0} - 1\right)(M+1) \quad pour \quad m = 1 \end{cases}$$

On peut aussi déduire ce résultat du lemme 3.3.1.

3.4.2. Etude du coût. — Notre but est de minimiser la variance pour un budget fixé, et d'en déduire les valeurs optimales de N,R_1,..,R_M, $P_1, .., P_{M+1}$ et M. Par conséquent, il s'agit de décrire le coût d'une simulation : chaque fois qu'une particule est générée, cela entraîne un coût moyen décrit par une fonction h. On suppose que

- le coût h d'une particule pour atteindre B_i partant de B_{i-1} dépend seulement de P_i (et non du niveau de départ ni du point de départ),
- h est croissante en x (ce qui signifie que plus la probabilité de transition est petite, plus difficile est la transition et par conséquent plus le coût est élevé),
- h est positive,
- h converge vers une constante (en général petite) quand x tend vers 1.

Le coût (moyen) est ainsi donné par la formule

$$C = \mathbb{E}(Nh(P_1) + R_1N_1h(P_2) + R_2N_2h(P_3) + .. + R_MN_Mh(P_{M+1})) \tag{30}$$

où N_i est le nombre de tirages qui ont atteint le niveau i. Finalement,

Proposition 3.4.2. — *En utilisant les notations introduites en Section 3.2, il vient*

$$C = N \sum_{i=0}^{M} h(P_{i+1})r_ip_i \tag{31}$$

Exemple 3.4.1. — *Etudions la marche aléatoire simple sur $\mathbb{Z}$ partant de 0 que l'on tue dès qu'elle atteint -1 ou k (succès si elle atteint k, échec sinon). Posons donc X_n tel que $X_0 = 0$ et $X_n = \sum_{i=1}^{n} Y_n$ où $\{Y_n\}$ est une suite de variables aléatoires sur $\{-1, 1\}$ avec $\mathbb{P}(Y_n = 1) = \mathbb{P}(Y_n = -1) = \frac{1}{2}$ et définissons $T_k = \inf\{n \geq 0 \ : \ X_n = -1 \ ou \ k\}$.*

On peut vérifier facilement que X_n et $X_n^2 - n$ sont des martingales. D'après le théorème d'arrêt de Doob, $\mathbb{E}(X_{T_k}) = 0$ et $\mathbb{E}(X_{T_k}^2) = \mathbb{E}(T_k)$, ce qui conduit à

$$p := \mathbb{P}(X_{T_k} = k) = \frac{1}{k+1} \quad et \quad \mathbb{E}(T_k) = k = \frac{1}{p} - 1$$

c-à-d le coût nécessaire pour atteindre le niveau supérieur est $\frac{1}{p} - 1$ si p est la probabilité de succès.

3.4.3. Minimisation de la variance à coût fixé. — Pour minimiser la variance de $\widehat{P}$, nous procédons en trois étapes :

1. On optimise la variance en $N, R_1, .., R_M$ à $P_1, .., P_{M+1}$ fixés (c-à-d aux niveaux B_i fixés).
2. On injecte ces valeurs optimales dans la variance et on déduit les valeurs optimales des P_i pour $i = 1 \cdots M+1$ sous la contrainte $\mathbb{P}(A) = P_1 \cdots P_{M+1}$.
3. On injecte ces valeurs optimales dans la variance et on déduit la valeur optimale de M, nombre de seuils.

1. Valeurs optimales de $N, R_1, .., R_M$.

En utilisant la méthode des multiplicateurs de Lagrange, on obtient

$$R_i = \frac{r_i}{r_{i-1}} = \sqrt{\frac{h(P_i)}{h(P_{i+1})}} \sqrt{\frac{1}{P_i P_{i+1}}} \sqrt{\frac{1 - P_{i+1}}{1 - P_i}} \quad i = 1, .., M \tag{32}$$

$$N = \frac{1}{\sqrt{h(P_1)}} \frac{C\sqrt{1/P_1 - 1}}{\sum_{i=1}^{M+1} \sqrt{h(P_i)} \sqrt{\frac{1}{P_i} - 1}} \tag{33}$$

2. Valeurs optimales de $P_1, .., P_{M+1}$.

Ainsi la variance devient

$$\mathrm{var}(\widehat{P}) = \frac{\mathbb{P}(A)^2}{C} \Big[\sum_{i=1}^{M+1} \sqrt{h(P_i)} \sqrt{\frac{1}{P_i} - 1} \Big]^2$$

En procédant de la même façon que précédemment sous la contrainte $\mathbb{P}(A) = P_1..P_{M+1}$, on obtient que les P_i vérifient

$$2\sqrt{C}\lambda \sqrt{h(x)(\frac{1}{x} - 1)} = h'(x)(1-x) - \frac{h(x)}{x}$$

Si nous supposons l'existence d'une unique solution à cette équation, on a $P_i = g(\lambda)$, d'où $\mathbb{P}(A) = g(\lambda)^{M+1}$ et $g(\lambda) = \mathbb{P}(A)^{\frac{1}{M+1}}$. Finalement,

$$P_i = \mathbb{P}(A)^{\frac{1}{M+1}} \quad i = 1, .., M+1 \tag{34}$$

3. Valeur optimale de M.

Les valeurs optimales des $P_1, .., P_{M+1}$ entraînent que les R_i optimaux sont égaux à $1/P_i = 1/\mathbb{P}(A)^{1/(M+1)}$, $i = 1, .., M$ et donc

$$\mathrm{var}(\widehat{P}) = \frac{\mathbb{P}(A)^2}{C}(M+1)^2 h(\mathbb{P}(A)^{1/M+1})(\mathbb{P}(A)^{-1/M+1} - 1)$$

qu'il s'agit de minimiser en M. Remarquons que $R_i P_i = 1$. Posons

$$f(M) = \frac{\mathbb{P}(A)^2}{C}(M+1)^2 h(\mathbb{P}(A)^{1/M+1})(\mathbb{P}(A)^{-1/M+1} - 1),$$

dont la dérivée s'annule en

$$F(y) := (2(1-e^y) + y)h(e^y) - y(1-e^y)e^y h'(e^y) = 0, \text{ avec } y = \frac{\ln \mathbb{P}(A)}{M+1} \tag{35}$$

En général, on n'obtient pas un entier. On a $y_0 = \frac{\ln \mathbb{P}(A)}{M+1}$ c-à-d $M+1 = [\frac{\ln \mathbb{P}(A)}{y_0}]$ ou $[\frac{\ln \mathbb{P}(A)}{y_0}] + 1$. Posons $\frac{\ln \mathbb{P}(A)}{y_0} = n + x$ avec $0 < x < 1$. Alors
- si $M+1 = n$, $y = \frac{\ln \mathbb{P}(A)}{n}$,
- si $M+1 = n+1$, $y = \frac{\ln \mathbb{P}(A)}{n+1}$.

et la valeur du rapport $\rho := \frac{f(n-1)}{f(n)}$ donne le choix optimal pour M :
- si ρ <1, $M = n-1$,
- si ρ >1, $M = n$.

Ainsi le nombre optimal de niveaux est donné par $M = [\frac{\ln \mathbb{P}(A)}{y_0}] - 1$ ou $M = [\frac{\ln \mathbb{P}(A)}{y_0}]$ avec y_0 solution de $F(y) = 0$. Alors M minimise

$$\mathrm{var}(\widehat{P}) = \frac{\mathbb{P}(A)^2}{C}\left[\ln \mathbb{P}(A)\right]^2 y^{-2} h(e^y)(e^{-y} - 1)$$

Exemple 3.4.2. — *Pour $h = 1$, on doit résoudre $y = 2(e^y - 1)$. On obtient $y_1 = 0$ et $y_2 \approx -1.5936$. y_2 est un minimum et la valeur optimale de M est*

$$M = [-0.6275 \ln P] - 1 \quad ou \quad [-0.6275 \ln \mathbb{P}(A)] \tag{36}$$

Avec $\mathbb{P}(A) = 10^{-k}$,

k	n	rapport(ρ) >1,<1	M	k	n	rapport(ρ) >1,<1	M
1	1	>	1	6	8	>	8
2	2	>	2	9	13	<	12
3	4	<	3	12	17	<	16
4	5	>	5	15	21	>	21
5	7	<	6	18	26	<	25

Notons que M croît lorsque $\mathbb{P}(A)$ décroît et avec cette valeur de M, chaque R_i et P_i deviennent

$$R_i \approx 5 \quad \text{et} \quad P_i \approx \frac{1}{5}. \tag{37}$$

Donc les nombres optimaux de retirage et les probabilités de transition optimales sont «indépendants» de la probabilité de l'événement rare.
De plus, asymptotiquement, $M = n = [\frac{\ln \mathbb{P}(A)}{y_0}] - 1$, donc

$$P_i = \mathbb{P}(A)^{\frac{1}{M+1}} = e^{\frac{\ln \mathbb{P}(A)}{M+1}} = e^{y_0} \quad \text{et} \quad \mathbb{P}(A) = e^{-(n+1)|y_0|}$$

Décrivons maintenant brièvement les solutions possibles de (35). Rappelons que nous voulons résoudre (35) c-à-d si $z = e^y$ et $z \neq 0, 1$,

$$\frac{h'(z)}{h(z)} = \frac{1}{z}(\frac{2}{\ln z} + \frac{1}{1-z})$$

En notant, $H(z) = \frac{h'(z)}{h(z)}$ et $L(z) = \frac{l'(z)}{l(z)} = \frac{1}{z}(\frac{2}{\ln z} + \frac{1}{1-z})$, il s'agit de résoudre

$$H(z) = L(z) \tag{38}$$

Tout d'abord, soit z_0 la solution de $2(z-1) = \ln z$. Puisque $h' \leq 0$, H est négatif et une étude rapide permet de montrer que L est positif sur $]0, z_0[$ et négatif sur $]z_0, 1[$. Par conséquent, les solutions de (38) sont dans $]z_0, 1[$, s'il en existe. Ainsi résoudre (35) est équivalent à étudier les intersections entre H et L. Une étude rapide de ces fonctions montre qu'il y a deux cas à distinguer (voir Figure 2) :

– cas 1 : nombre impair d'intersections entre L et H

$$\begin{aligned} &\Leftrightarrow \quad H(z) > L(z) \quad \text{près de} \quad 1 \\ &\Leftrightarrow \quad h''(1) < 0 \end{aligned}$$

– cas 2 : nombre pair d'intersections ou aucune entre L et H

Notons que $y = 0$ est une solution de (35). Dans le cas 1, cette solution correspond à un maximum et dans le cas 2, à un minimum. Et le second cas est exclu puisque nous avons fait l'hypothèse $h(1) > 0$.

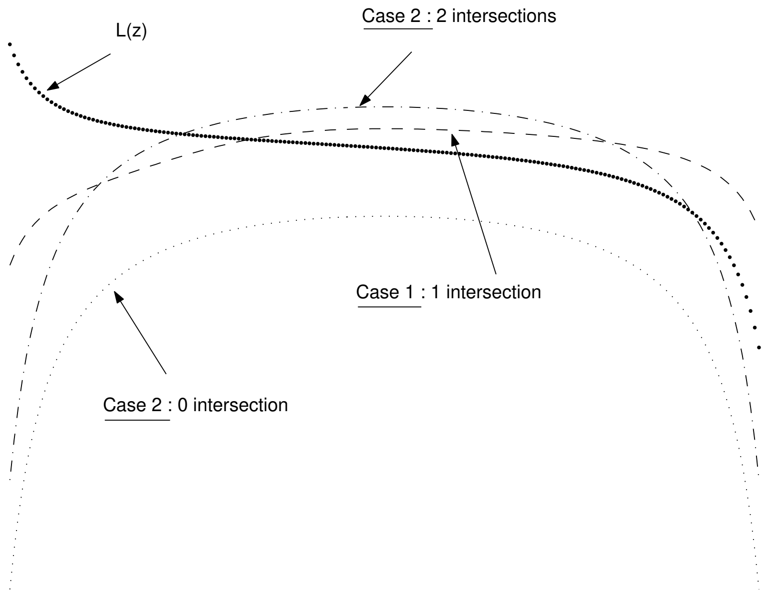


FIG. 2. Behavior of H and L

Remarque 3.4.3. — *La solution $y = 0$ correspond aux valeurs optimales suivantes*

$$M = \infty, \quad P_i = 1, \quad R_i = 1, \quad N \underset{M\to\infty}{\sim} \frac{C}{(M+1)h(1) + \ln(\mathbb{P}(A))h'(1)}$$

Mais $P_i = 1$ implique que $\mathbb{P}(A) = 1$ et $R_i = 1$ signifie que l'on réalise une simulation simple, sans retirage.

Exemple 3.4.3. — *Ici $\mathbb{P}(A) = 10^{-12}$ et $C = 10^4$.*
1) Dans l'exemple 3.4.1, $h(1) = 0$ et nous sommes dans le cas 2 : l'unique solution $y = 0$ est un minimum.
2) Posons $h(x) = \frac{1}{x} - 8x^2 + 12x - 5$. $h(1) = 0$ et nous sommes dans le cas 1 : $y = 0$ et $y \approx -0.9919$ sont les solutions. $y = 0$ est le maximum et l'autre solution le minimum. Prenons $y \approx -0.9919$, on obtient alors

$$M = 26, \quad P_0 \approx 0.3594, \quad R \approx 2.7826 \quad et \quad N \approx 22.9$$

et on peut prendre $R = 3$ et $N = 23$.
3) Prenons $h(x) = (\frac{1}{x} - 1)^2 e^{6x}$. $h(1) = 0$ et nous sommes dans le cas 2 : $y = 0$, $y_1 \approx -0.4612$ et $y_2 \approx -0.5645$ sont les solutions. $y = 0$ est le minimum et la seconde solution le maximum.
4) Posons $h(x) = \frac{1}{x}$. Ici $h(1) = 1$. On souhaite résoudre (35) dont les solutions

sont $y = 0$ et $y \approx -0.6438$. Prenons $y \approx -0.6438$, on obtient

$$M = 41, \quad P_0 \approx 0.5179, \quad R \approx 1.9307 \quad et \quad N \approx 34.5$$

et on peut prendre $R = 2$ et $N = 34$.

Remarque 3.4.4. — *Bien que nous aimerions prendre R_i tel que $R_i P_i = 1$, nous sommes contraints à choisir R_i entier et positif. Dans la suite de ce chapitre, nous supposerons que nous sommes dans le cas favorable où $R_i = 1/P_i$ est un entier.*

3.4.4. Remarques sur l'algorithme optimal. — Tout d'abord, notons que les valeurs optimales pour les $(R_i)_i$: $R_i = \frac{1}{P_0} := R$ et $(P_i)_i$: $P_i = P_0$ entraînent que

$$R_i P_{i+1} = 1$$

Ce résultat n'est pas surprenant puisqu'il signifie que le processus de branchement obtenu est un Galton-Watson critique et traduit un certain équilibre entre un nombre de répliques trop important qui entraînerait l'explosion du coût et un nombre de répliques insuffisant conduisant à un arrêt prématuré de l'algorithme.

Ensuite, avec ce choix optimal des paramètres, (Z_n) est une martingale (cf. : Remarque 3.2.1).

(Z_n) est aussi une chaîne de Markov de probabilités de transition

$$P(i,j) = \mathbb{P}(Z_{n+1} = j | Z_n = i) = \begin{cases} q_j^{\star i} & i \geq 1, \quad \leq 0 \\ \delta_{0j} & i = 0, \quad j \geq 0 \end{cases}$$

où $(q_k^{\star i})_k$ est la convolée n-ième de q_k, q_k étant définie par

$$q_k = \mathbb{P}(\mathrm{Bin}(R, P_0) = k).$$

Rappelons que la n-ième convolée de q_k est la loi de la somme de i variables indépendantes chacune régie par les probabilités $(q_k)_k$.

Enfin, lorsque $h \equiv 1$, la variance a alors l'expression suivante

$$\mathrm{var}(\widetilde{P}) = \frac{\mathbb{P}(A)^2}{N}(M+1)\left(\frac{1}{P_0} - 1\right)$$

et nous remarquons que, comme annoncé en section 2.4, la variance du modèle de branchement avec duplication des trajectoires a la même expression que la variance asymptotique des systèmes particulaires. C'est ce résultat qui avait motivé l'introduction du modèle étudié dans cette thèse.

3.5. Applications

3.5.1. File d'attente. — Afin d'analyser le comportement des différentes méthodes décrites ci-dessus, on les compare sur un exemple concret : on considère une file d'attente $M/M/1$ et on veut estimer la probabilité que la queue atteigne la longueur $L_0 = 150$. Les résultats sont présentés en Figure 3. Comme

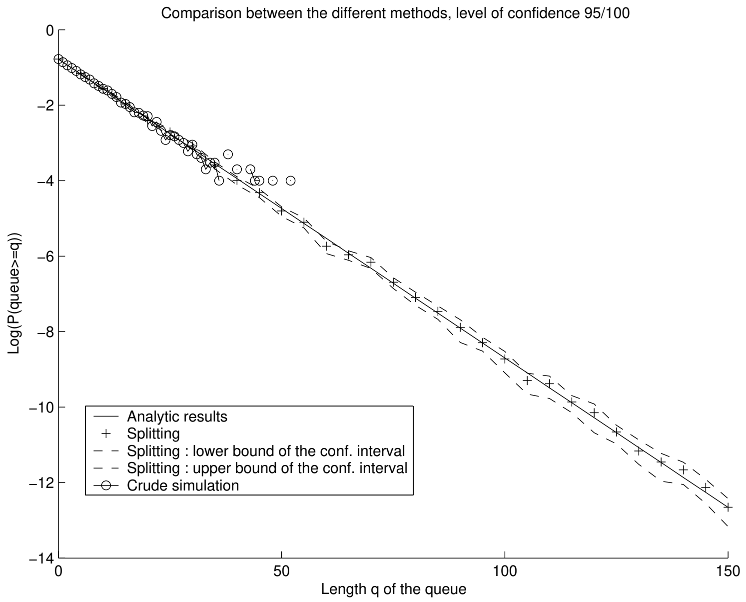


FIG. 3. Comparaison entre les différentes méthodes-Files d'attente

on s'y attendait et puisque on travaille à coût fixé C ($C = 10^4$), la simulation simple s'arrête après quelques itérations, faute d'un nombre suffisant d'échantillons. Par contre, le modèle de branchement avec duplication des trajectoires permet d'estimer la probabilité recherchée et les résultats sont proches des résultats théoriques.

3.5.2. Approximate counting. — Ce modèle peut aussi être appliqué à l'«approximate counting». Voir [40] et [22] pour plus de détails.

Le but est d'estimer le nombre de solutions du problème du sac à dos («Knapsack problem») c-à-d d'estimer le cardinal de Ω défini par

$$\Omega := \{x \in \{0,1\}^n \ : \ a.x := \sum_{i=1}^{n} a_i x_i \leq b\}$$

pour un vecteur réel et à coordonnées positives $a = (a_i)_{i=1}^n$ et b un nombre réel.

On applique l'algorithme appelé Markov Chain Monte Carlo (MCMC) (voir [52] et Annexe 3 pour plus de détails) : on construit une chaîne de Markov $\mathcal{M}_{Knap}$ d'espace d'état Ω et de transitions d'un état $x = (x_1, .., x_n) \in \Omega$ définies par

– avec probabilité $\frac{1}{2}$ posons $y = x$;
– sinon choisissons aléatoirement i uniformément dans $\{1, .., n\}$ et posons $y' = (x_1, .., x_{i-1}, 1 - x_i, x_{i+1}, .., x_n)$ Si $a.y' \leq b$ alors posons $y = y'$, sinon posons $y = x$.

Le nouvel état est y. Cette marche aléatoire sur l'hypercube tronqué par l'hyperplan $a.x = b$ converge vers la loi uniforme sur Ω. Ceci suggère une procédure pour sélectionner aléatoirement les solutions Knapsack «quasi-uniformément». Partant de $(0, .., 0)$, on simule $\mathcal{M}_{Knap}$ pendant un nombre suffisant de pas de façon à ce que la loi sur les états soit «proche» [(1)] de la loi uniforme, alors on renvoie l'état atteint. Bien sûr générer dans Ω n'est pas estimer la taille de Ω. Mais cela y conduit.

Gardons le vecteur a fixé mais autorisons b à varier. Notons $\Omega(b)$ et $\mathcal{M}_{Knap}(b)$ au lieu de Ω et $\mathcal{M}_{Knap}$ pour insister sur cette dépendance en b. Supposons sans perte de généralité que $a_1 \leq .. \leq a_n$ et définissons

$$b_1 = 0 \ \text{ et } \ b_i = \min\{b, \sum_{i=1}^{i-1} a_j\}.$$

On peut facilement vérifier que

$$|\Omega(b_{i-1})| \leq |\Omega(b_i)| \leq (n+1)|\Omega(b_{i-1})|$$

[(1)] Le problème est de borner le nombre de pas nécessaires pour que la chaîne de Markov $\mathcal{M}_{Knap}(b)$ soit «proche» de la stationnarité. Plus précisément, nous recherchons une borne du *«mixing time»* :

$$\tau_{mix}(\nu) := \min\{t : \Delta_x(t') \leq \nu \text{ pour tout } t' \geq t\}$$

où $\Delta_x(t) = \max_{S \subset \Omega} |P^t(x, S) - \Pi(S)|$ et Π est la loi stationnaire. Dans [40], il est montré que $\mathcal{O}(n^{9/2+\nu})$ pas suffisent.

Maintenant écrivons

(39) $|\Omega(b)| = |\Omega(b_{n+1})| = \dfrac{|\Omega(b_{n+1})|}{|\Omega(b_n)|}\dfrac{|\Omega(b_n)|}{|\Omega(b_{n-1})|}..\dfrac{|\Omega(b_2)|}{|\Omega(b_1)|}|\Omega(b_1)| := \rho_n^{-1}..\rho_1^{-1}$

Chaque rapport $\rho_i = \frac{|\Omega(b_i)|}{|\Omega(b_{i+1})|}$ peut être estimé en générant «quasi-uniformément» dans $\Omega(b_{i+1})$, en utilisant la chaîne de Markov $\mathcal{M}_{Knap}(b_{i+1})$ et en calculant la fraction de particules ayant atteint $\Omega(b_i)$.

Application numérique :

Prenons par exemple $a = [1, 2, 3, 4]$, $b = 3$, $h = 1$, $R = 5$ et $C = 2600$.
On choisit les niveaux comme suggéré précédemment : on définit d'abord $b_1 = 0$, $b_2 = 1$, $b_3 = 3$, $b_4 = 3$ et $b_5 = b$, puis $B_0 = \Omega$, $B_1 = \Omega(b_4)$, $B_2 = \Omega(b_3)$, $B_3 = \Omega(b_2)$ et $B_4 = \Omega(b_1)$.
Ainsi ici $M = n - 1$, $N = C/n$ et $n_{step} = 1020$.

De façon évidente, $\mathrm{Card}(\Omega) = 5$.

On réalise trois simulations différentes : d'abord la simulation simple, ensuite celle suggérée dans [40] et consistant à estimer les n rapports de façon indépendante et enfin celle basée sur le modèle de branchement avec duplication des trajectoires. On obtient les estimations suivantes de $\mathrm{Card}(\Omega)$:
- estimation par simulation simple = 4.088,
- estimation par les n ratios = 5.44,
- et estimation par branchement avec duplication des trajectoires = 5.019.

Bien que les niveaux n'ont pas été choisis de façon optimale, le modèle de branchement avec duplication des trajectoires apporte une amélioration.

3.5.3. Comparaison du modèle avec l'échantillonnage préférentiel. — On étudie, à nouveau, une file d'attente M/M/1 de paramètre d'arrivée λ et de paramètre de départ μ tels que $\lambda < \mu$ (puisque dépasser L_0 doit être rare). Ainsi les probabilités d'arrivée et de départ sont données par

$$\begin{cases} \mathbb{P}(\text{ arrivée }) = \frac{\lambda}{\lambda+\mu} \\ \mathbb{P}(\text{ départ }) = \frac{\mu}{\lambda+\mu} \end{cases}$$

Nous étudions dans cette section la probabilité que la longueur de la file dépasse la longueur L_0 avant d'être nulle et souhaitons comparer les modèles d'échantillonnage préférentiel et de branchement avec duplication des trajectoires.

D'après [38], le changement optimal dans le cadre de la file d'attente M/M/1 est donné par

$$\begin{cases} \lambda^* = \mu \\ \mu^* = \lambda \end{cases}$$

On étudie N files qui partent de 1. On les fait évoluer selon les paramètres d'arrivée et de départ λ^* et μ^* respectivement et l'estimateur sans biais de la probabilité recherchée est

$$\widehat{P}_{IS} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{Y_i \geq L_0} L(Y_i)$$

Posons

$$\begin{cases} p_\lambda = \frac{\lambda^*}{\lambda^*+\mu^*} \frac{\lambda+\mu}{\lambda} \\ p_\mu = \frac{\mu^*}{\lambda^*+\mu^*} \frac{\lambda+\mu}{\mu} \end{cases}$$

Le rapport de vraisemblance est actualisé à chaque nouvel événement dans le cycle par $L = L * p_\lambda$ ou $L * p_\mu$ selon que cet événement correspond à une arrivée ou un départ.

Pour la simulation avec le modèle de branchement avec duplication des trajectoires, on procède comme à l'accoutumée. On étudie N files qui partent de 1. Dès qu'une file atteint le seuil supérieur avant de retourner en 0, elle est dupliquée en R sous-files qui continuent d'évoluer à partir de ce seuil et ainsi de suite. L'estimateur sans biais de la probabilité recherchée est donné par

$$\widehat{P}_{Splitt} = \frac{1}{N} \sum_{i_0=1}^{N} \frac{1}{R^M} \sum_{i_1=1}^{R} .. \sum_{i_M=1}^{R} \mathbf{1}_{i_0} \mathbf{1}_{i_0 i_1} .. \mathbf{1}_{i_0 i_1 .. i_M}$$

où $\mathbf{1}_{i_0 i_1 .. i_j}$ représente le résultat du j-ième tirage (c-à-d il vaut 1 si on atteint B_j, 0 sinon).

Pour $L_0 = 20$ et en plaçant les seuils de 2 en 2, les résultats obtenus pour différentes valeurs de ρ et pour des coûts équivalents sont consignés dans les tableaux suivants.

Pour $\rho = 0.4$, $\mathbb{P} = 1.649267e - 08$ et pour C de l'ordre de $4e + 07$,

	IS	Splitting
Estimateur	$1.650651e-08$	–
Borne inf IC	$1.648443e-08$	–
Borne sup IC	$1.652859e-08$	–
Erreur relative	0.000839	–
Largeur rel. IC	0.002677	–

Pour $\rho = 0.6$, $\mathbb{P} = 2.437528e-05$ et pour C de l'ordre de $4e+07$,

	IS	Splitting
Estimateur	$2.434702e-05$	$2.230000e-05$
Borne inf IC	$2.429808e-05$	$5.853636e-06$
Borne sup IC	$2.439597e-05$	$2.522682e-05$
Erreur relative	0.001159	0.085139
Largeur rel. IC	0.004016	0.240146

Pour $\rho = 0.8$, $\mathbb{P} = 0.002916$ et pour C de l'ordre de $8e+06$,

	IS	Splitting
Estimateur	0.002909	0.00306500000000
Borne inf IC	0.002887	0.002957
Borne sup IC	0.002929	0.003173
Erreur relative	0.002765	0.051125
Largeur rel. IC	0.014195	0.074311

Commentaires : Tout d'abord, plus les probabilités de départ et d'arrivée sont éloignées de 1/2 et plus l'échantillonnage préférentiel sera efficace. Ensuite, l'avantage du modèle de branchement avec duplication des trajectoires est qu'en introduisant les niveaux intermédiaires, on réduit la longueur de chaque file et par conséquent le coût nécessaire pour atteindre A. On remarque enfin que l'échantillonnage préférentiel semble efficace pour toute valeur de ρ alors que la précision du modèle de branchement avec duplication des trajectoires se dégrade lorsque ρ décroit. Mais rappelons que l'échantillonnage préférentiel nécessite une très bonne connaissance du processus étudié et peut s'avérer par conséquent, délicat à appliquer.

3.5.4. Un exemple sur les arbres. — L'espace d'état est ici un arbre enraciné et les particules se déplacent de la racine vers les feuilles. Pour chaque noeud v de l'arbre, on associe à chacune de ses branches v_1, .., v_m les probabilités $p(v_1|v)$, .., $p(v_m|v)$; la particule issue de v choisit alors son chemin selon ces probabilités. Le but de l'algorithme est de trouver un noeud profond dans l'arbre c-à-d à la profondeur au moins d. En tronquant à la profondeur d, le problème devient alors "quelle est la probabilité que le plus profond noeud trouvé par l'algorithme soit à la profondeur maximale d ?"

De la même façon que dans les modèles étudiés précédemment, les probabilités de transition $p(v_i|v)$ sont inconnues ; par contre, étant donné un noeud v de l'arbre, on sait choisir une branche de v selon la loi de probabilité précédente.

Décrivons maintenant l'algorithme du modèle de branchement avec duplication des trajectoires. On part à l'étape 0 avec B particules à la racine. A l'étape i, il y a un nombre aléatoire de particules toutes en un noeud de l'arbre à la profondeur i. Si toutes les particules sont en des feuilles, on arrête l'algorithme. Sinon, à chaque particule en non-feuille, on ajoute $R_i - 1$ autres particules. On poursuit ensuite l'algorithme en choisissant aléatoirement (c-à-d selon les probabilités conditionnelles $p(.|v)$) pour chacune de ces particules une branche.

Le coût d'une telle simulation est donné par

$$C = N\Bigg[\sum_{i_1=1}^{r} p(v_{1i_1}|v_1) + R_1 \sum_{i_1=1}^{r}\sum_{i_2=1}^{r} p(v_{1i_1i_2}|v_{1i_1})p(v_{1i_1}|v_1)$$
$$+ \cdots + R_1\cdots R_{d-1}\sum_{i_1=1}^{r}\cdots\sum_{i_d=1}^{r} p(v_{1i_1\cdots i_d}|v_{1i_1\cdots i_{d-1}})\cdots p(v_{1i_1}|v_1)\Bigg]$$

Soit $x \in [0,1]$. Pour illustrer cet algorithme, on construit un arbre, représenté en Figure 4 et composé de 18 noeuds tel que

– chaque noeud a au plus trois branches,

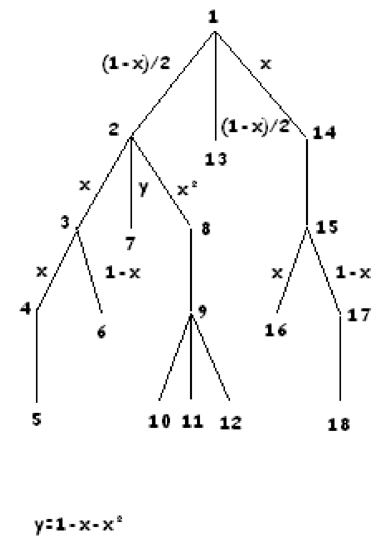


FIG. 4. Arbre considéré dans l'exemple

– la probabilité pour une particule au noeud i d'aller vers une des feuilles de ce noeud est donnée par la ligne i de la matrice P (dépendant de x) suivante :

$$P = \begin{bmatrix} (1-x)/2 & (1-x)/2 & x \\ x & 1-x-x^2 & x^2 \\ x & 1-x & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1-x & x & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

On calcule alors analytiquement la probabilité $\mathbb{P}(A)$ d'atteindre une feuille de profondeur maximale :

$$\mathbb{P}(A) = x^2(2 - x)$$

et en prenant, par exemple, $x = 10^{-3}$, on obtient

$$\mathbb{P}(A) = (2 - 10^{-3})10^{-6} \approx 2\ 10^{-6}$$

La simulation pour des coûts équivalents donne

- simulation simple : $\mathbb{P}(A)$ est estimé par $P_s = 2.5\ 10^{-6}$ (erreur de $5\ 10^{-7}$) avec un intervalle de confiance à 95% de longueur approchée $7\ 10^{-6}$.
- modèle de branchement avec duplication des trajectoires : $\mathbb{P}(A)$ est estimé par $P_{splitt} = 2\ 10^{-6}$ (erreur de 10^{-9}) avec un intervalle de confiance à 95% de longueur approchée $3\ 10^{-6}$.

Application 3.5.1. — *La discussion suivante (simplifiée et inexacte) fait ressortir l'intérêt de cette étude d'exploration d'arbre.*
Etant donnée une fonction f définie sur un espace d'état S, il existe une façon naturelle de définir un arbre tel que ses feuilles correspondent aux minima locaux de f. Chaque branche de l'arbre à la profondeur (hauteur) h correspond à une composante connexe de $\{s\ :\ f(s) \leq h\}$. Voir Figure 5.

Si on utilise un algorithme de Métropolis pour simuler la loi

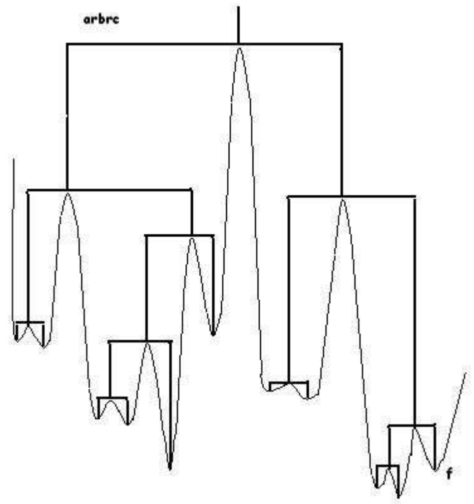


FIG. 5. Correspondance entre arbre et problème du recuit simulé

$$\pi_T(s) = c_T \exp\{-f(s)/T\},$$

on obtient en réalité la loi π_T *restreinte à la composante connexe* $\{s : f(s) \leq h(T)\}$ *qui contient le point de départ pour une certaine fonction* $h(T)$. *Identifions cette loi au point correspondant de l'arbre à hauteur* $h(T)$. *Ainsi on peut voir la progression de l'algorithme de recuit-simulé avec une température décroissante avec le temps comme une particule qui évolue lentement vers le bas de l'arbre, faisant des choix irrévocables et aléatoires aux points de branchement et s'arrêtant à une feuille.*

3.6. La fonction génératrice de Z_n

Nous étudions dans cette section l'itérée n-ième de la fonction génératrice f de la loi binomiale de paramètres (R, P_0) et donnons des encadrements précis de celle-ci. Pour alléger les notations, nous noterons l'itérée n-ième d'une fonction g par g_n.

3.6.1. Bornes de $f_n(s)$ pour $0 \leq s < 1$ et $m = 1$. —

Remarque 3.6.1. — *Notons que* f_n *et ses dérivées successives sont convexes. On en déduit, entre autre, que* f_n *et ses dérivées successives sont croissantes. De plus, pour tout* $0 \leq s \leq 1$, *on a* $s \leq f(s) \leq f(1) = 1$, *et par récurrence,* $s \leq f(s) \leq f_2(s) \leq .. \leq 1$. *Finalement on obtient* $f_n(s) \underset{n\to\infty}{\longrightarrow} 1$ *puisque* $f_n(s) \geq f_n(0)$.

Proposition 3.6.1. — *Soient* $\alpha_1 = \frac{f''(1)}{2}$, $C = \frac{c_0}{6\alpha_1} \max\limits_{s\in[0,1]} f'''(s) = \frac{c_0 f'''(1)}{6\alpha_1}$ *pour un certain* c_0 *et* $\gamma_{n+1} = n\alpha_1[1 - \frac{C}{n}\sum_{i=1}^{n}\frac{1}{i}]$. *Alors, pour s proche de 1 et n grand,*

$$1 - \frac{1-s}{1+\gamma_n(1-s)} \leq f_n(s) \leq 1 - \frac{(1-s)[1-\alpha_1(1-s)]}{1+\alpha_1(1-s)(n-1-\frac{\alpha_1^2(1-s)^2}{2})} \tag{40}$$

Démonstration. — Borne supérieure : Faisons un développement de Taylor de f au voisinage de 1. Alors il existe $\theta_n \in [f_n(s), f_n(1) = 1]$ tel que

$$\begin{aligned} f_{n+1}(s) = f(f_n(s)) &= f(1) + (f_n(s)-1)f'(1) + \frac{(f_n(s)-1)^2}{2} f''(\theta_n) \\ &= f_n(s) + \frac{(f_n(s)-1)^2}{2} f''(\theta_n), \end{aligned}$$

puisque $f'(1) = 1$. Soit $r_n = 1 - f_n(s)$, r_n satisfait

$$r_{n+1} = r_n - r_n^2 \frac{f''(\theta_n)}{2}.$$

Maintenant posons $\alpha_0 = \frac{f''(0)}{2}$. Définissons les suites décroissantes (a_n) et (b_n) qui vérifient

$$a_{n+1} = a_n - a_n^2 \alpha_1, \qquad b_{n+1} = b_n - b_n^2 \alpha_0, \qquad a_0 = b_0 = 1 - s \tag{41}$$

Alors, comme $\alpha_0 \leq \alpha_1$ par la Remarque 3.6.1,

$$a_n \leq r_n \leq b_n. \tag{42}$$

Il s'agit donc de borner a_n supérieurement et b_n inférieurement.

Tout d'abord, puisque $0 \leq b_j \leq 1$, on a

$$\frac{1}{b_n} = \frac{1}{b_{n-1}} + \alpha_0 \frac{1}{1 - \alpha_0 b_{n-1}} = \frac{1}{b_0} + \alpha_0 \sum_{j=0}^{n-1} \frac{1}{1 - \alpha_0 b_j} \geq \frac{1}{b_0} + n\alpha_0$$

Donc

$$b_n \leq \frac{1-s}{1 + \alpha_0 n(1-s)}, \tag{43}$$

d'où la majoration de b_n.

Appliquons maintenant la borne supérieure précédente à a_n (α_0 devenant α_1), on obtient

$$a_n \leq \frac{1-s}{1 + n\alpha_1(1-s)}. \tag{44}$$

En injectant (44) dans $\frac{1}{a_n} = \frac{1}{a_0} + \alpha_1 \sum_{j=0}^{n-1} \frac{1}{1-\alpha_1 a_j}$, on obtient la minoration suivante de a_n :

$$a_n \geq \frac{(1-s)[1 - \alpha_1(1-s)]}{1 + \alpha_1(1-s)(n - 1 - \frac{\alpha_1^2(1-s)^2}{2})}. \tag{45}$$

Finalement (42) et (45) induisent la borne supérieure de f_n dans (40).

Borne inférieure : On montre par induction que

$$h_{\gamma_n}(s) := 1 - \frac{1-s}{1 + \gamma_n(1-s)} \leq f_n(s) \quad \text{avec} \quad \begin{cases} \gamma_{n+1} = c_n + \gamma_n \\ \gamma_1 = 0 \\ c_n = \alpha_1(1 - \frac{C}{n}) \end{cases} \tag{46}$$

Pour $n = 1$, le résultat est donné par la Remarque 3.6.1. Notons tout d'abord que, pour tout $s \neq 1$,

$$h_n(s) = 1 - \frac{1-s}{1+n(1-s)} = 1 - \frac{1}{n}\frac{1}{1+\frac{1}{n(1-s)}} \geq 1 - \frac{1}{n}.$$

Notons ensuite que

$$n\alpha_1 - \gamma_{n+1} - C\alpha_1 \log n = C\alpha_1\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right),$$

et par conséquent,

$$\gamma_{n+1} \underset{n\to\infty}{\sim} n\alpha_1\left[1 - \frac{1}{n}(\log n + \gamma)\right],$$

où γ est la constante d'Euler. On en déduit que, pour n suffisamment grand, $h_{\gamma_n}(s) \geq 1 - c_0/n$ pour un certain c_0.

Enfin,

$$\begin{aligned}
h_{\gamma_{n+1}}(t) &= h_{c_n} \circ h_{\gamma_n}(t) = h_{c_n}(s), \quad \text{pour} \quad s = h_{\gamma_n}(t) \\
&= 1 + (s-1) + c_n(s-1)^2 + \frac{(s-1)^3}{6} h'''_{c_n}(\theta_n^1) \\
&\leq 1 + (s-1) + c_n(s-1)^2 \\
&= f(s) - (s-1)^2\left[\frac{s-1}{6} f'''(\theta_n^2) + \frac{C\alpha_1}{n}\right] \\
&\leq f(s) \qquad \text{par définition de} \quad C \\
&= f(h_{\gamma_n}(t)) = f \circ h_{\gamma_n}(t)
\end{aligned}$$

Mais par hypothèse de récurrence, on a $h_{\gamma_n}(s) \leq f_n(s)$, et donc, puisque f est croissante,

$$h_{\gamma_{n+1}}(t) \leq f_{n+1}(t).$$

□

Remarque 3.6.2. — *Remarquons que*

$$\gamma_{n+1} \underset{n\to\infty}{\sim} n\alpha_1\left[1 - \frac{1}{n}(\log n + \gamma)\right],$$

où γ est la constante d'Euler.

Dans le cas particulier de $f(s) = (P_0 s + 1 - P_0)^R$, on peut obtenir un résultat plus précis pour la borne inférieure :

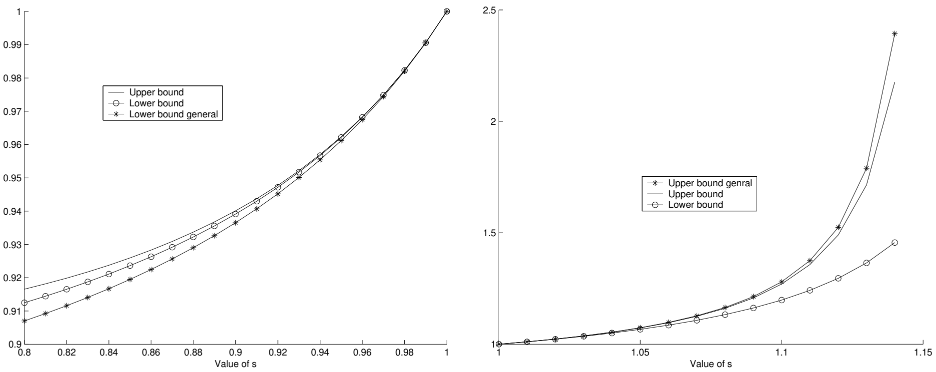
Proposition 3.6.2. — *Pour s proche de 1,*

$$1 - \frac{1-s}{1+n\alpha_1(1-s)} \leq f_n(s) \tag{47}$$

Remarquons l'égalité pour $s = 1$.

Démonstration. — Posons $h(s) = 1-(1-s)/(1+\alpha_1(1-s))$. Puisque $f(1) = h(1) = 1$, $f'(1) = h'(1) = 1$ et $f''(1) = h''(1) = 2\alpha_1$, le signe de $f-h$, au voisinage de 1, dépend de façon évidente du signe de la dérivée troisième de $f-h$, au voisinage de 1, qui est trivialement négative. Alors $h \leq f$. Puisque f est croissante, on déduit (47) par récurrence. $\square$

On a tracé en Figure 6.1 les bornes précédemment obtenues pour $P = 10^{-12}$ et s proche de 1.



1. pour $0 \leq s \leq 1$ 2. pour $1 \leq s$

FIG. 6. Bornes of $f_M(s)$

3.6.2. Bornes de $f_n(s)$ pour $1 \leq s$ et $m = 1$. —

Remarque 3.6.3. — *D'abord, notons que, par convexité, pour tout $s \geq 1$*

$$(s-1)f'(1) \leq f(s) - f(1) = f(s) - 1,$$

d'où $f(s) \geq s$ et par récurrence sur n

$$f_{n+1}(s) \geq f_n(s) \geq .. \geq f(s) \geq s \geq 1.$$

On remarque aussi que pour $s > 1$, la fonction itérée croît rapidement vers l'infini.

Proposition 3.6.3. — *Posons $\gamma'_n = n\alpha_1[1 + \frac{C}{n}(\log n + 1)] - \alpha_1$. Alors, pour s proche de 1 et n assez grand,*

$$1 + \frac{(s-1)}{1 - n\alpha_1 s_n^{P_0-2}(s-1)} \leq f_n(s) \leq 1 + \frac{s-1}{1 - n\gamma'_n(s-1)} \tag{48}$$

Démonstration. — En procédant comme pour la Proposition 3.6.1, on obtient la borne supérieure. Ici $f_n \underset{n\to+\infty}{\longrightarrow} \infty$ et on ne peut donc pas faire un développement de Taylor au voisinage de 1 comme précédemment. Pour contourner cette difficulté, considérons k_n la fonction réciproque de f_n, c'est l'itérée n-ième de la fonction génératrice k, fonction réciproque de f. Notons que k vaut 1 en 1, telle que sa dérivée première vaut 1 en 1, sa dérivée seconde est négative et $k_n \underset{n\to+\infty}{\longrightarrow} 1$. Ainsi par un développement de Taylor au voisinage de 1 et en utilisant les mêmes techniques que précédemment,

$$1 + \frac{(s-1)(1-\alpha_1(s-1))}{1+(n-1)\alpha_1(s-1)} \leq k_n(s) \leq 1 + \frac{s-1}{1 + n\alpha_1 s_n^{P_0-2}(s-1)}$$

où $\beta_2 = \frac{k"(s)}{2}$ et $s_n := 1 + \frac{1}{n\alpha_1}$. En utilisant le lien entre k_n et f_n et la borne supérieure de k_n, on obtient

$$1 + \frac{(s-1)}{1 - n\alpha_1 s_n^{P_0-2}(s-1)} \leq f_n(s). \tag{49}$$

La borne inférieure de k_n conduit à une borne supérieure de f_n ; mais elle n'apporte aucune précision supplémentaire. □

Comme précédemment, on peut obtenir une borne supérieure plus précise dans le cas où $f(s) = (P_0 s + 1 - P_0)^R$:

Proposition 3.6.4. — *Pour s proche de* 1,

$$f_n(s) \leq 1 + \frac{s-1}{1 - n\alpha_1(s-1)} \tag{50}$$

On a tracé en Figure 6.2 ces trois bornes pour $P = 10^{-12}$ et s proche de 1.

3.6.3. Conséquences. — Nous remarquons ainsi que, pour n grand et s proche de 1,

$$f_n(s) \approx 1 - \frac{1-s}{1+n\alpha_1(1-s)}$$

Notons que le membre de droite est simplement l'itérée de l'homographie qui vaut 1 en 1, telle que sa dérivée première vaut m en 1 et sa dérivée seconde vaut $f''(1)$ en 1.

Remarque 3.6.4. — *A propos de la loi géométrique.*
Si la loi de X est telle que les probabilités p_k sont en proportion géométrique : $p_k = \mathbb{P}(X=k) = bc^{k-1}$ pour $k = 1, 2\cdots$ et $p_0 = 1 - \sum_{i\geq 1} p_i$ avec $b, c > 0$ et $b \leq 1-c$, alors la fonction génératrice associée est une fraction rationnelle :

$$h(s) = 1 - \frac{b}{1-c} + \frac{bs}{1-cs}.$$

Posons $b = (1-c)^2$ et $c = \frac{\alpha_1}{1+\alpha_1}$, on obtient alors

$$h(s) = 1 + \frac{s-1}{1-\alpha_1(s-1)}$$

Ainsi on a comparé l'itérée n-ième de la fonction génératrice d'une loi Binomiale à celle d'une loi géométrique. Cela suggère de comparer les modèles de branchement avec duplication des trajectoires associés respectivement à une loi Binomiale et à une loi géométrique. Le second se déroule selon le protocole suivant : on génère les particules les unes après les autres. Tant que le niveau supérieur n'est pas atteint, on continue à générer des particules les unes après les autres. Dès que le niveau supérieur est atteint, on recommence depuis ce niveau (la loi géométrique est la loi du premier succès).

Ce lien est aussi souligné par Cosnard et Demongeot dans [17] : pour $m = 1$ et $\sigma^2 = f''(1) = 2\alpha_1$, le comportement asymptotique de f_{2^n} est le même que celui de l'itérée 2^n-ième de la fonction génératrice d'une loi géométrique de même variance (c-à-d h).

En généralisant pour $m \neq 1$ et en notant maintenant $\alpha_1 = \frac{f''(1)}{2m}$, on obtient finalement

Corollaire 3.6.1. —

$$f_n(s) \underset{n\to\infty}{\sim} 1 - \frac{m^n(1-s)}{1+\alpha_1(1-s)\frac{m^n-1}{m-1}}$$

3.6.4. Autres bornes. — Nous souhaitons dans cette section comparer les bornes obtenues précédemment à des bornes connues et démontrées par Alan Agresti dans [1] et Dean Fearn dans [24].

Tout d'abord, Agresti obtient les encadrements de la proposition 3.6.5, en procédant en trois étapes :
1) l'auteur encadre la fonction génératrice f par des fonctions appartenant à $A = \{p_0 + p_1 s + p_2 s^2 \ : \ p_1 + 2p_2 = f'(1)\}$
ou $B = \{p_0 + (1-p_0)s^k \ : \ k(1-p_0) = f'(1)\}$,
2) puis il encadre les fonctions de A et B par des homographies (dont on connaît explicitement les itérées),
3) enfin il itère ces encadrements.

Proposition 3.6.5. — *On a pour toute fonction génératrice $f(s) = \sum_{i\geq 0} p_i s^i$ telle que $m \leq 1$ et $f"(1) < \infty$ et $0 \leq s \leq 1$,*

$$(51)\quad 1 - \frac{m^n(1-s)}{1 + \min(\frac{1}{2}, \frac{p_0+m-1}{m})(1-s)\frac{1-m^n}{1-m}} \leq f_n(s) \leq 1 - \frac{m^n(1-s)}{1 + \frac{f"(1)}{m}(1-s)\frac{1-m^n}{1-m}}$$

Démonstration. — Eléments de preuve : On montre d'abord les minorations suivantes

$$\begin{cases} \text{pour} \quad p_0 \leq 1 - \frac{1}{2}m, \quad f(s) \geq p_0 + [2(1-p_0) - m]s + (p_0 + m - 1)s^2 \\ \text{pour} \quad p_0 \geq 1 - \frac{1}{2}m, \quad f(s) \geq 1 - \frac{1}{2}m + \frac{1}{2}ms^2 \end{cases}$$

et la borne supérieure

$$f(s) \leq 1 - \frac{m^2}{m + f"(1)} + \frac{m^2}{m + f"(1)} s^{\frac{f"(1)}{m}+1}$$

Ensuite, on encadre les fonctions de A et B par des homographies de la forme $f(b, c; s) := 1 - \frac{b(1-s)}{1+c(1-s)}$:

(i) Posons $b_1 = p_1 + 2p_2$, $c_1 = \frac{p_2}{p_1+2p_2}$ et $c_1' = \frac{p_2}{p_1+p_2}$. Alors

$$f(b_1, c_1; s) \leq p_0 + p_1 s + p_2 s^2 \leq f(b_1, c_1'; s)$$

(ii) Posons $b_2 = k(1-p_0)$, $c_1 = k - 1$ et $c_1' = \frac{k-1}{2}$. Alors

$$f(b_2, c_2; s) \leq p_0 + (1-p_0)s^k \leq f(b_2, c_2'; s)$$

Enfin, on itère les encadrements obtenus. □

On peut déduire de ces résultats le même type de bornes pour $m > 1$ de la façon suivante. Soit $\widetilde{q} < 1$ le point fixe de f. Posons $\bar{f}(s) := \frac{f(\widetilde{q}s)}{\widetilde{q}}$. Alors

$\bar{f}'(s) = f'(\widetilde{q}) < 1$, et donc $\bar{f}$ est la fonction génératrice d'un processus sous-critique et ainsi elle vérifie les encadrements précédents. On peut interpréter $\bar{f}$ comme la fonction génératrice du processus original conditionné par une extinction éventuelle.

Maintenant $f_n(s) = \widetilde{q}\bar{f}_n(s)$. D'où

Proposition 3.6.6. —

$$\widetilde{q}\left[1 - \frac{f'(\widetilde{q})^n(1-s)}{1+\bar{\alpha}\frac{1-f'(\widetilde{q})^n}{1-f'(\widetilde{q})}}\right] \leq f_n(s) \leq \widetilde{q}\left[1 - \frac{f'(\widetilde{q})^n(1-s)}{1+\widetilde{q}\frac{f''(\widetilde{q})}{f'(\widetilde{q})}(1-s)\frac{1-f'(\widetilde{q})^n}{1-f'(\widetilde{q})}}\right] \tag{52}$$

avec $\bar{\alpha} := \min\left\{\frac{1}{2}, \frac{\frac{f(0)}{\widetilde{q}}+f'(\widetilde{q})-1}{f'(\widetilde{q})}\right\}$.

Remarque 3.6.5. — *Nous pouvons aisément généraliser ces encadrements au cas de processus de branchement en environnement variable c'est à dire lorsque les lois de reproduction g_j sont indépendantes mais non identiquement distribuées. On obtient alors, en notant π_n la fonction génératrice de Z_n :*

$$\pi_n(s) = g_0 \circ \cdots \circ g_{n-1}(s),$$

$$1 - \left[\frac{1}{Q_n(1-s)} + \frac{1}{2}\sum_{j=0}^{n-1}\frac{g_j''(0)}{g_j'(1)Q_{j+1}}\right]^{-1} \leq \pi_n(s) \leq 1 - \left[\frac{1}{Q_n(1-s)} + \sum_{j=0}^{n-1}\frac{g_j''(1)}{g_j'(1)Q_{j+1}}\right]^{-1}$$

où $Q_n = \prod_{j=0}^{n-1} g_j'(1)$.

Citons maintenant les encadrements obtenus dans [24].

Proposition 3.6.7. — *On a pour toute fonction génératrice* $f(s) = \sum_{i\geq 0} p_i s^i$ *et* $0 \leq s \leq 1$,

$$1 - \frac{m^n(1-s)}{1+\frac{p_0+m-1}{m}(1-s)\frac{1-m^n}{1-m}} \leq f_n(s) \leq 1 - \frac{m^n(1-s)}{1+\frac{f''(1)}{2(1-p_0)}(1-s)\frac{1-m^n}{1-m}} \tag{53}$$

Remarquons que les bornes inférieures de Fearn et Agresti sont les mêmes.

Démonstration. — Eléments de preuve : On montre d'abord l'égalité suivante

$$\frac{m}{1-f(s)} = \frac{1}{1-s} + A(s)$$

avec $A(s) = \sum_{k=2}^{\infty} p_k \sum_{j=1}^{k-1} \sum_{l=0}^{j-1} \frac{s^l}{\sum_{k=1}^{\infty} p_k \sum_{j=0}^{k-1} s^j}$.

Ensuite, on encadre la fonction A. Enfin, on itère les encadrements obtenus. □

3.6.4.1. Commentaires. — Ces encadrements sont valables pour toute fonction génératrice f et tout $s \in [0,1]$ et nous permettront donc d'estimer la probabilité d'extinction comme nous le verrons dans la section suivante. Cependant, elles sont moins précises que les bornes obtenues en sections 3.6.1 et 3.6.2.

3.7. Probabilité d'extinction

Nous voulons estimer la probabilité d'extinction pour M donné :

$$\mathbb{P}\left(Z_{M+1} = 0 | Z_0 = N\right)$$

Nous allons utiliser les encadrements de fonctions génératrices quelconques par des homographies dus à Agresti [1], valables pour tout $s \in [0,1]$ et présentés en section 3.6.4.

Dans notre modèle, $f(s) = (P_0 s + 1 - P_0)^R$ et la probabilité d'extinction d'une particule initialement générée est

$$\mathbb{P}\left(Z_{M+1} = 0 | Z_0 = 1\right) = P_0 f_M(0) + 1 - P_0,$$

d'où

$$\begin{aligned} P_0\left(1 - \frac{m^M}{1 + \min(\frac{1}{2}, \frac{p_0+m-1}{m})\frac{1-m^M}{1-m}}\right) + 1 - P_0 &\leq \mathbb{P}\left(Z_{M+1} = 0 | Z_0 = 1\right) \\ \leq P_0\left(1 - \frac{m^M}{1 + \frac{f''(1)}{m}\frac{1-m^M}{1-m}}\right) + 1 - P_0 & \end{aligned}$$

En générant N particules indépendantes au départ, on obtient alors un encadrement de la probabilité d'arrêt de l'algorithme :

$$\left[P_0\left(1-\frac{m^M}{1+\alpha\frac{1-m^M}{1-m}}\right)+1-P_0\right]^N \leq \mathbb{P}\,(\text{ arrêt algorithme })$$
$$\leq \left[P_0\left(1-\frac{m^M}{1+\frac{f''(1)}{m}\frac{1-m^M}{1-m}}\right)+1-P_0\right]^N$$

On a donc $\mathbb{P}\,(\text{ arrêt algorithme }) = \Theta\,(\exp -NC)$.

<u>Application numérique</u> : (1) Pour $P = 10^{-15}$, $C = 10^4$, l'algorithme optimal donne

$$M = 20, P_0 = 0.1931 \quad \text{et} \quad N = 662$$

et en prenant $R = \lfloor 1/P_0 \rfloor = 5$ (alors $m = 0.9653 < 1$), on obtient

$$0.9831 \leq \mathbb{P}\,(Z_{M+1} = 0) \leq 0.9922$$

et

$$1.2796\ 10^{-5} \leq \mathbb{P}\,(\text{ arrêt algorithme }) \leq 5.7\ 10^{-3}$$

(2) Pour $P = 10^{-12}$, $C = 10^4$, l'algorithme optimal donne

$$M = 16, P_0 = 0.1968 \quad \text{et} \quad N = 588$$

et en prenant $R = 1/P_0 = 5.08$ (alors $m = 1$), on obtient

$$0.9701 \leq \mathbb{P}\,(Z_{M+1} = 0) \leq 0.9866$$

et

$$1.7657\ 10^{-8} \leq \mathbb{P}\,(\text{ arrêt algorithme }) \leq 3.5192\ 10^{-4}$$

Ensuite, si $\widetilde{q}$ est le point fixe de f et $\bar{\alpha} := \min\left\{\frac{1}{2}, \frac{\frac{f(0)}{\widetilde{q}}+f'(\widetilde{q})-1}{f'(\widetilde{q})}\right\}$,

$$P_0\widetilde{q}\left(1-\frac{f'(\widetilde{q})^M}{1+\bar{\alpha}\frac{1-f'(\widetilde{q})^M}{1-f'(\widetilde{q})}}\right)+1-P_0 \leq \mathbb{P}\,(Z_{M+1}=0|Z_0=1)$$
$$\leq P_0\widetilde{q}\left(1-\frac{f'(\widetilde{q})^M}{1+\widetilde{q}\frac{f''(\widetilde{q})}{f'(\widetilde{q})}\frac{1-f'(\widetilde{q})^M}{1-f'(\widetilde{q})}}\right)+1-P_0$$

En générant N particules indépendantes au départ, on obtient la probabilité d'arrêt de l'algorithme :

$$\left[P_0\widetilde{q}\left(1-\frac{f'(\widetilde{q})^M}{1+\bar{\alpha}\frac{1-f'(\widetilde{q})^M}{1-f'(\widetilde{q})}}\right)+1-P_0\right]^N \leq \mathbb{P}(\text{ arrêt algorithme })$$
$$\leq \left[P_0\widetilde{q}\left(1-\frac{f'(\widetilde{q})^M}{1+\widetilde{q}\frac{f''(\widetilde{q})}{f'(\widetilde{q})}\frac{1-f'(\widetilde{q})^M}{1-f'(\widetilde{q})}}\right)+1-P_0\right]^N$$

Ainsi, $\mathbb{P}(\text{ arrêt algorithme }) = \Theta(\exp -NC)$.

<u>Application numérique</u> : Pour $P = 10^{-9}$, $C = 10^4$, l'algorithme optimal donne

$$M = 12, P_0 = 0.2031 \quad \text{et} \quad N = 700$$

et en prenant $R = \lfloor 1/P_0 \rfloor = 5$ (alors $m = 1.0155 < 1$), on obtient, après estimation de $\widetilde{q}$,

$$0.9568 \leq \mathbb{P}(Z_{M+1} = 0) \leq 0.9756$$

et

$$3.7574\ 10^{-14} \leq \mathbb{P}(\text{ arrêt algorithme }) \leq 3.0194\ 10^{-8}$$

3.8. Résultats asymptotiques

Enonçons maintenant la démarche asymptotique présentée par Glasserman, Heidelberger, Shahabuddin et Zajic dans [31] pour le modèle de branchement avec duplication des trajectoires afin de faire le lien avec celle proposée dans cette thèse. Considérons une famille de probabilités $\mathbb{P} = \{P_k, k = 1, 2, \cdots\}$ et les probabilités $\gamma_k := P_k(A) \to 0$ lorsque $k \to \infty$, pour un certain borélien A. Remarquons que le contexte est le même que celui de la section 2.2.2.2 dans laquelle était présentée une approche grandes déviations. Nous souhaitons étudier γ_k lorsque $k \to \infty$. Il s'ensuit, pour revenir au modèle de branchement avec duplication des trajectoires, que cela revient à considérer que M, le nombre de niveaux, tend lui aussi vers l'infini. Un estimateur naturel de γ_k est l'estimateur de Monte Carlo : soit $X_k^{(1)}, \cdots, X_k^{(N)}$ un échantillon de N variables aléatoires indépendantes de loi commune P_k ; alors l'estimateur de $\gamma_k = P_k(A)$ correspondant est donné par

$$\widehat{\gamma}_k = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_A(X_k^{(i)})$$

Rappelons que par le théorème central limite, comme expliqué en section 2.2.1, le nombre de copies indépendantes nécessaire pour garantir une erreur relative fixée est proportionnel à $1/\gamma_k^2$ et croît infiniment lorsque $k \to \infty$ puisque

$$\mathbb{P}\left(|\gamma_k - \widetilde{\gamma}_k| \le \epsilon \gamma_k\right) = 1 - \alpha \Longrightarrow N = \left[\frac{q_{1-\alpha}\sigma(\widehat{\gamma}_k)}{\epsilon \gamma_k}\right]^2$$

où $q_{1-\alpha}$ est la quantile de la loi normale d'ordre $1-\alpha$.

Les auteurs définissent d'abord deux notions d'efficacité.

Définition 3.8.1. — *On dira qu'une famille d'estimateurs* $\{\widehat{\gamma}_k, k = 1, 2, \cdots\}$ *a une* erreur relative bornée *si*

$$\limsup_{k\to\infty} \frac{\sigma(\widehat{\gamma}_k)}{\gamma_k} < \infty,$$

où $\sigma(\widehat{\gamma}_k)$ *est l'écart type de* $\widehat{\gamma}_k$.

Sous cette condition, le nombre de particules initialement générées nécessaire pour avoir une erreur relative fixée reste borné. Une condition plus faible est

$$\lim_{k\to\infty} \frac{\log \mathbb{E}(\widehat{\gamma}_k^2)}{\log \gamma_k^2} = 1 \tag{54}$$

que l'on appellera *efficacité asymptotique* ou *condition d'optimalité asymptotique*. En d'autres termes, cette condition signifie que le taux exponentiel de décroissance du moment d'ordre 2 de l'estimateur est deux fois plus grand que celui du moment d'ordre 1. Remarquons que par positivité de la variance c'est la meilleure vitesse possible c-à-d la limsup du rapport ne peut excéder 2. Cette définition de l'efficacité asymptotique est similaire à celle énoncée en section 2.2.2.2 dans le contexte des grandes déviations avec $c = 1$. Elle garantit que la probabilité que l'erreur relative entre γ_k et son estimateur dépasse un certain α reste bornée.

Etudions maintenant l'efficacité asymptotique du modèle de branchement avec duplication des trajectoires. Rappelons que l'estimateur est donné par

$$\widehat{\gamma}_k = \frac{1}{NR_1 \cdots R_{k-1}} \sum_{i_0=1}^{N} \sum_{i_1=1}^{R_1} \cdots \sum_{i_{k-1}=1}^{R_{k-1}} \mathbf{1}_{i_0} \mathbf{1}_{i_0 i_1} .. \mathbf{1}_{i_0 i_1 \cdots i_{k-1}}$$

où $\mathbf{1}_{i_0 i_1 \cdots i_j}$ représente le résultat du j-ième tirage et N le nombre de copies indépendantes générées initialement.

Lorsque les $R_i = R$ pour $i = 1 \cdots M$, $P_i = P_0$ pour $i = 1 \cdots M+1$, $m = RP_0$ et $\gamma_k = P_0^{k+1}$, on obtient alors, en utilisant le lemme 3.3.1,

$$\operatorname{var}(\widehat{\gamma}_k) = \begin{cases} \frac{1}{N} \frac{P_0^{2(k+1)}}{m^k} \frac{m^{k+1}-1}{m-1} \left(\frac{1}{P_0} - 1\right) & m \neq 1 \\ \frac{1}{N} P_0^{2(k+1)} (k+1) \left(\frac{1}{P_0} - 1\right) & m = 1 \end{cases}$$

On en déduit que

$$\operatorname{var}(\widehat{\gamma}_k) = \begin{cases} \frac{1}{N} O\left(P_0^{2(k+1)}\right) & m > 1 \\ \frac{1}{N} P_0^{2(k+1)} (k+1) \left(\frac{1}{P_0} - 1\right) & m = 1 \\ \frac{1}{N} O\left(\frac{1}{m^k} P_0^{2(k+1)}\right) & m < 1 \end{cases}$$

Ainsi $\widehat{\gamma}_k$ est asymptotiquement optimal au sens de la définition (54) lorsque $m > 1$. Mais dans ce cas, la réduction de la variance se fait au détriment du nombre d'échantillons générés qui croît géométriquement. Les deux notions précédentes d'efficacité définissent l'efficacité seulement en fonction de la variance de l'estimateur et non en fonction d'un quelconque coût. La notion suivante d'efficacité asymptotique ou d'optimalité asymptotique semble plus appropriée au contexte, une fois de plus :

Définition 3.8.2. — *On dira qu'une famille d'estimateurs* $\{\widehat{\gamma}_k, k = 1, 2, \cdots\}$ *est* asymptotiquement efficace *ou* asymptotiquement optimale *si*

$$\lim_{k \to \infty} \frac{\log(var(\widehat{\gamma}_k) w(k))}{\log \gamma_k^2} = 1$$

où $w(k)$ désigne le coût moyen engendré lors de la simulation.

Il faut donc définir une notion de coût. Dans l'article, les auteurs proposent deux façons de définir ce dernier : la première en assignant un coût constant égal à une unité pour chaque génération d'une particule, la deuxième tient compte de l'effort nécessaire à une particule pour atteindre le niveau supérieur. Dans le premier cas, le coût moyen de la simulation est donné par

$$w(k) = \begin{cases} N \frac{m^k - 1}{m-1} & m \neq 1 \\ Nk & m = 1 \end{cases} = \begin{cases} NO\left(m^{k+1}\right) & m > 1 \\ N(k+1) & m = 1 \\ NO(1) & m < 1 \end{cases}$$

On remarque alors que l'estimateur $\widehat{\gamma}_k$ est le plus performant lorsque $m = 1$ puisque

$$\frac{\log(\operatorname{var}(\widehat{\gamma}_k) w(k))}{\log \gamma_k} \to 2.$$

Lorsque $m \neq 1$, la limite est inférieure à 2 en raison d'une variance trop grande ($m < 1$) ou d'un effort trop grand ($m > 1$). On retrouve donc l'équilibre déjà obtenu en section 3.4.4.

3.9. Lien avec le modèle d'échantillonnage préférentiel

Il est possible de dégager des parallèles entre le modèle de branchement avec duplication des trajectoires et d'autres modèles : comme nous l'avons déjà vu avec le modèle de Galton-Watson mais aussi avec le modèle d'échantillonnage préférentiel. Nous présentons, dans cette section, deux similarités entre ce dernier et le modèle de branchement avec duplication des trajectoires.

Tout d'abord, nous pouvons voir, dans un certain sens, le modèle étudié dans cette thèse comme un modèle d'échantillonnage préférentiel adaptatif. En effet, notons, comme précédemment, B_i les niveaux de transition. On estime la probabilité de l'événement rare A en M étapes comme suit :

1. Posons $k = 1$ et $n_k = N$.
2. Posons $A = B_k$ et $\mathbb{P}_k = \mathbb{P}(\bullet|B_{k-1})$.
3. Estimons $\mathbb{P}(A)$ en utilisant n_k échantillons avec

   $$\widetilde{P} = \frac{1}{n_k} \sum_{j=1}^{n_k} \mathbf{1}_A(j) \widetilde{L}(j, k)$$

   où $\widetilde{L}(j, k)$ est un estimateur de $L(j, k)$.
4. Posons $k = k + 1$ et n_k égal à R_k fois le nombre de succès précédent. Retourner à l'étape 2 tant que $k \leq M + 1$.

Le changement de mesure qui apparaît dans ce modèle est courant en échantillonnage préférentiel, comme expliqué en section 2.2.2.1. Après chaque étape k, la probabilité d'intérêt devient

$$\mathbb{P}_k = \mathbb{P}(\bullet|B_{k-1})$$

et par définition $\mathbb{P}_k(B_k) = P_k$ est plus grande que la probabilité de l'événement rare lui-même et donc plus facile à calculer. La fonction de vraisemblance est donnée par

$$L(j, k) = \frac{\mathbb{P}(B_k)}{\mathbb{P}_k(B_k)} = \mathbb{P}(B_{k-1}) = \prod_{i=1}^{k-1} P_k$$

mais comme L est inconnue, on l'estime par

$$\widetilde{L}(j,k) = \prod_{i=1}^{k-1} \widetilde{Q}_i$$

où les $\widetilde{Q}_i$ ont été obtenus durant les étapes précédentes.

Remarque 3.9.1. — *1) On ne connaît pas le rapport de vraisemblance actuel des nouvelles trajectoires et les rapports de vraisemblance sont égaux pour tout échantillon d'un même niveau, ce qui n'est pas le cas en général en échantillonnage préférentiel.*
2) Il n'y a pas ici de loi tiltée comme en échantillonnage préférentiel mais les mesures de probabilités sont seulement changées par conditionnement.

D'autre part, nous pouvons aussi voir la simulation d'échantillonnage préférentiel comme une procédure d'estimation faisant intervenir $n = N\prod_{k=1}^{M} R_k$ échantillons possibles à la dernière étape de l'échantillonnage préférentiel. On associe alors, comme vu précédemment, un poids à chaque échantillon. Pour chaque échantillon au niveau 1, le poids est 1. Quand un échantillon atteint le niveau k, il est divisé en R_k sous-trajectoires qui reçoivent donc chacune une fraction $1/R_k$ du poids initial. Ainsi le poids d'un échantillon à la dernière étape est

$$L = \prod_{i=1}^{M} \frac{1}{R_k}$$

et on estime $\mathbb{P}(A)$ par

$$\frac{1}{N}\sum_{i=1}^{n} \mathbf{1}_A(\omega_i)L(\omega_i)$$

Notons que cette formulation de l'estimateur est la même que celle de l'estimateur classique de l'échantillonnage préférentiel ([45]). Mais une fois de plus, cette comparaison est artificielle puisqu'on divise une somme de n estimateurs par seulement N. De plus, L est en fait une constante indépendante des ω_i. Bien qu'il n'y ait pas de changement de loi dans cette formulation, le fait que la fonction de poids L est en fait un rapport de vraisemblance est la principale analogie entre les deux modèles.

Remarque 3.9.2. — *Bayes dans [8] avait introduit cette notion de poids comme facteurs d'importance.*

3.10. Conclusion

La minimisation de la variance de l'estimateur à coût fixé a montré que les niveaux de transition devaient être uniformément répartis c-à-d que les probabilités conditionnelles devaient être égales et que les nombres de retirages devaient être égaux à l'inverse de cette valeur commune. Nous souhaitons maintenant obtenir des intervalles de confiance afin d'étudier la précision de l'algorithme. Nous verrons que ces intervalles de confiance dépendent de la transformée de Laplace de l'estimateur. Une autre démarche d'optimisation, sans doute plus naturelle mais plus complexe, serait donc de minimiser directement la transformée de Laplace afin de déduire un algorithme optimal. Mais nous montrerons a posteriori qu'asymptotiquement les deux démarches conduisent aux mêmes paramètres optimaux.

En pratique, les niveaux de transition sont généralement inconnus mais peuvent être estimés lors d'une première phase d'apprentissage (voir Chapitre 5). Quoi qu'il en soit, dans la plupart des cas, il ne pourra être garanti qu'elles seront toutes égales. C'est pourquoi nous étudions dans le chapitre suivant les intervalles de confiance dans le cadre général non critique ($m \neq 1$) et nous pourrons ainsi déterminer la sensibilité de ces bornes.

En outre, nous avons montré que les nombres de retirage devaient égaler une certaine valeur. Pourtant, il peut s'avérer que cette valeur ne soit pas entière alors que nous voulons dupliquer les trajectoires en un nombre entier. Par conséquent, nous proposons et analysons dans le chapitre suivant deux algorithmes visant à contourner ce problème.

CHAPITRE 4

TRANSFORMÉE DE LAPLACE ET INTERVALLES DE CONFIANCE

Les résultats de ce chapitre sont l'objet de l'article [44] soumis à la revue Stochastic Processes and their Applications.

4.1. Introduction

Le contrôle de la variance de $\widehat{P}$ donne un premier intervalle de confiance pour $\mathbb{P}(A)$: par application de l'inégalité de Markov,

$$\begin{aligned}
\mathbb{P}\left(\frac{|\widehat{P}-\mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) &\leq \frac{1}{\mathbb{P}(A)^2\alpha^2}\mathbb{E}\left(\left(\widehat{P}-\mathbb{E}\left(\widehat{P}\right)\right)^2\right) \\
&= \frac{\operatorname{var}(\widehat{P})}{\mathbb{P}(A)^2\alpha^2} \\
&\leq \frac{1}{\alpha^2 C}\left[(M+1)^2\left(\frac{1}{P_0}-1\right)h(P_0)\right] \\
&\approx \frac{4(M+1)}{\alpha^2 N}h(P_0)
\end{aligned}$$

Cette estimation est en général sans intérêt : elle peut s'avérer non pertinente puisque apparaît au dénominateur la probabilité de l'événement rare qui est très petite. Ainsi il peut arriver que cette borne soit plus grande que 1. Par exemple, pour $h=1$, $M=12$ et $\alpha=10^{-2}$, la borne obtenue vaut $\approx \frac{5.10^5}{N}$; afin d'avoir une borne inférieure à 1, il faut $N \geq 5.10^5$.

En d'autres termes, en notant que $\mathrm{var}(\widehat{P}) = \frac{1}{N^2}\sum_{i=1}^N \mathrm{var}(\widetilde{Q}_i) = \frac{1}{N}\mathrm{var}(\widetilde{Q}) := \frac{s^2}{N}$,

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^N \widetilde{Q}_i - \mathbb{E}(\widetilde{Q})\right| \geq \mathbb{P}(A)\alpha\right) \leq \frac{s^2}{N\mathbb{P}(A)^2\alpha^2}$$

Cette inégalité, à la base de la loi faible des grands nombres, n'est pas satisfaisante puisque sous des conditions supplémentaires de régularité, le théorème de la limite centrale implique

$$\mathbb{P}\left(\sqrt{\frac{N}{s^2}}\left(\frac{1}{N}\sum_{i=1}^N \widetilde{Q}_i - \mathbb{E}(\widetilde{Q})\right) \geq y\right) \underset{N\to\infty}{\longrightarrow} 1-\phi(y)$$

Ce type de résultat n'est pas satisfaisant car asymptotique alors que nous avons besoin de bornes exactes en fonction des paramètres.

Mais par le Lemme2 du ChapitreVII de [25], il vient

$$1-\phi(y) \approx \frac{1}{\sqrt{2\pi}}\frac{e^{-y^2/2}}{y}.$$

On a même, plus précisément,

$$\frac{1}{\sqrt{2\pi}}\left[\frac{1}{y}-\frac{1}{y^3}\right]e^{-y^2/2} \leq 1-\phi(y) \leq \frac{1}{\sqrt{2\pi}}\frac{e^{-y^2/2}}{y}.$$

De quoi on peut espérer une approximation de la forme

$$\mathbb{P}\left(\sqrt{\frac{N}{s^2}}\left(\frac{1}{N}\sum_{i=1}^N \widetilde{Q}_i - \mathbb{E}(\widetilde{Q})\right) \geq y\right) \underset{N\to\infty}{\sim} e^{-Ny^2/(2s^2)} \tag{55}$$

L'inégalité de Markov n'étant pas suffisamment précise, on va donc chercher à affiner le résultat. Une amélioration de cette borne est obtenue grâce à la méthode de Chernoff, fréquemment utilisée en théorie des grandes déviations : pour tout $u > 0$,

$$\begin{aligned}
\mathbb{P}\left(\widehat{P} \geq \mathbb{P}(A)(1+\alpha)\right) &= \mathbb{P}\left(\frac{1}{N}\sum_{i=1}^N \widetilde{Q}_i \geq \mathbb{P}(A)(1+\alpha)\right) \\
&= \mathbb{P}\left(\sum_{i=1}^N W_{M+1}^i \geq N(1+\alpha)\right) \\
&= \mathbb{P}\left(e^{u\sum_{i=1}^N W_{M+1}^i} \geq e^{uN(1+\alpha)}\right) \leq e^{-uN(1+\alpha)}\mathbb{E}\left(e^{uW_{M+1}}\right)^N \\
&\leq \exp\left\{N\left[\log\left(\mathbb{E}\left(e^{uW_{M+1}}\right)\right) - u(1+\alpha)\right]\right\}
\end{aligned}$$

L'optimisation en $u > 0$ implique que

$$\mathbb{P}\left(\widehat{P} \geq (1+\alpha)\mathbb{P}(A)\right) \leq \exp\left\{N \inf_{u>0}\left[\log \mathbb{E}\left(e^{uW_{M+1}}\right) - u(1+\alpha)\right]\right\} \tag{56}$$

De même,

$$\mathbb{P}\left(\widehat{P} \leq (1-\alpha)\mathbb{P}(A)\right) \leq \exp\left\{N \inf_{u<0}\left[\log \mathbb{E}\left(e^{uW_{M+1}}\right) - u(1-\alpha)\right]\right\} \tag{57}$$

Remarque 4.1.1. — *Notons que les paramètres optimaux de l'algorithme déduits par minimisation de la variance coïncident avec ceux déduits par minimisation de la transformée de Laplace. En effet, avec la Remarque 3.4.2 et (31), on obtient la borne gaussienne suivante*

$$2\exp\left(-\frac{\alpha^2}{8}\frac{N}{var(W_{M+1})}\right) \tag{58}$$

Dans notre modèle, la fonction génératrice de Z_{M+1} est $P_0 f_M(s) + 1 - P_0$ où $f(s) = (P_0 s + 1 - P_0)^R$ (fonction génératrice d'une loi binomiale de paramètres (R, P_0)).

La transformée de Laplace de W_{M+1} est alors donnée par

$$\mathbb{E}\left(e^{uW_{M+1}}\right) = P_0 f_M\left(e^{\frac{u}{P_0 m^M}}\right) + 1 - P_0. \tag{59}$$

Posons maintenant

$$\psi(u) := \log\left(f(e^u)\right) = R\log\left(P_0 e^u + 1 - P_0\right),$$

on observe que $\psi = h^{-1} o f o h$ avec $h(u) = \exp(u)$; et ainsi $\psi_{M+1} = h^{-1} o f_{M+1} o h$ et

$$\mathbb{E}\left(e^{uW_{M+1}}\right) = \exp\left\{\frac{1}{R}\psi_{M+1}\left(\frac{u}{P_0 m^M}\right)\right\} \tag{60}$$

Par conséquent, (56) et (57) se traduisent par

$$\mathbb{P}\left(\frac{|\widehat{P} - \mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq \exp\left\{N \inf_{u>0} F^+_{M+1}(u)\right\} + \exp\left\{N \inf_{u<0} F^-_{M+1}(u)\right\}$$

où

$$\begin{cases} F^+_{M+1}(u) := \frac{1}{R}\psi_{M+1}\left(\frac{u}{P_0 m^M}\right) - (1+\alpha)\,u \\ F^-_{M+1}(u) := \frac{1}{R}\psi_{M+1}\left(\frac{u}{P_0 m^M}\right) - (1-\alpha)\,u \end{cases}$$

Nous cherchons donc à obtenir des estimées de la transformée de Laplace de W_{M+1} que l'on réécrit sous la forme

$$\mathbb{E}\left(e^{uW_{M+1}}\right) = \exp\left\{\frac{1}{R}\psi_{M+1}\left(\frac{u}{P_0 m^M}\right)\right\}$$

où ψ_{M+1} est l'itérée $M+1$-ième d'une certaine fonction ψ.

L'expression de ψ, quoique simple, ne permet pas de donner une forme explicite de son itérée $M+1$-ième. En pratique, le nombre optimal de niveaux M n'est pas très grand et par conséquent, des estimées asymptotiques seraient imprécises. Pour la même raison, une approche numérique pourrait être envisagée ; cependant, elle s'avère impraticable puisque le nombre optimal M est inconnu à l'avance et est justement dérivé à partir de ces bornes.

C'est pourquoi, dans cette section, nous cherchons des estimées précises de la transformée de Laplace de W_{M+1} et plus précisément, des estimées précises de ψ_{M+1} et d'abord de ψ. Ce qui est fait selon la technique suivante : au lieu d'utiliser une seule fonction pour borner ψ, nous en utilisons plusieurs afin d'obtenir des bornes supérieures et inférieures plus précises. Nous choisissons ces bornes dans les groupes de Lie de faible dimension des fonctions homographiques et des fonctions affines. Plus la dimension du groupe de Lie est grande et plus précise sera notre approximation, puisque la dimension décrit le nombre de paramètres à ajuster (par exemple, les k premières dérivées en un point). Malheureusement, nous n'avons pas trouvé de groupe de Lie de plus grande dimension ayant la propriété de monotonicité requise pour itérer directement les inégalités obtenues pour ψ. L'intérêt d'utiliser de telles fonctions réside aussi dans le fait que leurs itérées ont une forme explicite. En utilisant cette technique, nous dérivons, dans cette section, des bornes précises de la probabilité de l'erreur relative $\mathbb{P}(A)$ et son estimateur. En pratique, l'ajustement des P_i proches de la valeur optimale peut être fait dans une première phase. La proportion du coût consacré à cette partie d'apprentissage est le sujet du chapitre 5. Mais il apparaît rapidement que, même dans le cas où les P_i sont proches de la valeur optimale, le fait que le nombre de retirage n'est pas entier dégrade rapidement la précision de l'algorithme : dans ce cas, on prend R égal à l'entier le plus proche de la valeur optimale mais quelle que soit sa valeur, la criticalité du processus de Galton-Watson est perdue et la perte de précision est significative comme le montre la Figure 1 illustrant la sensibilité des bornes en m ou encore le tableau suivant donnant la longueur de l'intervalle de confiance à 95% pour la variance et la transformée de Laplace pour

– le modèle optimal (*Optimal*) : $m = 1$,

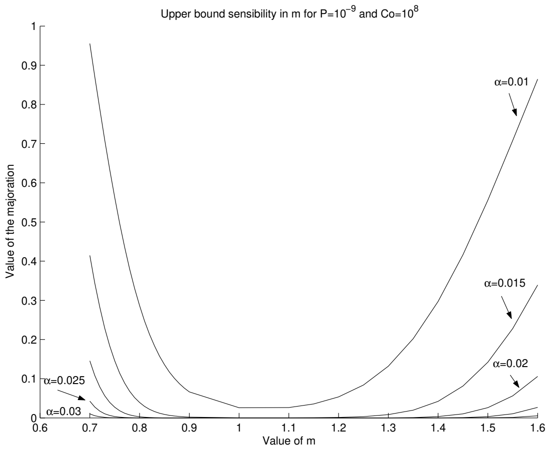


FIG. 1. Sensibilité des bornes en m pour différentes valeurs de α

– le modèle où R est l'entier le plus proche de la valeur optimale (*Déterministe*) : $m \neq 1$,

dans deux cas :

– pour $\mathbb{P}(A) = 3.5\ 10^{-11}$ et $C = 10^8$ (*Cas surcritique*),

– pour $\mathbb{P}(A) = 5\ 10^{-11}$ et $C = 10^8$ (*Cas sous-critique*).

	$\mathbb{P}(A) = 3.5\ 10^{-11}$ et $C = 10^8$		$\mathbb{P}(A) = 5.10^{-11}$ et $C = 10^8$	
	Cas surcritique : $m = 1.111$		Cas sous-critique : $m = 0.909$	
	Variance	Laplace	Variance	Laplace
Optimal	0.05080	0.03948	0.04938	0.03835
Déterministe	0.05702	0.04428	0.05428	0.04215

TAB1. *Longueur de l'intervalle de confiance à* 95%

Ce chapitre est aussi consacré à l'étude de différents moyens pour contourner ce problème. Guidés par [3], nous introduisons deux algorithmes considérant le nombre de retirage comme une variable aléatoire dans l'espoir d'améliorer la simulation et d'être le plus proche du modèle optimal. Plus précisément, posons $k = \lfloor 1/P_0 \rfloor$. Au lieu d'utiliser une valeur déterministe unique pour R, nous générons des variables aléatoires de Bernoulli i.i.d. R_i sur $\{k, k+1\}$ de paramètre $p = \mathbb{P}(R_1 = k)$. Dans un premier modèle, nous générons des variables aléatoires de Bernoulli i.i.d. R_i à chaque fois qu'une particule atteint le niveau supérieur i partant du niveau $i-1$ et nous décidons d'ajuster p de façon à avoir $m=1$. Un second modèle consiste à générer un environnement aléatoire décrit par la suite $(R_1, R_2, .., R_M)$ de M variables aléatoires de Bernoulli i.i.d. R_i sur $\{k, k+1\}$ de paramètre p, obtenu par la technique d'optimisation utilisée précédemment mais avec une constante additionnelle (lien entre R et son inverse).

Nous faisons dans les trois parties suivantes l'étude des trois différents modèles. Dans un dernière section, nous illustrons les résultats grâce à une application numérique.

4.2. Premier modèle : déterministe

4.2.1. Le modèle. — Dans cette section, nous considérons le processus de branchement de Galton-Watson associé au modèle de branchement avec duplication des trajectoires dans le cas où R est déterministe avec $m \neq 1$ et nous souhaitons obtenir des estimées précises des intervalles de confiance.

Rappelons que

$$\mathbb{P}\left(\widehat{P} \geq (1+\alpha)\mathbb{P}(A)\right) \leq \exp\left\{N \inf_{u>0}\left[\log \mathbb{E}\left(e^{uW_{M+1}}\right) - u(1+\alpha)\right]\right\} \tag{61}$$

et

$$\mathbb{P}\left(\widehat{P} \leq (1-\alpha)\mathbb{P}(A)\right) \leq \exp\left\{N \inf_{u<0}\left[\log \mathbb{E}\left(e^{uW_{M+1}}\right) - u(1-\alpha)\right]\right\} \tag{62}$$

Mais puisque, avec $\psi(u) = R\log\left(P_0 e^u + 1 - P_0\right),$

$$\mathbb{E}\left(e^{uW_{M+1}}\right) = \exp\left\{\frac{1}{R}\psi_{M+1}\left(\frac{u}{P_0 m^M}\right)\right\}$$

ces deux inégalités conduisent à

$$\mathbb{P}\left(\frac{|\widehat{P} - \mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq \exp\left\{N \inf_{u>0} F^+_{M+1}(u)\right\} + \exp\left\{N \inf_{u<0} F^-_{M+1}(u)\right\}$$

où

$$\begin{cases} F^+_{M+1}(u) := \frac{1}{R}\psi_{M+1}\left(\frac{u}{P_0 m^M}\right) - (1+\alpha)\, u \\ F^-_{M+1}(u) := \frac{1}{R}\psi_{M+1}\left(\frac{u}{P_0 m^M}\right) - (1-\alpha)\, u \end{cases}$$

4.2.2. La transformée de Laplace de W_{n+1}. —

4.2.2.1. Introduction. — Nous cherchons à obtenir des bornes précises sur la transformée de Laplace de W_{n+1} donnée par (60) (où M est remplacé par n). Par conséquent, nous sommes intéressés par le comportement de ψ_{n+1} et puisque ψ_{n+1} n'a pas d'écriture explicite, nous sommes amenés à estimer ψ dans un premier temps et d'en déduire des bornes de ψ_{n+1} par itération.

Heuristique pour l'estimation de ψ : au lieu d'utiliser une seule fonction pour estimer ψ, on propose des encadrements par morceaux afin d'obtenir des estimées plus précises. Près de l'origine, on bornera ψ par des fonctions homographiques et ailleurs par des droites affines.

Par exemple, pour $u \geq 0$ et $m > 1$, près de l'origine une «bonne» fonction homographique (même valeur en 0, mêmes première et seconde dérivées en 0) majore finement ψ. Pour de grandes valeurs de u, par contre, il vaut mieux utiliser la direction asymptotique. Ainsi, on obtient

$$\psi(u) \leq \begin{cases} mu/\left(1 - \frac{1-P_0}{2}u\right) & \text{pour} \quad 0 \leq u \leq u_0 \\ Ru \quad \text{pour} \quad u_0 \leq u \end{cases}$$

pour un u_0 tel que $mu_0/\left(1 - \frac{1-P_0}{2}u_0\right) = Ru_0$.

Notons $h(u; x, y) = \frac{xu}{1-yu}$ et $g(u; x, y) = xu + y$. Plus généralement, on obtient des majorations et minorations précises de la forme

$$\begin{cases} h(u; x, y) & \text{pour} \quad u \quad \text{près de} \quad 0 \\ g(u; x, y) & \text{sinon} \end{cases}$$

Comme nous l'avons dit précédemment, nous choisissons les fonctions pour les encadrements dans les groupes de Lie de petite dimension des fonctions affines et des fonctions homographiques. L'intérêt d'utiliser de telles fonctions réside dans leur monotonicité mais aussi dans le fait que leurs itérées ont une forme explicite. En effet, il vient facilement,

$$h_n(u;x,y) = \frac{x^n u}{1 - yu\frac{x^n-1}{x-1}} \quad \text{et} \quad g_n(u;x,y) = x^n u + y\frac{x^n - 1}{x-1}$$

Représentons maintenant toute homographie h définie par $h(u) = \frac{au+b}{cu+d}$ par la matrice

$$M_h = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

et de même toute fonction affine g définie par $g(u) = au + b$ par la matrice

$$M_g = \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}$$

Et ainsi nous sommes ramenés à étudier l'action de $Gl(2)$ sur $\mathbb{C}$ définie par

$$\alpha.z = \frac{az+b}{cz+d}, \qquad \text{pour} \;\; \alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Le problème est de décrire l'action du groupe sur un z_0 qui sort d'un domaine ; ainsi le point crucial sera de déterminer le nombre d'itérations après lequel il y aura un changement de domaine et donc un passage des homographies aux fonctions affines.

Heuristique pour l'estimation de ψ_n : Puisque les estimées de ψ sont de deux types, nous bornerons ψ_n avec trois fonctions différentes. En effet, partant d'un u proche de 0, on itèrera h ; tandis que partant d'un u loin de 0, on itèrera la fonction affine g. Partant d'une valeur intermédiaire pour u, nous rencontrerons un changement de régime que décrirons par la suite.

Plus précisément et par exemple pour $m > 1$ et $u > 0$, nous verrons dans la section suivante que 0 est un point fixe répulsif. Donc partant de $u \geq 0$, $\psi_n(u)$ diverge lorsque n tend vers ∞ et il existe différents régimes pour les estimées supérieures :

- si $u \geq u_0$, on itère simplement $u \mapsto Ru$ et l'itérée n-ième de ψ est inférieure à l'itérée n-ième de Ru,
- si $u \leq u_0$, on commence par itérer $h\left(.; m, \frac{1-P_0}{2}\right)$ et si après n itérations $h_n\left(u; m, \frac{1-P_0}{2}\right)$ est toujours inférieure à u_0 alors l'itérée n-ième de ψ est inférieure à l'itérée n-ième de h. Dans le cas contraire, après k itérations, $h_k\left(u; m, \frac{1-P_0}{2}\right)$ est plus grand que u_0, alors on continue en itérant $u \mapsto Ru$. La valeur k_0 qui apparaît dans le lemme suivant correspond à ce changement de régime.

Finalement, les estimées de ψ_n reflètent ce changement de régime et ont la forme suivante

$$\begin{cases} h_n(u;x,y) \text{ pour } u \text{ près de } 0 \\ g_n(u;x,y) \text{ pour } u \text{ loin de } 0 \\ au^\alpha \text{ sinon} \end{cases}$$

4.2.2.2. Comportement et estimées de ψ pour $m > 1$. — La Figure 2 représente le comportement de ψ dans le cas $m > 1$. Notons que ψ a deux points fixes : 0 qui est répulsif et un certain $\widetilde{q} = \psi(\widetilde{q}) < 0$ qui est attractif.

En procédant comme expliqué en Section 4.2.2.1, des études de fonctions

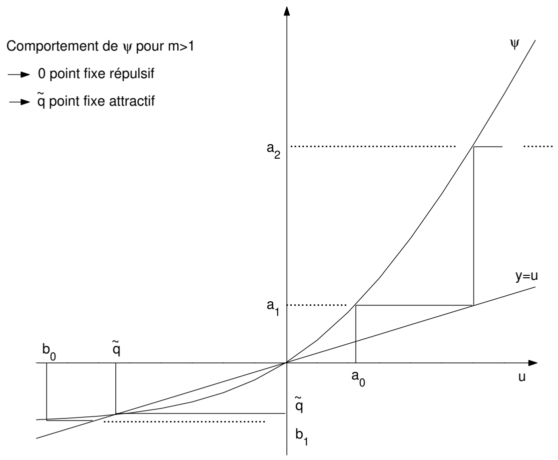


FIG. 2. Comportement de ψ pour $m > 1$

classiques mènent aux lemmes suivants.

Lemme 4.2.1. — *Estimées de ψ pour $u \geq 0$ et $m > 1$*

$$\psi(u) \leq \begin{cases} h\left(u; m, \frac{1-P_0}{2}\right) & \textit{pour } 0 \leq u \leq 2 \\ Ru \quad \textit{pour } u \geq 2 \end{cases}$$

$$\psi(u) \geq \begin{cases} mu \quad \textit{pour } u \leq v_0 \\ g\left(u; R, R\log\left(P_0\right)\right) & \textit{pour } u \geq v_0 \end{cases}$$

où $v_0 := -\log(P_0)/(1-P_0) \geq 0$.

Lemme 4.2.2. — *Estimées de ψ pour $u \leq 0$ et $m > 1$*
Rappelons que $\widetilde{q} = \psi(\widetilde{q})$ et posons
– b *tel que* $h\,(b; m, 1-P_0) < b$,
– $d := [\widetilde{q} - h\,(b; m, 1-P_0)]/(\widetilde{q} - b)$,
– $m_{\widetilde{q}} := \psi'\,(\widetilde{q}) = me^{\widetilde{q}\left(1-\frac{1}{R}\right)}$,
– $u_{-\infty} := R\log(1-P_0)$,
– $v_1 := arg_u\{h\,(u; m, 1-P_0) = g\,(u; m_{\widetilde{q}}, \widetilde{q}(1-m_{\widetilde{q}}))\}$,
– $v_2 := arg_u\{g\,(u; m_{\widetilde{q}}, \widetilde{q}(1-m_{\widetilde{q}})) = u_{-\infty}\} = \widetilde{q} + (u_{-\infty} - \widetilde{q})/m_{\widetilde{q}}$.

$$\psi\,(u) \leq \begin{cases} h\,(b; m, 1-P_0) \quad pour \quad b \leq u \leq 0 \\ g\,(u; d, \widetilde{q}(1-d)) \quad pour \quad \widetilde{q} \leq u \leq b \\ \widetilde{q} + h\left(u - \widetilde{q}; m_{\widetilde{q}}, (1-P_0)e^{-\frac{\widetilde{q}}{R}}\right) \quad pour \quad u_2 \leq u \leq \widetilde{q} \quad pour\ un\ certain \quad u_2 \leq 0 \end{cases}$$

$$\psi\,(u) \geq \begin{cases} h(u; m, \frac{1-P_0}{2}) \quad pour \quad v_1 \leq u \leq 0 \\ g\,(u; m_{\widetilde{q}}, \widetilde{q}(1-m_{\widetilde{q}})) \quad pour \quad v_2 \leq u \leq v_1 \\ u_{-\infty} \quad pour \quad u \leq v_2 \end{cases}$$

4.2.2.3. Estimées de ψ_n pour $m > 1$. —

Théorème 4.2.1. — *Estimées de ψ_n pour $u \geq 0$ et $m > 1$*
Posons $u_{\star,n} := h_n^{-1}\left(2; m, \frac{1-P_0}{2}\right) = 2/[m^n(1+\sigma_n^2/R)]$ et $\gamma := \log R/\log m > 1$.
Alors il existe δ_i for $i = 0, 1$ tels que

$$\psi_n\,(u) \leq \begin{cases} h_n(u; m, \frac{1-P_0}{2}) \quad pour \quad 0 \leq u \leq u_{\star,n} \\ 2R^{n-\delta_0}\left(\frac{u}{1+\frac{1-P_0}{2(m-1)}u}\right)^{\gamma} \quad pour \quad u_{\star,n} \leq u \leq 2 \\ R^n u \quad pour \quad 2 \leq u \end{cases}$$

$$\psi_n\,(u) \geq \begin{cases} m^n u \quad pour \quad 0 \leq u \leq \frac{v_0}{m^n} \\ R^{n-\delta_1}u^{\gamma}\,(v_0 - q_g) \quad pour \quad \frac{v_0}{m^n} \leq u \leq v_0 \\ g_n\,(u; R, R\log P_0) \quad pour \quad v_0 \leq u \end{cases}$$

Démonstration. — Borne supérieure
Par souci de clarté, nous noterons h au lieu de $h(.; m, \frac{1-P_0}{2})$, pendant la preuve.
Si $\mathbf{u} \leq \mathbf{u_{\star,n}}$,

-puisque $m > 1$, $u \leq h(u)$ et puisque h est croissante, $u \leq h(u) \leq \cdots \leq h_n(u)$.
-puisque $u \leq u_{\star,n}$, $h_n(u) \leq 2$ alors

$$\psi_n(u) = \psi_{n-1}(\psi(u)) \leq \psi_{n-1}(h(u)) \leq \cdots \leq h_n(u) = \frac{m^n u}{1 - \frac{1-P_0}{2} u \frac{m^n - 1}{m-1}}$$

If $\mathbf{u} \geq \mathbf{u_0}$, since $R \geq 1$, $2 \leq u \leq Ru \leq \cdots \leq R^n u$ et donc $\psi_n(u) \leq R^n u$

Si $\mathbf{u} \in [\mathbf{u_{\star,n}}, \mathbf{2}]$, puisque $u \leq 2$, on commence par itérer h et puisque $u \leq h(u)$ et $u \geq h_n^{-1}(2)$, les itérées $h_k(u)$ finalement deviennent supérieures à 2 et alors on itère $u \mapsto Ru$. Posons k_0 tel que $h_{k_0}(u) \leq 2$ et $h_{k_0+1}(u) > 2$. k_0 correspond au changement de régime et

$$k_0 \quad := \quad \log\left(\frac{2}{1 + \frac{1-P_0}{2(m-1)} 2}\right) / \log m - \log\left(\frac{u}{1 + \frac{1-P_0}{2(m-1)} u}\right) / \log m$$

Posons $\delta_0 = \log\left(\frac{2}{1 + \frac{1-P_0}{2(m-1)} 2}\right) / \log m$, on a

$$\begin{aligned} \psi_n(u) &= \psi_{n-k_0}(\psi_{k_0}(u)) \leq R^{n-k_0} h_{k_0}(u) \leq R^{n-k_0} 2 \\ &\leq \cdots \leq 2R^{n-\delta_0} \left(\frac{u}{1 + \frac{1-P_0}{2(m-1)} u}\right)^{\gamma} \end{aligned}$$

□

Un comportement similaire étant observé dans les autres cas, nous omettrons les preuves des théorèmes suivants.

Théorème 4.2.2. — *Estimées de ψ_n pour $u \leq 0$ et $m > 1$*
Posons
- $u_{\star\star,n} := h_n^{-1}(b; m, 1 - P_0) = b/[m^n(1 + \sigma_n^2 b/R)]$,
- $v_{\star,n} := h_n^{-1}\left(v_1; m, \frac{1-P_0}{2}\right) = v_1/[m^n(1 + \sigma_n^2 v_1/(2R))]$,
- $v_{\star\star,n} := g_n^{-1}(v_2; m_q, q(1 - m_q)) = q - (q - v_2)/m^n$,
- $\eta := \log d / \log m$ *et* $\mu := \log m / \log m_q$.

Alors il existe δ_i pour $i = 2, 3$ tels que

$$\psi_n(u) \leq \begin{cases} h_n(u; m, 1-P_0) & pour \;\; u_{\star\star,n} \leq u \leq 0 \\ q + d^{n-\delta_2}\left(\frac{u}{1+\frac{1-P_0}{m-1}}\right)^{\eta}(b-q) & pour \;\; b \leq u \leq u_{\star\star,n} \\ g_n(u; d, q(1-d)) & pour \;\; q \leq u \leq b \\ q + h_n\left(u-q; m_q, (1-P_0)e^{-\frac{q}{R}}\right) & pour \;\; u_2 \leq u \leq q \end{cases}$$

$$\psi_n(u) \geq \begin{cases} h_n(u; m, \frac{1-P_0}{2}) & pour \;\; v_{\star,n} \leq u \leq 0 \\ q + m_q^{n-\delta_3}\left(\frac{u}{1+\frac{1-P_0}{2(m-1)}u}\right)^{-\mu}(v_1-q) & pour \;\; v_1 \leq v_{\star,n} \\ g_n(u; m_q, q(1-m_q)) & pour \;\; v_{\star\star,n} \leq u \leq v_1 \\ u_{-\infty} & pour \;\; u \leq v_{\star\star,n} \end{cases}$$

4.2.2.4. Estimées de ψ_n pour $m < 1$. — En procédant de la même façon, on obtient le même genre d'inégalités.

4.2.3. Intervalles de confiance. —

Théorème 4.2.3. — *Bornes de $\mathbb{P}\left(\frac{|\widehat{P}-\mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right)$*

Pour α assez petit et $m \neq 1$,

$$\mathbb{P}\left(\frac{|\widehat{P}-\mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq h_+ + h_-$$

où

$$\begin{cases} h_+ := \exp\left\{-\frac{2C}{h(P_0)(1-P_0)R}m^{M+1}\left(\frac{1-m}{1-m^{M+1}}\right)^2\left(\sqrt{1+\alpha}-1\right)^2\right\} \\ h_- := \exp\left\{-\frac{C}{h(P_0)(1-P_0)R}m^{M+1}\left(\frac{1-m}{1-m^{M+1}}\right)^2\left(1-\sqrt{1-\alpha}\right)^2\right\} \end{cases}$$

Démonstration. — Nous observons dans tous les cas le même comportement : nous ferons donc seulement la preuve pour le cas $m > 1$ et $u \geq 0$ par exemple. Nous avons $F^+_{M+1}(u) := \frac{1}{R}\psi_{M+1}\left(\frac{u}{P_0 m^M}\right) - (1+\alpha)u$. Alors

$$\inf_{u>0} F^+_{M+1}(u) = \inf\left\{\inf_{u\in[0,u_{\star,M+1}]} F^+_{M+1}(u), \inf_{u\in[u_{\star,M+1},2]} F^+_{M+1}(u), \inf_{u>2} F^+_{M+1}(u)\right\}$$

Pour $\mathbf{u} \in [\mathbf{0}, \mathbf{u}_{\star,\mathbf{M+1}}]$,

$$F_{M+1}^{+}(u) \leq \frac{u}{1-\sigma_{M+1}^{2}u/2} - u(1+\alpha) := \phi_{M+1}(u)$$

où σ_{M+1}^2 est la variance de W_{M+1} définie à la Remarque 3.4.2.

L'annulation de la dérivée première de ϕ_{M+1} donne le minimum de ϕ_{M+1} :

$$u_s = \frac{2}{\sigma_{M+1}^2}\left(1-\frac{1}{\sqrt{1+\alpha}}\right)$$

qui est inférieur à $u_{\star,M+1}$ pour α assez petit c-à-d

$$\alpha \leq \left(1-u_{\star,M+1}\sigma_{M+1}^2/2\right)^{-2}-1$$

Pour $\mathbf{u} \in [\mathbf{u}_{\star,\mathbf{M+1}}, \mathbf{2}]$, $F_{M+1}^{+}(u) \leq R^{M-\delta_0}2\left(\frac{u}{1+\frac{1-P_0}{2(m-1)}\frac{u_{\star,M+1}}{P_0m^M}}\right)^{\gamma} - (1+\alpha)u := \phi_{M+1}(u)$. L'annulation de la dérivée première de ϕ_{M+1} donne le minimum de ϕ_{M+1} :

$$\widetilde{u}_s = \left\{\frac{(1+\alpha)}{\gamma 2R^{M-\delta_0}}\left(P_0m^M+\frac{1-P_0}{2(m-1)}u_{\star,M+1}\right)^{\gamma}\right\}^{\frac{1}{\gamma-1}}$$

qui est inférieur à $u_{\star,M+1}$ pour

$$\alpha \leq \frac{\gamma 2R^{M-\delta_0}}{u_{\star,M+1}}\left(\frac{u_{\star,M+1}}{P_0m^M+\frac{1-P_0}{2(m-1)}u_{\star,M+1}}\right)^{\gamma}-1$$

Et donc la solution sort de l'intervalle.

Pour $\mathbf{u} \geq \mathbf{2}$, $F_{M+1}^{+}(u) \leq R^M u/\left(P_0m^M\right) - u(1+\alpha) = u\left[1/P_0^{M+1}-(1+\alpha)\right]$, et $1/P_0^{M+1}-(1+\alpha)$ est positif pour M suffisamment grand ; donc le minimum est atteint en 2 et la solution sort de l'intervalle.

Finalement, $\mathbb{P}\left(\widehat{P} \geq P(1+\alpha)\right)$ est borné par

$$\exp\left\{\frac{N}{R}F_{M+1}^{+}(u_s)\right\} = \exp\left\{-\frac{N}{R}\frac{1}{1-P_0}m^{M+1}\frac{1-m}{1-m^{M+1}}\left(\sqrt{1+\alpha}-1\right)^2\right\}$$

On obtient ainsi le résultat en notant que

$$N = \begin{cases} \frac{C}{h(P_0)}\frac{m-1}{m^{M+1}-1} & \text{pour } m \neq 1 \\ \frac{C}{h(P_0)}\frac{1}{M+1} & \text{pour } m = 1. \end{cases}$$

□

Théorème 4.2.4. — *Pour $m = 1$ et α assez petit, on a*

$$\mathbb{P}\left(\frac{|\widehat{P} - \mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq g_+ + g_-$$

où

$$\begin{cases} g_+ := \exp\left\{-\frac{2C}{h(P_0)(1-P_0)R}\frac{1}{(M+1)^2}\left(\sqrt{1+\alpha}-1\right)^2\right\} \\ g_- := \exp\left\{-\frac{C}{h(P_0)(1-P_0)R}\frac{1}{(M+1)^2}\left(1-\sqrt{1-\alpha}\right)^2\right\} \end{cases}$$

On vérifie que cette borne est meilleure que celle obtenue dans le cas $m \neq 1$.

Le Corollaire suivant découle du Théorème 4.2.3 et fournit une borne supérieure précise et exacte de $\mathbb{P}\left(|\widehat{P} - \mathbb{P}(A)|/\mathbb{P}(A) \geq \alpha\right)$ faisant intervenir la variance de W_{M+1}.

Corollaire 4.2.1. — *Pour α comme dans le Théorème 4.2.3,*

$$\mathbb{P}\left(\frac{|\widehat{P} - \mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq 2\exp\left\{-\frac{\alpha^2}{4}\frac{N}{var(W_{M+1})}\left(1-\frac{\alpha}{2}\right)\right\}$$

Démonstration. — Par le Théorème 4.2.3, la définition de N et la valeur de $\mathrm{var}(W_{M+1})$ donnée dans la Remarque 3.4.2, on déduit que $\mathbb{P}\left(|\widehat{P} - \mathbb{P}(A)|/\mathbb{P}(A) \geq \alpha\right)$ est majoré par

$$\exp\left\{-\frac{2N}{\mathrm{var}(W_{M+1})}\left(\sqrt{1+\alpha}-1\right)^2\right\} + \exp\left\{-\frac{N}{\mathrm{var}(W_{M+1})}\left(1-\sqrt{1-\alpha}\right)^2\right\}$$

Mais puisque $\sqrt{1+\alpha} \leq 1 + \frac{\alpha}{2}$ et $\sqrt{1-\alpha} \leq 1 - \frac{\alpha}{2}$,

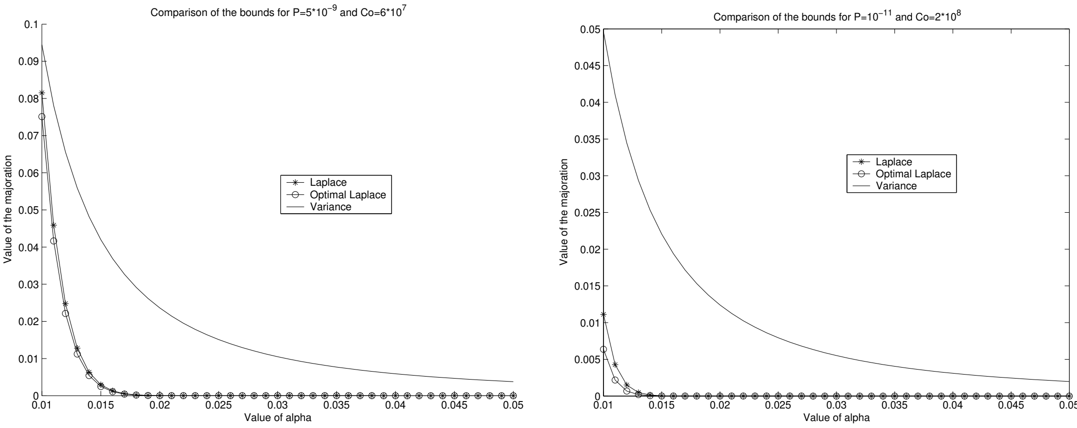
$$\left(\sqrt{1+\alpha}-1\right)^2 = \frac{\alpha^2}{2+\alpha+2\sqrt{1+\alpha}} \geq \frac{\alpha^2}{4}\frac{1}{1+\frac{\alpha}{2}} \geq \frac{\alpha^2}{4}\left(1-\frac{\alpha}{2}\right)$$
$$\left(1-\sqrt{1-\alpha}\right)^2 = \frac{\alpha^2}{2-\alpha+2\sqrt{1-\alpha}} \geq \frac{\alpha^2}{4}\left(1+\frac{\alpha}{2}\right) \geq \frac{\alpha^2}{4}\left(1-\frac{\alpha}{2}\right)$$

qui conduit au résultat cherché. □

Découle aussi des Théorèmes 4.2.3 et 4.2.4 la remarque suivante qui donne l'allure asymptotique des bornes précédentes :

Remarque 4.2.1. — *Pour M grand et $c_1 = 1/[8R(1-P_0)]$, la borne supérieure de $\mathbb{P}\left(|\widehat{P} - \mathbb{P}(A)|/\mathbb{P}(A) \geq \alpha\right)$ est de l'ordre de*

$$2\exp\left\{-c_1 \frac{C\ \alpha^2}{h(P_0)} \frac{1}{m^{M+1}}\right\} \quad pour \quad m > 1$$
$$2\exp\left\{-c_1 \frac{C\ \alpha^2}{h(P_0)} \frac{1}{(M+1)^2}\right\} \quad pour \quad m = 1$$
$$2\exp\left\{-c_1 \frac{C\ \alpha^2}{h(P_0)} m^{M+1}\right\} \quad pour \quad m < 1$$



1. Cas surcritique   2. Cas sous-critique

FIG. 3. Estimées supérieures pour $m \neq 1$

4.2.4. Illustration numérique. — Nous avons tracé en Figure 3 les différentes bornes utilisant :
- la variance,
- la transformée de Laplace dans le cas optimal : $m = 1$,
- la transformée de Laplace dans l'autre cas : $m \neq 1$,

pour les deux jeux de valeurs suivants
- pour $\mathbb{P}(A) = 5\ 10^{-9}$ et $C = 6\ 10^7$, alors $m = 1.055$ (cas surcritique),
- pour $\mathbb{P}(A) = 10^{-11}$ et $C = 2\ 10^8$, alors $m = 0.924$ (cas sous-critique).

4.2.5. Comparaisons avec des bornes connues. — Nous souhaitons, dans cette section, comparer les bornes précédemment trouvées à d'autres déjà connues.

4.2.5.1. L'inégalité de Hoeffding. — Notons $S_n = \sum_{i=1}^n X_i$ où les X_i sont des variables aléatoires indépendantes. On s'intéresse à

$$\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t)$$

Nous utilisons ici la méthode de Chernoff pour borner supérieurement cette probabilité, méthode qui s'avère très efficace dans le cadre de somme de variables aléatoires indépendantes. En effet, on obtient pour tout $s > 0$,

$$\begin{aligned}\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) &\leq e^{-st}\mathbb{E}\left(\exp\left\{s\sum_{i=1}^n [X_i - \mathbb{E}(X_i)]\right\}\right) \\ &= e^{-st}\prod_{i=1}^n \mathbb{E}\left(\exp\left\{s[X_i - \mathbb{E}(X_i)]\right\}\right) \qquad (63)\end{aligned}$$

Il s'agit donc de trouver des bornes supérieures précises de la transformée de Laplace des variables aléatoires $X_i - \mathbb{E}(X_i)$. On peut procéder de différentes façons. Une version élégante dans le cadre des variables aléatoires bornées est proposée par Hoeffding [39] :

Lemme 4.2.3. — *Soit X une variable aléatoire centrée telle que $a \leq X \leq b$. Alors pour tout $s > 0$,*

$$\mathbb{E}\left(e^{sX}\right) \leq e^{s^2(b-a)^2/8}$$

Démonstration. — Tout d'abord par convexité de la fonction exponentielle, pour $a \leq x \leq b$,

$$e^{sx} \leq \frac{x-a}{x-b}e^{sb} + \frac{b-x}{b-a}e^{sa}$$

Comme $\mathbb{E}(X) = 0$ et en posant $p = -a/(b-a)$ et $u = s(b-a)$,

$$\begin{aligned}\mathbb{E}(e^{sX}) &\leq \frac{-a}{x-b}e^{sb} + \frac{b}{b-a}e^{sa} \\ &= \left(1 - p + pe^{s(b-a)}\right)e^{-ps(b-a)} \\ &= e^{\varphi(u)}\end{aligned}$$

avec $\varphi(u) = -pu + \log(1 - p + pe^u)$. On a $\varphi(0) = \varphi'(0) = 0$ et

$$\varphi''(u) = \frac{p(1-p)e^{-u}}{(p + (1-p)e^{-u})^2} \leq \frac{1}{4}$$

Ainsi par un développement de Taylor, pour $\theta \in [0, u]$,

$$\varphi(u) = \varphi(0) + u\varphi'(0) + \frac{u^2}{2}\varphi''(\theta) \leq \frac{u^2}{8} = \frac{s^2(b-a)^2}{8}$$

D'où le résultat. □

On en déduit le théorème suivant :

Théorème 4.2.5. — *Soient $X_1, X_2, \cdots, X_n$ des variables aléatoires bornées indépendantes telles que $\mathbb{P}(X_i \in [a_i, b_i]) = 1$ pour tout i. Alors pour tout $t > 0$,*

$$\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq e^{-2t^2/\sum_{i=1}^n (b_i - a_i)^2}$$

et

$$\mathbb{P}(S_n - \mathbb{E}(S_n) \leq -t) \leq e^{-2t^2/\sum_{i=1}^n (b_i - a_i)^2}$$

Démonstration. — C'est immédiat en injectant le résultat du Lemme 4.2.3 dans (63), puis en optimisant sur s (prendre $s = 4t/\sum_{i=1}^n (b_i - a_i)^2$). □

Remarque 4.2.2. — *Tout repose sur l'inégalité de convexité de la fonction exponentielle, la majoration de $\mathbb{E}(e^{sX})$ par $e^{\varphi(u)}$ et la majoration de $\varphi(u)$.*

Cette inégalité a la forme espérée en (55) à ceci près que la variance moyenne $\sigma^2 = \frac{1}{n}\sum_{i=1}^n \mathrm{var}(X_i)$ est remplacée par la borne supérieure $1/4\sum_{i=1}^n (b_i - a_i)^2$. En d'autres termes, l'inégalité de Hoeffding ignore l'information sur la variance des X_i. Les inégalités de la section suivante fournissent une amélioration en ce sens.

Revenons à notre cas particulier : on a tout d'abord

$$\begin{aligned}
\mathbb{P}\left(\widehat{P} - \mathbb{P}(A) \geq \alpha\mathbb{P}(A)\right) &= \mathbb{P}\left(\frac{\mathbb{P}(A)}{N}\sum_{i=1}^N \frac{Z^i_{M+1}}{P_0 m^M} - \mathbb{P}(A) \geq \mathbb{P}(A)\alpha\right) \\
&= \mathbb{P}\left(\sum_{i=1}^N Z^i_{M+1} - \mathbb{E}(\sum_{i=1}^N Z^i_{M+1}) \geq N P_0 m^M \alpha\right) \\
&\leq \exp\left\{-s[N P_0 m^M \alpha]\right\} \mathbb{E}\left(\exp\left\{s[Z_{M+1} - \mathbb{E}(Z_{M+1})]\right\}\right)^N \qquad (64)
\end{aligned}$$

par indépendance et équidistribution.

En vu d'appliquer le Théorème 4.2.5, posons $X_i = Z^i_{M+1} - \mathbb{E}(Z^i_{M+1}) = Z^i_{M+1} - P_0 m^M$. On a alors

$$-P_0 m^M \leq X_i \leq R^M - P_0 m^M$$

et ainsi par application du théorème

$$\mathbb{P}\left(\widehat{P} - \mathbb{P}(A) \geq \alpha\mathbb{P}(A)\right) \leq \exp\{-2\alpha^2 N\mathbb{P}(A)^2\}$$

et

$$\mathbb{P}\left(\widehat{P} - \mathbb{P}(A) \leq -\alpha\mathbb{P}(A)\right) \leq \exp\{-2\alpha^2 N\mathbb{P}(A)^2\}$$

D'où

Proposition 4.2.1. — *Pour tout $\alpha \geq 0$,*

$$\mathbb{P}\left(\frac{|\widehat{P} - \mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq 2\exp\{-2\alpha^2 N\mathbb{P}(A)^2\}$$

Remarque 4.2.3. — *En analysant la preuve dans notre cas, on remarque que tout repose sur l'inégalité*

$$P_0 f_M(e^s) + 1 - P_0 \leq \mathbb{P}(A)e^{sR^M} + 1 - \mathbb{P}(A)$$

où le terme de droite est la fonction génératrice des moments de Z_{M+1} avec $f(s) = (P_0 s + 1 - P_0)^R$. Résultat que l'on obtient par une inégalité de Jensen grâce à la convexité de la fonction $x \mapsto x^R$.

4.2.5.2. Inégalités de Bennett et de Bernstein. — On suppose maintenant sans perte de généralité que $\mathbb{E}(X_i) = 0$ pour tout $i = 1..n$. On part de nouveau de l'inégalité (63) et donc nous avons besoin à nouveau de borner supérieurement $\mathbb{E}(e^{sX_i})$. Notons $\sigma_i^2 = \mathrm{var}(X_i) = \mathbb{E}(X_i^2)$. On obtient alors l'inégalité due à Bennett [11] :

Théorème 4.2.6. — *Soient $X_1, \cdots, X_n$ des variables aléatoires réelles centrées telles que $|X_i| \leq c$ pour tout i avec la probabilité 1. Alors pour tout $t > 0$,*

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i > t\right) \leq \exp\left\{-\frac{n\sigma^2}{c^2} h\left(\frac{ct}{n\sigma^2}\right)\right\}$$

avec $h(u) = (1+u)\log(1+u) - u$ pour $u \geq 0$.

Démonstration. — Posons tout d'abord

$$F_i = \sum_{r=2}^{+\infty} \frac{s^{r-2}\mathbb{E}(X_i^R)}{r!\sigma_i^2}.$$

Puisque $e^{sx} = 1 + sx + \sum_{r=2}^{+\infty} \frac{s^r x^r}{r!}$, on peut écrire

$$\begin{aligned}
\mathbb{E}(e^{sX_i}) &= 1 + s\mathbb{E}(X_i) + \sum_{r=2}^{+\infty} \frac{s^r \mathbb{E}(X_i^r)}{r!} \\
&= 1 + s^2\sigma_i^2 F_i \quad \text{puisque} \quad \mathbb{E}(X_i) = 0 \\
&\leq \exp\left\{s^2\sigma_i^2 F_i\right\}
\end{aligned}$$

Comme les X_i sont bornés, on obtient pour tout $r \geq 2$,

$$\mathbb{E}(X_i^r) \leq c^{r-2}\sigma_i^2 \qquad \text{et} \qquad F_i \leq \frac{1}{(sc)^2} \sum_{r=2}^{+\infty} \frac{(sc)^r}{r!} = \frac{e^{sc} - 1 - sc}{(sc)^2}$$

Par conséquent,

$$\mathbb{E}(e^{sX_i}) \leq \exp\left\{s^2\sigma_i^2 \frac{e^{sc} - 1 - sc}{(sc)^2}\right\}$$

et en injectant dans (63)

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i > t\right) \leq \exp\left\{n\sigma^2(e^{sc} - 1 - sc)/c^2 - st\right\}$$

En optimisant sur s (prendre $s = \frac{1}{c}\log\left(1 + \frac{tc}{n\sigma^2}\right)$), on obtient le résultat du théorème. □

Le résultat est plus parlant si l'on minore de façon élémentaire $h(u)$ par $u^2/(2+2u/3)$ car alors on obtient l'inégalité de Bernstein [12] :

Théorème 4.2.7. — *Sous les conditions du théorème précédent, pour tout $\varepsilon > 0$,*

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i > \varepsilon\right) \leq \exp\left\{-\frac{n\varepsilon^2}{2\sigma^2 + 2c\varepsilon/3}\right\}$$

On remarque que excepté le terme $2c\varepsilon/3$ qui apparaît au dénominateur, l'inégalité de Bernstein est bonne comparée au résultat (55) donné par le théorème de la limite centrale. En outre, l'inégalité de Bernstein fait apparaître un phénomène intéressant : si $\sigma^2 < \varepsilon$, la borne supérieure se comporte en $e^{-n\varepsilon}$ au lieu de $e^{-n\varepsilon^2}$ garanti par l'inégalité de Hoeffding. L'intérêt de cette borne est qu'elle capture à la fois un comportement asymptotique gaussien et un comportement poissonien.

Revenons à notre cas particulier : on repart de (64). Remarquons d'abord que pour tout i,

$$|X_i| = |Z^i_{M+1} - \mathbb{E}(Z^i_{M+1})| \leq R^M =: c$$

puis en notant $\widetilde{\sigma}^2_{M+1} = \mathrm{var}(Z_{M+1})$, on obtient

(i) Inégalité de Bennett

$$\mathbb{P}\left(\widehat{P} - \mathbb{P}(A) \geq \alpha\mathbb{P}(A)\right) \leq \exp\left\{-\frac{N\sigma^2_{M+1}}{R^{2M}} h\left(\frac{\alpha\mathbb{P}(A)R^{2M}}{\sigma^2_{M+1}}\right)\right\}$$

(ii) Inégalité de Bernstein

$$\mathbb{P}\left(\widehat{P} - \mathbb{P}(A) \geq \alpha\mathbb{P}(A)\right) \leq \exp\left\{-\frac{N\alpha^2\mathbb{P}(A)^2R^{2M}}{2\sigma^2_{M+1} + 2/3\alpha\mathbb{P}(A)R^{2M}}\right\}$$

Or $\widetilde{\sigma}^2_{M+1} = \mathrm{var}(Z_{M+1}) = \left(P_0 m^M\right)^2 \mathrm{var}(W_{M+1})$ et on en déduit, par la Remarque 3.4.2, que

$$\widetilde{\sigma}^2_{M+1} = \begin{cases} \left(P_0 m^M\right)^2 \left(\frac{1}{P_0} - 1\right) \frac{m^{M+1}-1}{m-1} \frac{1}{m^M} & \text{pour} m \neq 1 \\ P_0(1-P_0)(M+1) & \text{pour} m = 1 \end{cases}$$

Finalement, asymptotiquement en M,

$$\begin{cases} \widetilde{\sigma}^2_{M+1} \underset{M\to+\infty}{\sim} P_0(1-P_0)\frac{m^{2M+1}}{m-1} & \text{pour } \; m > 1 \\ \widetilde{\sigma}^2_{M+1} = P_0(1-P_0)(M+1) & \text{pour } \; m = 1 \\ \widetilde{\sigma}^2_{M+1} \underset{M\to+\infty}{\sim} P_0(1-P_0)\frac{m^M}{1-m} & \text{pour } \; m < 1 \end{cases}$$

De quoi on déduit les bornes asymptotiques suivantes

Corollaire 4.2.2. — *Asymptotiquement en M, pour toute valeur de m, la borne supérieure de*

$$\mathbb{P}\left(\widehat{P} - \mathbb{P}(A) \geq \alpha\mathbb{P}(A)\right)$$

de type Bernstein est de l'ordre de

$$\exp\left\{-\frac{3}{2}N\alpha\mathbb{P}(A)\right\}$$

Pour plus de détails et d'autres inégalités de concentration voir [49].

4.2.5.3. Commentaires. — Les bornes supérieures obtenues dans ce paragraphe ont certes l'avantage d'être valables pour tout α ; cependant elles se dégradent très vite pour de petites valeurs de $\mathbb{P}(A)$ et en particulier dans notre cadre où nous étudions des événements rares. En outre, pour de petites valeurs de α, elles sont moins bonnes que celles obtenues avec la méthode ad hoc. Enfin, les techniques utilisées pour les inégalités de Bennett et de Bernstein permettant d'obtenir des majorations de $\mathbb{P}\left(\widehat{P}-\mathbb{P}(A)\geq\alpha\mathbb{P}(A)\right)$ semblent être inutisables pour majorer $\mathbb{P}\left(\widehat{P}-\mathbb{P}(A)\leq-\alpha\mathbb{P}(A)\right)$.

Le tableau suivant donne les différentes bornes obtenues dans cette section ainsi que les bornes optimales et déterministes obtenues dans la section précédente, pour $\alpha = 0.01$:
- les bornes optimales (*Laplace optimal*),
- les bornes déterministes c-à-d issues du modèle où R est l'entier le plus proche de la valeur optimale (*Laplace*),
- les bornes de Hoeffding (*Hoeffding*),
- les bornes de Bennett (*Bennett*),
- les bornes de Bernstein (*Bernstein*),

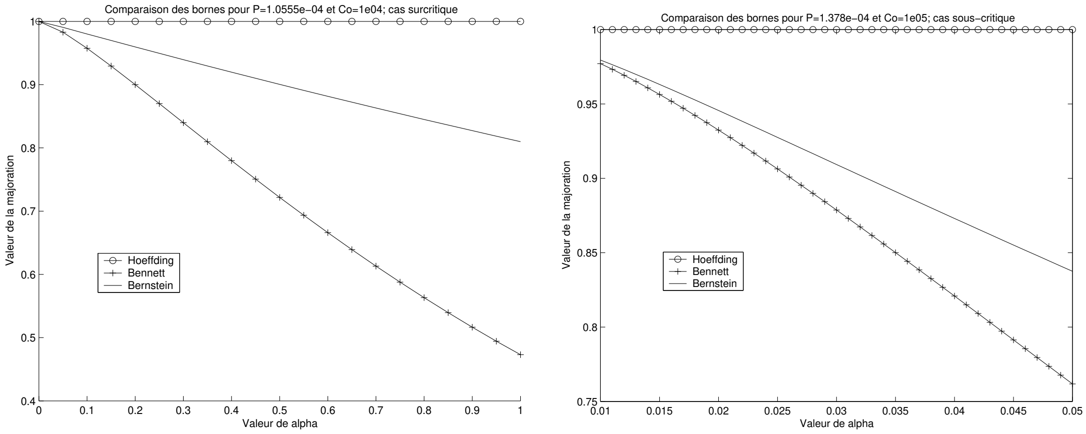
dans deux cas :
- pour $\mathbb{P}(A) = 5\ 10^{-9}$ et $C = 6\ 10^7$ (*Cas surcritique* $m = 1.0556$),
- pour $\mathbb{P}(A) = 10^{-11}$ et $C = 2\ 10^8$ (*Cas sous-critique* $m = 0.9239$).

	Laplace optimal	Laplace	Hoeffding	Bennett	Bernstein
Cas surcritique	0.0052	0.0060	1.0000	0.9980	0.9997
Cas sous-critique	0.0001	0.0003	1.0000	1.0000	1.0000

TAB2. *Comparaison des différentes bornes*

Nous avons tracé en Figure 4 les bornes obtenues dans ce paragraphe :
- par l'inégalité de Hoeffding,
- par l'inégalité de Bennett,
- par l'inégalité de Bernstein,

pour les deux jeux de valeurs suivants
- pour $\mathbb{P}(A) = 10^{-9}$ et $C = 10^8$, alors $m = 1.0155$ (cas surcritique),
- pour $\mathbb{P}(A) = 10^{-10}$ et $C = 10^8$, alors $m = 0.9653$ (cas sous-critique).



1. Cas surcritique 2. Cas sous-critique

FIG. 4. Bornes de Hoeffding, Bennett et Bernstein

4.3. Second modèle : générer à chaque succès

4.3.1. Le modèle. — Soit $k \in \mathbb{N}$ et $\delta \in [0,1[$ tels que $1/P_0 = k + \delta$. Pour s'approcher au plus de l'algorithme optimal, on autorise R à varier c-à-d on considère R comme une variable aléatoire. Plus précisément, chaque fois qu'une particule atteint un niveau supérieur, on génère une réalisation R d'une loi de Bernoulli de paramètre $p = 1 - q$. Notons $\xi = RP_0$.

Ici, l'estimateur est donné par :

$$\widehat{P} = \frac{\mathbb{P}(A)}{N} \sum_{i=1}^{N} W_{M+1}^i = \frac{1}{N} \sum_{i=1}^{N} \frac{Z_{M+1}^i}{\mathbb{E}(Z_{M+1})} = \frac{\mathbb{P}(A)}{N} \sum_{i=1}^{N} \frac{Z_{M+1}^i}{P_0 \mathbb{E}(\xi)^M}$$

4.3.2. Etude de la variance et optimisation de l'algorithme. —

4.3.2.1. Etude de la variance. —

Proposition 4.3.1. — *$\widehat{P}$ est trivialement un estimateur sans biais de variance*

$$var(\widehat{P}) = \frac{\mathbb{P}(A)^2}{N} \left[\left(\frac{1}{P_0} - 1 \right) \sum_{i=0}^{M} \frac{1}{\mathbb{E}(\xi)^i} + \frac{var(\xi)}{P_0 \mathbb{E}(\xi)} \sum_{i=1}^{M} \frac{1}{\mathbb{E}(\xi)^i} \right] \tag{65}$$

Démonstration. — Tout d'abord, $\operatorname{var}(\widehat{P}) = \frac{\mathbb{P}(A)^2}{N} \frac{1}{[P_0 \mathbb{E}(\xi)^M]^2} \operatorname{var}(Z_{M+1})$. Ainsi, le calcul de $\operatorname{var}(\widehat{P})$ se ramène à celui de $\operatorname{var}(Z_{M+1})$.

Notons que pour toute X variable aléatoire et toute filtration $\mathcal{F}$,

$$\operatorname{var}(X) = \operatorname{var}(\mathbb{E}(X|\mathcal{F})) + \mathbb{E}(\operatorname{var}(X|\mathcal{F})) \tag{66}$$

Appliquons (66) à $X = Z_{M+1}$ et $\mathcal{F} = \sigma(Z_M)$ puis à $X = X_M$ et $\mathcal{F} = \sigma(R_M)$ et puisque $X_k^i \sim Bin(R_k, P_0)$, il vient

$$\begin{aligned}
\operatorname{var}(Z_{M+1}) &= \operatorname{var}(\sum_{i=1}^{Z_M} X_M^i) = \operatorname{var}(Z_M \mathbb{E}(X_M|Z_M)) + \mathbb{E}(Z_M \operatorname{var}(X_M|Z_M)) \\
&= \mathbb{E}(X_M)^2 \operatorname{var}(Z_M) + \operatorname{var}(X_M)\mathbb{E}(Z_M) \\
&= \mathbb{E}(\mathbb{E}(X_M|R_M))^2 \operatorname{var}(Z_M) + \operatorname{var}(\mathbb{E}(X_M|R_M)\mathbb{E}(Z_M) + \mathbb{E}(\operatorname{var}(X_M|R_M))\mathbb{E}(Z_M) \\
&= \mathbb{E}(\xi)^2 \operatorname{var}(Z_M) + \operatorname{var}(\xi)\mathbb{E}(Z_M) + (1 - P_0)\mathbb{E}(\xi)\mathbb{E}(Z_M)
\end{aligned}$$

Par une récurrence descendante,

$$\operatorname{var}(Z_{M+1}) = \mathbb{E}(\xi)^{2M} \operatorname{var}(Z_1) + \operatorname{var}(\xi) \sum_{k=1}^{M} \mathbb{E}(Z_k)\mathbb{E}(\xi)^{2(M-k)} + (1-P_0)\mathbb{E}(\xi) \sum_{k=1}^{M} \mathbb{E}(Z_k)\mathbb{E}(\xi)^{2(M-k)}$$

Mais $\operatorname{var}(Z_1) = P_0(1 - P_0)$ et $\mathbb{E}(Z_k) = P_0\mathbb{E}(\xi)^{k-1}$, ce qui conduit finalement au résultat attendu. □

4.3.2.2. Optimisation en les paramètres. — Comme fait en Section 3.4.3, un algorithme optimal se déduit par minimisation de la variance de $\widehat{P}$ pour un budget C donné. Le coût (moyen) est maintenant

$$C = Nh(P_0) \sum_{i=0}^{M} \mathbb{E}(\xi)^i + NP_0 h(1) \sum_{i=0}^{M-1} \mathbb{E}(\xi)^i \tag{67}$$

En négligeant le coût introduit par la génération, à chaque succès, d'un nombre aléatoire de retirage, nous supposerons dans la suite que

$$C = Nh(P_0) \sum_{i=0}^{M} \mathbb{E}(\xi)^i, \qquad (h(1) \ll h(P_0)). \tag{68}$$

En pratique, ce surcoût est effectivement négligeable : par exemple dans le cas de la simulation de trajectoires d'Ornstein-Uhlenbeck.

Supposons de plus $\mathbb{E}(\xi) \neq 1$, alors nous pouvons réécrire la variance sous la forme suivante

$$\begin{aligned}
\mathrm{var}(\widehat{P}) &= \frac{\mathbb{P}(A)^2}{N}\left[\left(\frac{1}{P_0} - 1 + \frac{\mathrm{var}(\xi)}{P_0\mathbb{E}(\xi)}\right)\sum_{i=0}^{M}\frac{1}{\mathbb{E}(\xi)^i} - \frac{\mathrm{var}(\xi)}{P_0\mathbb{E}(\xi)}\right] \\
&= \frac{\mathbb{P}(A)^2}{N}\left[\left(\frac{1}{P_0} - 1 + \frac{\mathrm{var}(\xi)}{P_0\mathbb{E}(\xi)}\right)\frac{1}{\mathbb{E}(\xi)^M}\frac{\mathbb{E}(\xi)^{M+1}-1}{\mathbb{E}(\xi)-1} - \frac{\mathrm{var}(\xi)}{P_0\mathbb{E}(\xi)}\right]
\end{aligned}$$

et le coût s'écrit

$$C = Nh(P_0)\frac{\mathbb{E}(\xi)^{M+1}-1}{\mathbb{E}(\xi)-1}$$

La valeur optimale de N est alors donnée par

$$N = \frac{C}{h(P_0)}\frac{\mathbb{E}(\xi)-1}{\mathbb{E}(\xi)^{M+1}-1}$$

et il s'agit donc de minimiser

$$\mathrm{var}(\widehat{P}) = \frac{\mathbb{P}(A)^2}{CP_0}h(P_0)\frac{\mathbb{E}(\xi)^{M+1}-1}{\mathbb{E}(\xi)-1}\left[\left(1 - P_0 + \frac{\mathrm{var}(\xi)}{\mathbb{E}(\xi)}\right)\frac{1}{\mathbb{E}(\xi)^M}\frac{\mathbb{E}(\xi)^{M+1}-1}{\mathbb{E}(\xi)-1} - \frac{\mathrm{var}(\xi)}{\mathbb{E}(\xi)}\right]$$

dont le terme principal proportionnel à

$$\frac{1}{\mathbb{E}(\xi)^M}\left[\frac{\mathbb{E}(\xi)^{M+1}-1}{\mathbb{E}(\xi)-1}\right]^2$$

est minimal pour $\mathbb{E}(\xi) = 1$.

4.3.2.3. Intervalle de confiance. — A partir de maintenant, nous prendrons

$$m = f'(1) = \mathbb{E}(RP_0) = P_0(k+q) = 1$$

qui équivaut à $p = 1-\delta$. La fonction génératrice de la loi de reproduction dans ce cas devient

$$f(s) = (P_0e^u + 1 - P_0)^k(P_0qe^u + 1 - P_0q) := \exp\{\psi(\log s)\}$$

et son expression reflète bien le fait que R est choisi aléatoirement entre k et $k+1$ selon le paramètre q.

Ensuite, puisque $\log(P_0s + 1 - P_0) \leq \frac{1}{k+q}\log f(s)$ et $m = 1$,

$$\mathbb{E}\left(e^{uW_{M+1}}\right) \leq \exp\left\{\frac{1}{k+q}\log f_{M+1}(e^{\frac{u}{P_0}})\right\} = \exp\left\{\frac{1}{k+q}\psi_{M+1}(\frac{u}{P_0})\right\}$$

En utilisant de nouveau la méthode de Chernoff,

$$\mathbb{P}\left(\frac{|\widehat{P}-\mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \quad = \quad \exp\left\{N \inf_{u>0} F^{+}_{M+1}(u)\right\} + \exp\left\{N \inf_{u<0} F^{-}_{M+1}(u)\right\}$$

où

$$\begin{cases} F^{+}_{M+1}(u) := \frac{1}{k+q}\psi_{M+1}\left(\frac{u}{P_0}\right) - (1+\alpha)\, u \\ F^{-}_{M+1}(u) := \frac{1}{k+q}\psi_{M+1}\left(\frac{u}{P_0}\right) - (1-\alpha)\, u \end{cases}$$

Une fois de plus, nous nous intéressons à la transformée de Laplace de W_{M+1} et nous sommes conduits à étudier le comportement des itérées de la fonction ψ.

4.3.3. La transformée de Laplace de W_{n+1}. — Posons

$$\beta := \left[k(1-P_0) + q(1-P_0 q)\right]/(k+q).$$

Nous procédons comme expliqué en Section 4.2.2.1.

4.3.3.1. Estimées de ψ. —

Lemme 4.3.1. — *Estimées de ψ pour $u \geq 0$*
Il existe $u_0 > 0$ tel que

$$\psi\,(u) \leq \begin{cases} h\left(u;1,\frac{\beta}{2}\right) & \text{pour } \ 0 \leq u \leq u_0 \\ (k+1)u & \text{pour } \ u \geq u_0 \end{cases}$$

Lemme 4.3.2. — *Estimées de ψ pour $u \leq 0$*
Il existe $u^0 < 0$ tel que

$$\psi\,(u) \leq h(u;1,\beta) \quad \text{pour } \ u \leq u^0$$

4.3.3.2. Estimées de ψ_n. —

Proposition 4.3.2. — *Estimées de ψ_n pour $u \geq 0$*
Soit $u_\star := h_n^{-1}(u_0;1,\frac{\beta}{2})$.

$$\psi_n\,(u) \leq \begin{cases} h_n\left(u;1,\frac{\beta}{2}\right) & \text{pour } \ 0 \leq u \leq u_\star \\ u_0(k+1)^{n-\frac{2}{\beta}\left(\frac{1}{u}-\frac{1}{u_0}\right)} & \text{pour } \ u_\star \leq u \leq u_0 \\ (k+1)^n u & \text{pour } \ u_0 \leq u \end{cases}$$

Proposition 4.3.3. — *Estimées de ψ_n pour $u \leq 0$*

$$\psi_n(u) \leq h_n(u; 1, \beta) \quad pour \quad u \leq u^0$$

4.3.4. Intervalle de confiance. — Finalement, par le même type de techniques que dans le cas déterministe, on déduit

Théorème 4.3.1. — *Pour α assez petit, on a*

$$\mathbb{P}\left(\frac{|\widehat{P} - \mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq h_+ + h_- \tag{69}$$

où

$$\begin{cases} h_+ := \exp\left\{-\frac{2C}{(k+q)\beta[h(P_0)+P_0h(1)M/(M+1)]}\frac{1}{(M+1)^2}\left(\sqrt{1+\alpha}-1\right)^2\right\} \\ h_- := \exp\left\{-\frac{C}{(k+q)\beta[h(P_0)+P_0h(1)M/(M+1)]}\frac{1}{(M+1)^2}\left(1-\sqrt{1-\alpha}\right)^2\right\} \end{cases}$$

D'où l'on déduit en utilisant le même genre d'arguments que pour le Corollaire 4.2.1 :

Corollaire 4.3.1. — *Pour α tel que dans le Théorème 4.3.1 et*

$$c_2 = 1/[8(k+q)\beta],$$

$$\mathbb{P}\left(|\widehat{P} - \mathbb{P}(A)|/\mathbb{P}(A) \geq \alpha\right) \leq 2\exp\left\{-c_2 \frac{C\ \alpha^2}{h(P_0)}\frac{1}{(M+1)^2}\right\}$$

Le Corollaire 4.3.1 fournit de meilleurs résultats que ceux obtenus dans la Section 4.2 puisque ici la borne supérieure est en $\exp\{-cst\frac{1}{(M+1)^2}\}$ comme dans le cas optimal.

4.4. Troisième modèle : générer un environnement aléatoire

4.4.1. Le modèle. — Dans ce modèle, on décide de générer un environnement aléatoire décrit par la suite $(R_1, R_2, \cdots, R_M)$ au début de la simulation et d'optimiser l'algorithme en $\mathbb{E}(R)$. Plus précisément, soient $r_i = \{R_j^{(i)}, j = 1..M\}$ pour $i = 1 \cdots N$, N suites de variables aléatoires de loi de Bernoulli sur $\{k, k+1\}$. r_i est la i-ième suite «environnementale». Etant données ces suites environnementales, on considère N processus de branchement indépendants

$(Z_n^{(i)})_{n\geq 0}$, $i = 1, \cdots, N$ de la même façon que dans le cas déterministe à ceci près que chaque R_j est une variable aléatoire associée à une fonction génératrice aléatoire (f.g.) $f_j(s)$ qui représente la loi de reproduction de la j-ième génération et $f_j(s) = (P_0 s + 1 - P_0)^{R_j}$.

(Z_n) est maintenant un processus de branchement en environnement aléatoire (Branching Process in Randon Environment : BPRE).

De même que dans les modèles à environnement aléatoire, il faudra distinguer l'aléa provenant de l'environnement et celui du processus lui-même.

Puisque à chaque processus $(Z_n^{(i)})_{n\leq 0}$ $i = 1..N$ issu des $N > 1$ particules initiales, on associe un environnement aléatoire r_i, les N particules initiales et les processus associés se reproduisent indépendamment. Par conséquent, on peut réécrire $\widehat{P}$ l'estimateur de $\mathbb{P}(A)$ comme la somme de N processus de branchement indépendants :

$$\widehat{P} = \frac{1}{N}\sum_{i=1}^{N} \widetilde{Q}_i = \frac{\mathbb{P}(A)}{N}\sum_{i=1}^{N} \frac{Z_{M+1}^{(i)}}{\xi_0\xi_1\cdots\xi_M} = \frac{\mathbb{P}(A)}{N}\sum_{i=1}^{N} W_{M+1}^{(i)}$$

dont on veut minimiser la variance à coût fixé pour déduire le paramètre optimal pour R. Rappelons qu'ici $\xi_i = R_i P_0$ est une variable aléatoire pour tout $i = 1\cdots M$.

4.4.1.1. Rappels sur les processus de branchement en environnement aléatoire. — Supposons maintenant que $r = \{r_j, j \geq 0\}$ est une suite de variables aléatoires indépendantes et identiquement distribuées qui déterminent la suite des fonctions génératrices $\{f_j(s), j \geq 0\}$ de reproduction d'un processus de branchement en environnement aléatoire. Supposons que $Z_0 = 1$ et que toutes les espérances intervenant dans la suite du paragraphe sont finies.

Soit $q(r)$ la probabilité d'extinction pour un environnement r fixé :

$$q(r) := \mathbb{P}\left(Z_n \underset{n\to\infty}{\longrightarrow} 0 | r\right) = \lim_{n\to\infty} f_0 \circ f_1 \circ \cdots \circ f_{n-1}(0)$$

Alors, d'après Athreya-Karlin [5] par exemple,

Proposition 4.4.1. —

$$q(r) = \lim_{n\to\infty} f_0 \circ f_1 \circ \cdots \circ f_{n-1}(s) \quad p.s. \quad pour \quad 0 \leq s < 1$$

et ou bien $\mathbb{P}(q(r)=1)=1$ *ou bien* $\mathbb{P}(q(r)<1)=1$. *De plus,* $q(r)$ *est l'unique solution* $\widetilde{q}(r)$ *de l'équation fonctionnelle*

$$\widetilde{q}(r)=f_1(\widetilde{q}(r))$$

satisfaisant $\mathbb{P}(q(r)<1)=1$.

De même que pour le modèle de Galton-Watson classique, on a (voir par exemple Smith et Wilkinson [55])

***Théorème 4.4.1* (Extinction ou explosion).** — $Z_n \underset{n\to\infty}{\to} 0$ *avec probabilité* $q(r)$ *et* $Z_n \underset{n\to\infty}{\to} \infty$ *avec probabilité* $1-q(r)$.

En environnement aléatoire, le comportement du processus de branchement Z_n est déterminé par la valeur de $\mathbb{E}(\log f_0'(1))$ au lieu de $m=f'(1)$ dans le modèle classique. Et on a d'après [56] par exemple :

Théorème 4.4.2. — *(i) Si* $\mathbb{E}(\log f_0'(1))\leq 0$, $q(r)=1$ *p.s.*
(ii) Si $\mathbb{E}(\log f_0'(1))>0$, $q(r)<1$ *p.s.*

Définition 4.4.1. — *Un processus de branchement en environnement aléatoire est dit* critique *si* $\mathbb{E}(\log f_0'(1))=0$, surcritique *si* $\mathbb{E}(\log f_0'(1))>0$ *et* sous-critique *si* $\mathbb{E}(\log f_0'(1))<0$.

Cas critique : $\mathbb{E}(\log \mathbf{f_0'(1)})=\mathbf{0}$
En analogie avec le modèle de Galton-Watson classique, on peut obtenir une estimée de la probabilité de non-extinction due à Kolmogorov et la loi de Z_n renormalisé conditionné sur la non-extinction due à Yaglom (voir [32]) :

Théorème 4.4.3. — *Soit* Z_n *un processus de branchement dans un environnement aléatoire avec* $\mathbb{E}(\log f_0'(1))=0$, $\mathbb{P}(f_0'(1)=1)<1$, $\mathbb{E}(\log f_0'(1))^2<+\infty$ *et* $\mathbb{E}\left(\frac{f_0''(1)}{f_0'(1)^2}\right)^{\epsilon}<+\infty$ *pour* $\epsilon>0$. *Alors il existe* $0<C<+\infty$ *tel que*

$$\lim_{n\to+\infty} n^{1/2}\mathbb{P}(Z_n>0)=C$$

De plus, la loi conditionnelle de la variable aléatoire $\frac{Z_n}{\mathbb{E}(Z_n|Z_n>0)}$ *conditionnée par* $\{Z_n>0\}$ *converge vers une loi non-dégénérée* $\nu\neq\delta_0$ *d'espérance 1 et de variance finie.*

Remarque 4.4.1. — *Rappelons que la probabilité de survie* $\mathbb{P}(Z_n > 0)$ *se comporte en* n^{-1} *en environnement déterministe alors qu'elle se comporte en* $n^{-1/2}$ *en environnement aléatoire comme nous venons de le voir.*

Cas sous-critique : $\mathbb{E}(\log \mathbf{f}_0'(\mathbf{1})) < \mathbf{0}$
Ici nous avons, voir [4],

Théorème 4.4.4. — *Un processus de branchement sous-critique en environnement aléatoire a la propriété suivante*

$$\lim_{n\to\infty} \mathbb{P}(Z_n = k | Z_n > 0, r_1, .., r_{n-1}) = b_k(r)$$

pour une suite $(b_k(r))$ *qui détermine une loi de probabilité non-dégénérérée.*

Cas surcritique : $\mathbb{E}(\log \mathbf{f}_0'(\mathbf{1})) > \mathbf{0}$
Tanny [56] établit un premier résultat sur la vitesse de croissance Z_n sur $\{Z_n \underset{n\to\infty}{\to} \infty \text{ p.s. }\}$:

Proposition 4.4.2. —

$$\frac{1}{n} \log Z_n \underset{n\to\infty}{\to} \mathbb{E}(\log f_0'(1))$$

En analogie avec le théorème de Kesten et Stigum sur les processus de Galton-Watson surcritiques en environnement déterministe, on a [4] :

Théorème 4.4.5. — *Soit* $W_n := \frac{Z_n}{\prod_{i=0}^{n-1} f_i'(1)}$ $n \geq 1$. *La famille* $\{W_n, (r, Z_0, .., Z_n)\}$ *constitue une martingale positive et donc*

$$\lim_{n\to\infty} W_n = W$$

existe p.s. Avec une condition de régularité supplémentaire,
(i) $\lim_{n\to\infty} \mathbb{E}\left(e^{-uW_n} | r\right) = \psi(u, r)$ *où* $\psi(u, r)$ *est l'unique solution de l'équation fonctionnelle : (*T *est l'opérateur de shift du processus définissant l'environnement)*

$$\psi(u, r) = f_0\left(\psi\left(\frac{u}{f_0'(1)}, T(r)\right)\right)$$

parmi celles satisfaisant $\lim_{u\to 0} u^{-1}(1 - \psi(u, r)) = 1$.
(ii) $\mathbb{E}(W|r) = 1$
(iii) $\mathbb{P}(W = 0|r) = q(r)$

4.4.2. Etude de la variance et optimisation de l'algorithme. —

4.4.2.1. Etude de la variance. —

Proposition 4.4.3. — *$\widehat{P}$ est trivialement un estimateur sans biais de variance*

$$var(\widehat{P}) = \frac{\mathbb{P}(A)^2}{N}\left(\frac{1}{P_0}-1\right)\sum_{i=0}^{M}\mathbb{E}\left(\frac{1}{\xi}\right)^i \tag{70}$$

Démonstration. — Tout d'abord, $\operatorname{var}(\widehat{P}) = \frac{\mathbb{P}(A)^2}{N}\operatorname{var}(W_{M+1})$. Ainsi, le calcul de $\operatorname{var}(\widehat{P})$ se ramène à celui de $\operatorname{var}(W_{M+1})$.

En appliquant (66) à $X = W_{M+1} = \frac{Z_{M+1}}{P_0\prod_{i=1}^{M}\xi_i}$ et $\mathcal{F} = \sigma(Z_M, r)$ et puisque W_M est une martingale, on a

$$\begin{aligned}
\operatorname{var}(W_{M+1}) &= \operatorname{var}(W_M) + \mathbb{E}\left(\frac{1}{[P_0\prod_{i=1}^{M}\xi_i]^2}\operatorname{var}(Z_{M+1}|\sigma(Z_M,r))\right)\\
&= \operatorname{var}(W_M) + \mathbb{E}\left(\frac{Z_M}{[P_0\prod_{i=1}^{M}\xi_i]^2}\operatorname{var}(X_M|\sigma(Z_M,r))\right)\\
&= \operatorname{var}(W_M) + \mathbb{E}\left(\frac{Z_M}{[P_0\prod_{i=1}^{M}\xi_i]^2}R_M P_0(1-P_0)\right)\\
&\qquad\qquad\text{puisque } X_M^i \sim Bin(R_M, P_0)\\
&= \operatorname{var}(W_M) + (1-P_0)\mathbb{E}\left(\frac{W_M}{P_0\prod_{i=1}^{M}\xi_i}\right)
\end{aligned}$$

Par une récurrence descendante,

$$\operatorname{var}(W_{M+1}) = \operatorname{var}(W_0) + (1-P_0)\sum_{k=0}^{M}\mathbb{E}\left(\frac{W_k}{P_0\prod_{i=1}^{k}\xi_i}\right)$$

Il reste à calculer l'espérance de $\frac{W_k}{P_0\prod_{i=1}^{k}\xi_i}$ par récurrence et à l'étape k on a

$$\mathbb{E}\left(\frac{W_k}{P_0\prod_{i=1}^{k}\xi_i}\right) = \frac{1}{P_0}\mathbb{E}\left(\frac{1}{\xi}\right)^k \tag{71}$$

Clairement la formule est vraie pour $k = 0$: $\mathbb{E}\left(\frac{W_0}{P_0}\right) = \frac{1}{P_0}$. Pour le passage de $k-1$ à k, supposons (71) pour $k-1$.

$$\begin{aligned}\mathbb{E}\left(\frac{W_k}{P_0\prod_{i=1}^{k}\xi_i}\right) &= \mathbb{E}\left(\frac{Z_{k-1}}{[P_0\prod_{i=1}^{k-1}\xi_i]^2\xi_k}\mathbb{E}(X_{k-1}|\sigma(Z_{k-1},r))\right) \\ &= \mathbb{E}\left(\frac{W_{k-1}}{P_0\prod_{i=1}^{k}\xi_i}\right) = \mathbb{E}\left(\frac{W_{k-1}}{P_0\prod_{i=1}^{k-1}\xi_i}\right)\mathbb{E}\left(\frac{1}{\xi}\right)\end{aligned}$$

puisque les R_i sont i.i.d. Et on obtient finalement le résultat en utilisant l'hypothèse de récurrence. □

4.4.2.2. Optimisation en les paramètres. —

Remarque 4.4.2. — *D'abord par une inégalité de Jensen,*

$$\mathbb{E}\left(\frac{1}{R}\right)\mathbb{E}(R) \geq 1 \quad \textit{et} \quad \mathbb{E}\left(\frac{1}{R}\right) = \frac{2k+1-\mathbb{E}(R)}{k(k+1)}$$

Alors puisque $R = k$ ou $k+1$ et $\lfloor 1/P_0 \rfloor = k$, on a

$$\frac{k}{k+1} \leq m = \mathbb{E}(RP_0) < 1 \quad (R = k) \ \textit{ou}\ 1 < m \leq \frac{k+1}{k} \quad (R = k+1)$$

Comme en Section 3.4.3, un algorithme optimal se déduit par minimisation de la variance de $\widehat{P}$ pour un budget C donné. Le coût (moyen) est maintenant

$$C = N\left[h(P_0)\sum_{i=0}^{M}\mathbb{E}(\xi)^i + Mh(1)\right] \tag{72}$$

En négligeant le coût introduit par la génération de l'environnement aléatoire, nous supposons, dans la suite, que

$$C = N\left[h(P_0)\sum_{i=0}^{M}\mathbb{E}(\xi)^i\right], \qquad (h(1) \ll h(P_0)). \tag{73}$$

En pratique, ce surcoût est effectivement négligeable : par exemple dans le cas de la simulation de trajectoires d'Ornstein-Uhlenbeck.

Une fois les cas triviaux écartés, on peut supposer que $\mathbb{E}(R) \neq \frac{1}{P_0}$ et $\mathbb{E}(\frac{1}{R}) \neq P_0$, alors la variance s'écrit

$$\mathrm{var}(\widehat{P}) = \frac{\mathbb{P}(A)^2}{N}(\frac{1}{P_0} - 1)\frac{\mathbb{E}(\frac{1}{\xi})^{M+1} - 1}{\mathbb{E}(\frac{1}{\xi}) - 1}$$

et le coût a la forme suivante

$$C = Nh(P_0)\frac{\mathbb{E}(\xi)^{M+1}-1}{\mathbb{E}(\xi)-1}$$

La valeur optimale de N est donnée par ce dernier :

$$N = \frac{C}{h(P_0)}\frac{\mathbb{E}(\xi)-1}{\mathbb{E}(\xi)^{M+1}-1}$$

et il s'agit de minimiser l'expression suivante

$$\mathrm{var}(\widehat{P}) = \frac{\mathbb{P}(A)^2}{C}h(P_0)\left(\frac{1}{P_0}-1\right)\frac{\mathbb{E}(\frac{1}{\xi})^{M+1}-1}{\mathbb{E}(\frac{1}{\xi})-1}\frac{\mathbb{E}(\xi)^{M+1}-1}{\mathbb{E}(\xi)-1}$$

sous la contrainte $\mathbb{E}(R) = 2k+1-k(k+1)\mathbb{E}(\frac{1}{R})$. Soient
- $u = \mathbb{E}(\xi)-1$ et $v = \mathbb{E}(\frac{1}{\xi})-1$,
- $\rho = \frac{(k+\delta)^2}{k(k+1)}$ et $\alpha = \frac{\delta(1-\delta)}{k(k+1)}$.

Avec ces notations, la contrainte devient $v = \alpha - \rho u$ et par les multiplicateurs de Lagrange, il s'agit finalement de résoudre

$$F(M,u) = \rho F(M,v) \tag{74}$$

où $F(M,u) = \frac{(1+u)^M(Mu-1)+1}{u((1+u)^{M+1}-1)}$.

Remarque 4.4.3. — *Notons que u et v dépendent de M et varient respectivement dans les intervalles suivants $[-\frac{\delta}{k+\delta}, \frac{1-\delta}{k+\delta}]$ et $[-\frac{1-\delta}{k+\delta}, \frac{\delta}{k}]$.*

Nous déduisons les résultats asymptotiques suivants :

Proposition 4.4.4. — *Asymptotiquement en M, le problème d'optimisation a seulement trois solutions :*
- $u \to u_1 = 0$ *et* $v \to v_1 = \alpha$
- $u \to u_2 = \frac{\alpha}{\rho}$ *et* $v \to v_2 = 0$
- $u \to u_3 = \frac{1+\alpha-\rho}{2\rho}$ *et* $v \to v_3 = \frac{\alpha+\rho-1}{2}$

Démonstration. — • Supposons d'abord que u et v ne convergent pas vers 0 quand $M \to +\infty$. Ainsi nous devons analyser trois cas (le cas $u < 0, v < 0$ étant exclu par la Remarque4.4.2) :
- $u < 0$ et $v > 0$

– $u > 0$ et $v < 0$
– $u > 0$ et $v > 0$

Cas 1 : Si $u < 0$ et $v > 0$ (ou par symétrie $u > 0$ et $v < 0$), puisque u ne converge pas vers 0, il existe une sous-suite de u telle que $u < -\epsilon \ \forall M$. Alors asymptotiquement en M, (74) est équivalent à

$$-\frac{1}{u} = \rho \frac{M}{1+v}$$

ce qui est absurde puisque $v \neq -1$ et u ne converge pas vers 0.

Cas 2 : Si $u > 0$ et $v > 0$, de la même façon, il existe une sous-suite de u et il existe une sous-suite de v telles que $u > \epsilon$ et $v > \epsilon$. Alors asymptotiquement en M, (74) est équivalent à

$$\frac{M}{1+u} = \rho \frac{M}{1+v}$$

Et étant donnée la contrainte, u doit nécessairement converger ves $\frac{1+\alpha-\rho}{2\rho}$ et v vers $\frac{\alpha+\rho-1}{2}$. Puisque la suite u a une seule valeur d'adhérence et reste dans un compact, elle converge vers $u_3 = \frac{1+\alpha-\rho}{2\rho}$ et similairement v converge vers $v_3 = \frac{\alpha+\rho-1}{2}$. Pour garantir $u > 0$ et $v > 0$, δ doit être dans $[\frac{k}{1+2k}, \frac{1}{2}] := [\delta_1, \delta_2]$.

• Supposons maintenant que u converge vers 0 quand $M \to +\infty$. Nous avons deux cas à analyser : Mu borné ou pas.

Si Mu n'est pas borné, on peut en extraire une sous-suite divergente. Alors

$$F(M,u) = \frac{1 - \frac{1}{Mu} + \frac{1}{Mu(1+u)^M}}{1 - \frac{1}{(1+u)^{M+1}}} \frac{1}{u(1+u)} \sim \frac{1}{u(1+u)}$$

et donc (74) est équivalent à

$$\frac{1}{u(1+u)} = \rho \frac{M}{1+v}$$

ce qui est absurde.

Ainsi Mu est borné et on peut en extraire une sous-suite qui converge vers 0 ou vers une constante C_1. Dans le premier cas,

$$F(M,u) \sim \frac{e^{Mu}(Mu-1)+1}{u(e^{Mu}-1)} \sim \frac{(1+Mu)(Mu-1)+1}{u(1+Mu-1)} \sim M$$

et (74) est équivalent à

$$M = \rho \frac{M}{1+v}$$

ce qui est absurde puisque $v > -1$. Donc la sous-suite de Mu converge vers une constante C_1 solution de

$$\frac{e^{C_1}(C_1-1)+1}{C_1(e^{C_1}-1)} = \frac{\rho}{1+\alpha}$$

Soit $f(x) = \frac{e^x(x-1)+1}{x(e^x-1)} - \frac{\rho}{1+\alpha}$. Notons que f est strictement croissante, converge vers $1 - \frac{\rho}{1+\alpha} \geq 0$ quand $x \to +\infty$ et vers $-\frac{\rho}{1+\alpha} \leq 0$ quand $x \to -\infty$ pour $\delta \leq \delta_2$. Ainsi C_1 est défini de manière unique et la suite Mu a un seul point limite et reste dans un compact, donc elle converge vers C_1, u converge vers $u_1 = 0$ et v vers $v_1 = \alpha$.

Par symétrie on obtient un résultat équivalent pour v : v converge vers $v_2 = 0$ et u vers $u_2 = \frac{\alpha}{\rho}$. □

Rappelons que nous voulons minimiser la variance de $\widehat{P}$ à coût fixé.

Proposition 4.4.5. — *Soit $\delta_0 := \sqrt{k}(\sqrt{k+1} - \sqrt{k})$. Alors asymptotiquement en M,*
- *si $\delta \in [0, \delta_0]$, (u_1, v_1) minimise la variance,*
- *si $\delta \in [\delta_0, 1]$, (u_2, v_2) minimise la variance.*

Démonstration. — Soient $\beta = \frac{P^2}{C} h(P_0) \left(\frac{1}{P_0} - 1\right)$ et pour $i = 1, 2, 3$,

$$\mathrm{var}_i = \beta \frac{[1+u_i]^{M+1} - 1}{u_i} \frac{[1+v_i]^{M+1} - 1}{v_i}$$

Tout d'abord, comparons var_1 et var_2. Pour $i = 1, 2$,

$$\mathrm{var}_i \sim \beta \frac{e^{(M+1)(1+C_i-1/M)} - 1}{u_i} \frac{[1+v_i]^{M+1}}{v_i} \sim \beta M \frac{e^{C_i} - 1}{C_i} \frac{[1+v_i]^{M+1}}{v_i}$$

Donc

$$\frac{\mathrm{var}_1}{\mathrm{var}_2} \sim \frac{e^{C_1} - 1}{C_1} \frac{C_2}{e^{C_2} - 1} \frac{u_2}{v_1} \left(\frac{1+v_1}{1+u_2}\right)^{M+1}$$

Mais $\frac{1+v_1}{1+u_2} < 1 \Leftrightarrow \delta \leq \delta_0 := \sqrt{k}(\sqrt{k+1} - \sqrt{k})$. Par conséquent,

$$\mathrm{var}_1 \leq \mathrm{var}_2 \Leftrightarrow \delta \leq \delta_0$$

Maintenant remarquons que var_3 n'est pas défini sur $[0,\delta_1]\cup[\delta_2,1]$ et δ_0 appartient clairement à $]\delta_1,\delta_2[$. Donc pour $\delta\in[\delta_1,\delta_2]$, il suffit de comparer

- var_1 et var_3 sur $[\delta_1,\delta_0]$,
- var_2 et var_3 sur $[\delta_0,\delta_2]$.

Mais $\frac{\mathrm{var}_1}{\mathrm{var}_3}\sim M\frac{e^C-1}{C}\frac{u_3v_3}{\alpha}\left[4\rho\frac{1+\alpha}{(1+\alpha+\rho)^2}\xrightarrow[M\to\infty]{}\right]^{M+1}$ et $4\rho\frac{1+\alpha}{(1+\alpha+\rho)^2}=\frac{4\frac{\rho}{1+\alpha}}{(1+\frac{\rho}{1+\alpha})^2}<1$. Ainsi $\frac{\mathrm{var}_1}{\mathrm{var}_3}\xrightarrow[M\to\infty]{}0$ et $\mathrm{var}_1\leq\mathrm{var}_3$ asymptotiquement en M. De la même façon, $\frac{\mathrm{var}_2}{\mathrm{var}_3}\xrightarrow[M\to\infty]{}0$. $\square$

4.4.2.3. Criticalité. — Par des arguments élémentaires de convexité,

Proposition 4.4.6. — *Asymptotiquement en M,*
- *pour tout $\delta\in[0,\delta_0]$, on est dans le cas sous-critique,*
- *pour tout $\delta\in[\delta_0,1]$, on est dans le cas surcritique.*

Proposition 4.4.7. — *Le cas critique est défini par $p_c:=\mathbb{P}(R=k)$*

$$p_c=\frac{\log\left(\frac{k+\delta}{k+1}\right)}{\log\left(\frac{k}{k+1}\right)}$$

et
- $p>p_c\Leftrightarrow\mathbb{E}(\log f_1'(1))<0\Leftrightarrow$ *cas sous-critiuqe,*
- $p=p_c\Leftrightarrow\mathbb{E}(\log f_1'(1))=0\Leftrightarrow$ *cas critique,*
- $p<p_c\Leftrightarrow\mathbb{E}(\log f_1'(1))>0\Leftrightarrow$ *cas surcritique.*

Notons que

$$p\underset{M\to\infty}{\rightarrow}p_a:=\begin{cases}1-\delta & \text{pour }\ \delta<\delta_0\\ \frac{k(1-\delta)}{k+\delta} & \text{pour }\ \delta>\delta_0\end{cases}$$

Traçons les graphes de p_a de p_c en Figure 5 et nous retrouvons par un autre moyen la distinction entre les cas sous-critique et surcritique.

4.4.2.4. Intervalle de confiance. — De même que dans les modèles précédents, nous cherchons à borner

$$\mathbb{P}\left(|\widehat{P}-\mathbb{P}(A)|/\mathbb{P}(A)\geq\alpha\right)$$

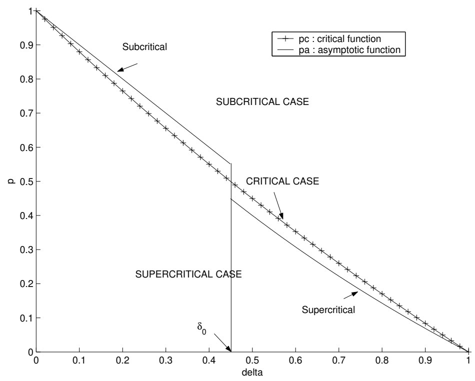


FIG. 5. Graphe p=f(δ) asymptotiquement en M

Posons alors $\psi_i(u) := R_i \log(P_0 e^u + 1 - P_0)$, $f_{0,i} := f_0 \circ \cdots \circ f_i$ et $\psi_{0,i} := \psi_0 \circ \cdots \circ \psi_i$, prenons $\xi_i = R_i P_0$ pour tout $i = 1, \cdots, M$ et $\xi_0 = P_0$,

$$\mathbb{E}\left(e^{uW_{M+1}}\right) = \mathbb{E}\left[P_0 f_{1,M}\left(e^{\frac{u}{\prod_{i=0}^{M}\xi_i}}\right) + 1 - P_0\right] = \mathbb{E}\left[\exp\left\{\psi_{0,M}\left(u/\prod_{i=0}^{M}\xi_i\right)\right\}\right]$$

Finalement,

$$\mathbb{P}\left(\frac{|\widehat{P} - \mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq \mathbb{E}\left(\exp\left\{\inf_{u>0} F^{+}_{M+1}(u)\right\}\right)^N + \mathbb{E}\left(\exp\left\{\inf_{u<0} F^{-}_{M+1}(u)\right\}\right)^N$$

où

$$\begin{cases} F^{+}_{M+1}(u) := \psi_{0,M}\left(\frac{u}{\prod_{i=0}^{M}\xi_i}\right) - (1+\alpha)\,u \\ F^{-}_{M+1}(u) := \psi_{0,M}\left(\frac{u}{\prod_{i=0}^{M}\xi_i}\right) - (1-\alpha)\,u \end{cases}$$

Une fois de plus, nous nous intéressons à la transformée de Laplace de W_{M+1} et nous sommes conduits à étudier le comportement des itérées de la fonction aléatoire ψ.

4.4.3. La transformée de Laplace de W_{n+1}. —

4.4.3.1. Estimées de $\psi_{0,M}$. — Rappelons que nous cherchons à estimer $\psi_{0,M} = \psi_0 \circ \cdots \circ \psi_M$ et donc nous devons itérer

$$\begin{cases} \psi^{(1)}(u) = k\log(P_0 e^u + 1 - P_0) \\ \quad \text{et} \\ \psi^{(2)}(u) = (k+1)\log(P_0 e^u + 1 - P_0) \end{cases}$$

Heuristique : Quel que soit le cas étudié ($m > 1$, $m < 1$), nous utiliserons seulement deux fonctions pour borner ψ. Plus précisément, par exemple, pour $u \geq 0$, jusqu'à un certain a (intersection entre l'homographie inférieure ($m < 1$) et la première bissectrice) nous bornerons ψ par une fonction homographique et ensuite par la direction asymptotique dans les autres cas.

- Pour $u \geq 0$. Soient
  - $m^{(1)} = kP_0 < 1$ et $m^{(2)} = (k+1)P_0 > 1$,
  - $a^+ := \arg_{u \neq 0}\{h(u; m^{(1)}, \frac{1-P_0}{2}) = u\} = 2\frac{1-m^{(1)}}{1-P_0}$,
  - $b^+ := h(a^+; m^{(2)}, \frac{1-P_0}{2})$,
  - $u^+ := h^{-1}_{M+1}(a^+; m^{(2)}, \frac{1-P_0}{2})$.

En suivant la démarche adoptée en Section 4.2.2.1 dans une version simplifiée, on déduit

$$\psi^{(1)}(u) \leq \begin{cases} h\left(u; m^{(1)}, \frac{1-P_0}{2}\right) & \text{pour } 0 \leq u \leq a^+ \\ g(u; k, a^+(1-k)) & \text{pour } u \geq a^+ \end{cases}$$

et

$$\psi^{(2)}(u) \leq \begin{cases} h\left(u; m^{(2)}, \frac{1-P_0}{2}\right) & \text{pour } 0 \leq u \leq a^+ \\ g(u; k+1, b^+ - a^+(k+1)) & \text{pour } u \geq a^+ \end{cases}$$

ce qui conduit à la proposition suivante

Proposition 4.4.8. — *Si $a_0 = a^+$ et pour tout $i = 1 \cdots M$,*

$$a_i := \begin{cases} a^+ & si \;\; R_i = k \\ b^+ & si \;\; R_i = k+1 \end{cases}$$

$$\psi_{0,M}(u) \leq \begin{cases} h^+_{0,M}(u) & pour \;\; 0 \leq u \leq u^+ \\ g^+_M(u) & pour \;\; u^+ \leq u \leq a^+ \\ \prod_{i=0}^M R_i u + \sum_{i=0}^M \left(\prod_{j=0}^{i-1} R_j\right) a_i - \sum_{i=0}^M \left(\prod_{j=0}^{i} R_j\right) a^+ & pour \;\; u \geq a^+ \end{cases}$$

où
- $h^+_{0,M}(u) = u\left(\prod_{i=0}^M \xi_i\right) / \left[1 - \frac{1-P_0}{2} u \left(1 + \sum_{i=1}^M \xi_i \cdots \xi_M\right)\right]$ *et*
- $g^+_M(u) = \begin{cases} h^+_{0,M}(u) & avec\ proba \;\; \mathbb{P}\left(0 \leq h^+_{0,M}(u) \leq a^+\right) \\ b^+ & avec\ proba \;\; 1 - \mathbb{P}\left(0 \leq h_{0,M}(u) \leq a^+\right) \end{cases}$

- Pour $u \leq 0$. Soit

- $a^- := \arg_{u\neq 0}\{h(u; m^{(2)}, 1-P_0) = u\} = \frac{1-m^{(2)}}{1-P_0}$,
- $b^- := h(a^-; m^{(1)}, 1-P_0)$,
- $u^- := h^{-1}_{M+1}(a^-; m^{(1)}, 1-P_0)$.

De la même façon, on déduit

$$\psi^{(1)}(u) \leq \begin{cases} h\left(u; m^{(1)}, 1-P_0\right) = \frac{m^{(1)}u}{1-(1-P_0)u} & \text{pour } \; a^- \leq u \leq 0 \\ b^- \quad \text{pour } \; u \leq a^- \end{cases}$$

et

$$\psi^{(2)}(u) \leq \begin{cases} h\left(u; m^{(2)}, 1-P_0\right) & \text{pour } \; a^- \leq u \leq 0 \\ b^- \quad \text{pour } \; u \leq a^- \end{cases}$$

ce qui conduit à la proposition suivante

Proposition 4.4.9. —

$$\psi_{0,M}(u) \leq \begin{cases} h^-_{0,M}(u) & \textit{pour } \; u^- \leq u \leq 0 \\ g^-_M(u) & \textit{pour } \; a^- \leq u \leq u^- \\ b^- & \textit{pour } \; u \leq a^- \end{cases}$$

où

- $h^-_{0,M}(u) = u\left(\prod_{i=0}^M \xi_i\right) / \left[1-(1-P_0)u\left(1+\sum_{i=1}^M \xi_i\cdots\xi_M\right)\right]$ *et*
- $g^-_M(u) = \begin{cases} h^-_{0,M}(u) & \textit{avec proba } \; \mathbb{P}\left(b \leq h^-_{0,M}(u) \leq 0\right) \\ b^- & \textit{avec proba } \; 1-\mathbb{P}\left(b \leq h^-_{0,M}(u) \leq 0\right) \end{cases}$

4.4.3.2. Sur la marche aléatoire sur le groupe affine et quelques conséquences. — Nous souhaitons estimer

$$\mathbb{P}\left(0 \leq h^+_{0,M}(u) \leq a\right) \quad \text{pour } \; u \geq 0 \;\; \text{et } \; \mathbb{P}\left(b \leq h^-_{0,M}(u) \leq 0\right) \quad \text{pour } \; u \leq 0$$

Rappelons que $\xi_i = R_i P_0$, posons $v = \frac{1-P_0}{2}$ et

$$\begin{cases} x_{n+1}(y_0) = \xi_0\cdots\xi_n y_0 + v[1+\sum_{i=1}^n \xi_i\cdots\xi_n] \\ y_{n+1}(y_0) = \xi_0\cdots\xi_n y_0 + v[1+\sum_{i=0}^{n-1} \xi_0\cdots\xi_i] \end{cases}$$

Marche aléatoire sur le groupe affine : Considérons les transformations affines $g_k(x) = \xi_k x + v$ ($x \in \mathbb{R}$) et la marche aléatoire $(g_0 \circ g_1 \circ \cdots \circ g_n)$ sur le groupe affine. Immédiatement

$$\begin{cases} g_0 \circ g_2 \circ \cdots \circ g_n(x) = y_{n+1}(x) \\ g_n \circ g_{n-1} \circ \cdots \circ g_0(x) = x_{n+1}(x) \end{cases}$$

et

Lemme 4.4.1. — *(i)* $y_n \overset{\mathcal{L}}{=} x_n$
(ii) $\{0 \leq h_{0,n}(u) \leq a\} = \{0 \leq y_{n+1}\,(1/a) \leq 1/u\}$
(iii) $\{b \leq h_{0,n}(u) \leq 0\} = \{\frac{1}{u} \leq y_{n+1}\,(1/b) \leq 0\}$

D'où le lien entre la marche aléatoire sur le groupe des transformations homographiques et la marche aléatoire sur le groupe affine.

Par conséquent, nous sommes intéressés par les propriétés asymptotiques des lois $g_0 \circ g_1 \circ \cdots \circ g_n$ qui clairement sont caractérisées par la loi de ξ. De même que pour les processus de branchement en milieu aléatoire, on distingue trois régimes différents déterminés par le signe de $\mathbb{E}(\log \xi)$.

Pour plus de détails sur les marches aléatoires sur le groupe affine, le lecteur est renvoyé, par exemple, à Vervaat [57], Brandt [16] pour le cas unidimensionnel et à [42], Bougerol and Picard [14], Babillot et al. [7] pour le cas d-dimensionnel.

Cas sous-critique : $\mathbb{E}\,(\log \xi) < \mathbf{0}$
Tout d'abord, Diaconis et Freedman [21] établissent que le processus «backward» $\psi_{0,M}(u) = \psi_0 \circ \cdots \circ \psi_M(u)$ converge presque sûrement, avec une vitesse exponentielle vers une variable aléatoire limite qui ne dépend pas du point de départ u.

Concernant la marche aléatoire sur le groupe affine dans notre cas particulier ($\xi \overset{\mathcal{L}}{=} Ber$ sur $\{kP_0, (k+1)P_0\}$), nous avons par Maksimov [51]

$$y_n \xrightarrow[n\to\infty]{} v(1+D)$$

où $D = \sum_{i=0}^{\infty} \xi_0 \cdots \xi_i$. En général, rien n'est connu sur la distribution de D, à ceci près qu'elle satisfait l'équation fonctionnelle suivante

$$F(\lambda) = pF\left(\frac{\lambda}{kP_0} - 1\right) + (1-p)F\left(\frac{\lambda}{(k+1)P_0} - 1\right)$$

avec $p = \mathbb{P}(R = k)$.

Cependant, Kesten [42] donne un résultat asymptotique sur la distribution de D : il existe $C > 0$ et $\kappa > 0$,

$$\mathbb{P}(D \geq x) \underset{x\to\infty}{\sim} \frac{C}{x^\kappa};$$

d'où l'on déduit

Corollaire 4.4.1. — *Posons $\kappa = -\log[p(1-p)]/\log[k(k+1)P_0^2]$,*

$$\mathbb{P}\left(0 \leq h_{0,n}(u) \leq a\right) \underset{n\to\infty}{\sim} \mathbb{P}\left(v(1+D) \leq \frac{1}{u}\right) \underset{n\to\infty, u\to 0}{\sim} 1 - \frac{C}{(\frac{1}{uv}-1)^\kappa}$$

Cas critique : $\mathbb{E}\left(\log\xi\right) = \mathbf{0}$
Maksimov affirme dans [51] que la loi de $\log y_n/\sqrt{n}$ approche, quand $n \to \infty$, la distribution normale tronquée pour tout point de départ de la marche.

Corollaire 4.4.2. — *Soient Z une variable aléatoire suivant la loi normale tronquée et $\sigma^2 = var(\xi)$.*

$$\mathbb{P}\left(0 \leq h_{0,n}(u) \leq a\right) \underset{n\to\infty}{\sim} \mathbb{P}\left(Z \leq -\frac{1}{\sqrt{n}}\log u\right) \underset{n\to\infty}{\sim} \left[1 - e^{-\frac{\log u^2}{n}}\right]^{\frac{1}{2}}$$

Cas surcritique : $\mathbb{E}\left(\log\xi\right) > \mathbf{0}$
Selon Maksimov [51], la distribution de $(\frac{y_n}{v})^{1/\sqrt{n}}e^{-M\sqrt{n}}$ approche, lorsque n tend vers $+\infty$, la distribution log normale pour tout point de départ de la marche.

Corollaire 4.4.3. — *Soient $Z \overset{\mathcal{L}}{=} N(0,1)$ et $\sigma^2 = var(\xi)$.*

$$\mathbb{P}\left(0 \leq h_{0,n}(u) \leq a\right) \underset{n\to\infty}{\sim} \frac{1}{2}\exp\left\{-\frac{1}{2\sigma^2}n\left[\frac{1}{n}\log(uv) + \mathbb{E}(\log\xi)\right]\right\}$$

4.4.4. Intervalle de confiance. — Finalement, avec les majorations obtenues dans la section précédente, on déduit

Théorème 4.4.6. — *Pour α assez petit,*

$$\mathbb{P}\left(\frac{|\widehat{P} - \mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq h_+^N + h_-^N \tag{75}$$

où

$$\begin{cases} h_+ := \mathbb{E}\left(\exp\left\{-\frac{2}{1-P_0}\frac{1}{\sum_{i=0}^{M}\xi_0^{-1}..\xi_i^{-1}}\left(\sqrt{1+\alpha}-1\right)^2\right\}\right) \\ h_- := \mathbb{E}\left(\exp\left\{-\frac{1}{1-P_0}\frac{1}{\sum_{i=0}^{M}\xi_0^{-1}..\xi_i^{-1}}\left(1-\sqrt{1-\alpha}\right)^2\right\}\right) \end{cases}$$

avec $N = C/\left[h(P_0)\frac{\mathbb{E}(\xi)^{M+1}-1}{\mathbb{E}(\xi)-1} + h(1)M\right]$.

Par une inégalité de Jensen,

$$\mathbb{P}\left(\frac{|\widehat{P}-\mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq g_+^N + g_-^N \tag{76}$$

où

$$\begin{cases} g_+ := \exp\left\{-\frac{2}{1-P_0}\frac{1}{\mathbb{E}(\xi^{-1})}\frac{\mathbb{E}(\xi^{-1})-1}{\mathbb{E}(\xi^{-1})^{M+1}-1}\left(\sqrt{1+\alpha}-1\right)^2\right\} \\ g_- := \exp\left\{-\frac{1}{1-P_0}\frac{1}{\mathbb{E}(\xi^{-1})}\frac{\mathbb{E}(\xi^{-1})-1}{\mathbb{E}(\xi^{-1})^{M+1}-1}\left(1-\sqrt{1-\alpha}\right)^2\right\} \end{cases}$$

En utilisant le même type d'argument que pour le Corollaire 4.2.1 :

Corollaire 4.4.4. — *Pour* $c_3 = 1/[8(1-P_0)]$ *et* α *tel que dans le Théorème 4.4.6,*

$$\mathbb{P}\left(\frac{|\widehat{P}-\mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq 2\mathbb{E}\left(\exp\left\{-c_3\frac{\alpha^2}{\sum_{i=0}^{M}\xi_0^{-1}..\xi_i^{-1}}\right\}\right)^N$$

Par une inégalité de Jensen, on obtient la borne supérieure

$$2\exp\left\{-c_3\frac{C\ \alpha^2}{\left[h(P_0)\frac{\mathbb{E}(\xi)^{M+1}-1}{\mathbb{E}(\xi)-1}+h(1)M\right]}\frac{1}{\mathbb{E}(\xi^{-1})}\frac{\mathbb{E}(\xi^{-1})-1}{\mathbb{E}(\xi^{-1})^{M+1}-1}\right\}$$

4.5. Illustration numérique

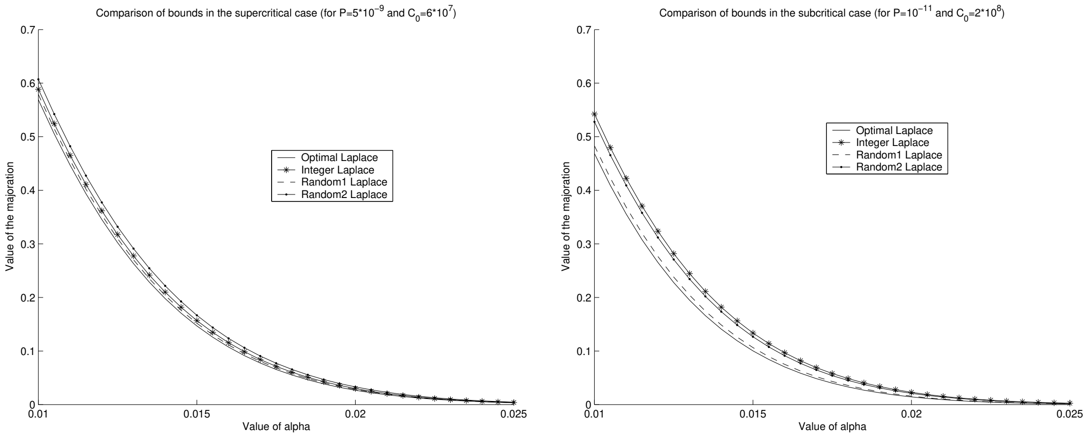
On a tracé en Figure 6 les bornes faisant intervenir la transformée de Laplace de W_{M+1} pour les différents modèles :

- modèle optimal : $m = 1$ (tracé plein),
- modèle déterministe où $m \neq 1$ (tracé $- * -$),
- modèle où un nouveau R est généré à chaque succès (tracé pointillé),
- modèle où un environnement aléatoire est généré (tracé $-\cdot-$),

dans deux cas

- pour $\mathbb{P}(A) = 5\ 10^{-9}$ et $C = 6\ 10^7$, alors m=1.055 (cas surcritique),
- pour $\mathbb{P}(A) = 10^{-11}$ et $C = 2\ 10^8$, alors m=0.924 (cas sous-critique).

Le tableau suivant présente la longueur de l'intervalle de confiance à 95%



1. Cas surcritique   2. Cas sous-critique

FIG. 6. Intervalle de confiance

donnée par la variance et la transformée de Laplace pour les différents modèles :

- modèle optimal (*Optimal*) : $m = 1$,
- modèle déterministe où $m \neq 1$ (*Déterministe*) : $m \neq 1$,
- modèle où un nouveau R est généré à chaque succès (*Aléatoire 1*),
- modèle où un environnement aléatoire est généré (*Aléatoire 2*),

dans deux cas

- pour $\mathbb{P}(A) = 3.5\ 10^{-11}$ et $C = 10^8$ (cas surcritique),
- pour $\mathbb{P}(A) = 5\ 10^{-11}$ et $C = 10^8$ (cas sous-critique).

	$\mathbb{P}(A) = 3.5\ 10^{-11}$ et $C = 10^8$		$\mathbb{P}(A) = 5.10^{-11}$ et $C = 10^8$	
	Cas surcritique : $m = 1.111$		Cas sous-critique : $m = 0.909$	
	Variance	Laplace	Variance	Laplace
Optimal	0.05080	0.03948	0.04938	0.03835
Déterministe	0.05702	0.04428	0.05428	0.04215
Aléatoire 1	0.05134	0.03990	0.04990	0.03878
Aléatoire 2	0.05388	0.04026	0.05235	0.03910

TAB3. *Longueur de l'intervalle de confiance à* 95%

On remarque que la longueur de l'intervalle de confiance donnée par la variance est plus grande que celle donnée par la transformée de Laplace, le premier algorithme aléatoire est le plus efficace alors que le modèle déterministe fournit les moins bons résultats.

4.6. Conclusion

Dans ce chapitre, la simplicité relative du modèle nous a permis d'établir des résultats explicites comme des bornes de Chernoff de l'erreur relative entre la probabilité d'intérêt $\mathbb{P}(A)$ et son estimateur $\widehat{P}$. En poursuivant les calculs et en utilisant l'heuristique présentée précédemment, on peut aussi déduire un théorème central limite ainsi que des bornes de Berry-Esseen. Nous avons ensuite étudié la sensibilité des bornes de Chernoff en fonction du choix des nombres de retirage R dans trois algorithmes différents, tous trois basés sur le modèle de branchement avec duplication des trajectoires : il apparaît alors que, lorsqu'on ne peut être exactement dans le cas critique (qui correspond à l'algorithme optimal), la manière façon de procéder est de considérer R comme une variable aléatoire que l'on génère à chaque succès dans la simulation. Cette procédure est d'ailleurs couramment utilisée en pratique, voir par exemple [3]. Cependant, le gain apporté par cet algorithme par rapport au modèle déterministe n'étant que peu significatif comme le montre l'application numérique précédente, nous utiliserons en pratique le modèle déterministe qui consiste à prendre l'entier le plus proche de la valeur optimale. Notons quand même que le surcoût engendré par le retirage n'a pas été pris en compte dans cette étude. Dans la pratique, par exemple la simulation de trajectoires d'Ornstein-Uhlenbeck, ce surcoût est négligeable.

4.7. Compléments

4.7.1. Sur l'intégrabilité exponentielle de W. — Notons tout d'abord que pour $u \leq 0$, $\mathbb{E}\left(e^{uW_n}\right) \leq 1$ et donc en notant W la loi limite de W_n lorsque $n \to +\infty$,

$$\mathbb{E}\left(e^{uW}\right) \leq 1, \qquad \text{pour} \quad u \leq 0.$$

Que se passe-t-il maintenant pour les valeurs de u positives ? Nous avons

$$\psi_{n+1}\left(\frac{u}{P_0 m^n}\right) = R \log \mathbb{E}\left(e^{uW_{n+1}}\right).$$

Par conséquent, toute information sur $\psi_{n+1}\left(\frac{u}{P_0 m^n}\right)$ nous fournira des résultats sur l'exponentielle intégrabilité de W_{n+1} et W. Le Théorème 4.2.1 nous permet alors de conclure :

Corollaire 4.7.1. — *La variable aléatoire $uW^{\gamma'}$ est exponentiellement intégrable, indépendamment de la valeur u, avec $\gamma' = \frac{\gamma}{\gamma-1} \geq 1$ c-à-d*

$$\mathbb{E}\left(e^{uW^{\gamma'}}\right) < +\infty \quad \textit{pour tout} \quad u \in \mathbb{R}.$$

4.7.2. Sur la transformée de Laplace de Z_n. — Grâce aux bornes obtenues sur la fonction génératrice de Z_n obtenues au chapitre précédent, nous pouvons aussi déduire des résultats sur l'intervalle de confiance, cependant moins précis que ceux obtenus dans ce chapitre.

En reprenant les notations et la démarche de la Section 4.1, il vient que pour tout $\lambda > 0$,

$$\begin{aligned}
\mathbb{P}\left(\widehat{P} \geq \mathbb{P}(A)(1+\alpha)\right) &= \mathbb{P}\left(\frac{1}{N}\sum_{i=1}^{N}\widetilde{Q}_i \geq \mathbb{P}(A)(1+\alpha)\right) \\
&= \mathbb{P}\left(e^{\lambda\sum_{i=1}^{N}\widetilde{Q}_i} \geq e^{\lambda N\mathbb{P}(A)(1+\alpha)}\right) \\
&\leq \exp\{-\lambda N\mathbb{P}(A)(1+\alpha)\}\,\mathbb{E}\left(e^{\lambda\widetilde{Q}}\right)^N \\
&\leq \exp\{-N[\lambda\mathbb{P}(A)(1+\alpha)-\gamma(\lambda)]\}
\end{aligned}$$

où $\gamma(\lambda) = \mathbb{E}\left(e^{\lambda\widetilde{Q}}\right)$ est la log-Laplace de $\widetilde{Q}$. L'optimisation en $\lambda > 0$ conduit à

$$\mathbb{P}\left(\widehat{P} \geq \mathbb{P}(A)(1+\alpha)\right) \leq \exp\left\{-N\sup_{\lambda>0}[\lambda\mathbb{P}(A)(1+\alpha)-\gamma(\lambda)]\right\}$$

De la même façon,

$$\mathbb{P}\left(\widehat{P} \leq \mathbb{P}(A)(1+\alpha)\right) \leq \exp\left\{-N\sup_{\lambda<0}[\lambda\mathbb{P}(A)(1-\alpha)-\gamma(\lambda)]\right\}$$

Posons γ^* la transformée de Cramér de γ : $\gamma^*(\tau) = \sup_\lambda[\lambda\tau-\gamma(\lambda)]$. Alors

$$\begin{aligned}
\mathbb{P}\left(\frac{|\widehat{P}-\mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) &\leq \exp\{-N\gamma^*(\mathbb{P}(A)(1-\alpha))\}+\exp\{-N\gamma^*(\mathbb{P}(A)(1+\alpha))\} \\
&\leq 2\exp\{-N\min\{\gamma^*(\mathbb{P}(A)(1+\alpha)),\gamma^*(\mathbb{P}(A)(1-\alpha))\}\}
\end{aligned}$$

Ainsi, ne pouvant obtenir une expression explicite de γ^*, nous allons chercher à l'estimer. En considérant les gradients des fonctions à optimiser, nous pouvons prouver que le suprémum est atteint pour des valeurs de $\lambda \geq 0$ proches de 0. Ainsi en utilisant les bornes de f_M obtenues au chapitre précédent et valables au voisinage de 1, il vient

$$\begin{cases}\gamma^*(\mathbb{P}(A)(1+\alpha)) \geq F(\mathbb{P}(A)(1+\alpha)), \\ \gamma^*(\mathbb{P}(A)(1-\alpha)) \geq G(\mathbb{P}(A)(1-\alpha)),\end{cases}$$

où

$$F(x) = \sup_{\lambda>0}\left[\lambda x - \ln\left(1+P_0\frac{e^{\lambda/R^M}-1}{1-M\alpha_1\left(e^{\lambda/R^M}-1\right)}\right)\right]$$

et en posant $u_0 = 1+[2/3(M-1)]^{3/2}$,

$$G(x) = \sup_{\lambda<0}\left[\lambda x - \ln\left(1-P_0\frac{\left(1-e^{\lambda/R^M}\right)\left[1-\alpha_1\left(1-e^{\lambda/R^M}\right)\right]}{u_0}\right)\right].$$

Finalement,

$$\mathbb{P}\left(\frac{|\widehat{P}-\mathbb{P}(A)|}{\mathbb{P}(A)} \geq \alpha\right) \leq 2\exp\left\{-N\min\left\{F(\mathbb{P}(A)(1+\alpha)), G(\mathbb{P}(A)(1-\alpha))\right\}\right\}.$$

On peut alors facilement déduire les expressions explicites mais complexes de $F(x)$ et $G(x)$. Nous avons tracé en Figure 7 les bornes supérieures obtenues grâce à la transformée de Laplace et la variance pour différentes valeurs de α, l'erreur relative de l'intervalle de confiance, avec $P = 10^{-9}$ et les valeurs optimales déduites de l'algorithme optimal. Notons que la borne supérieure

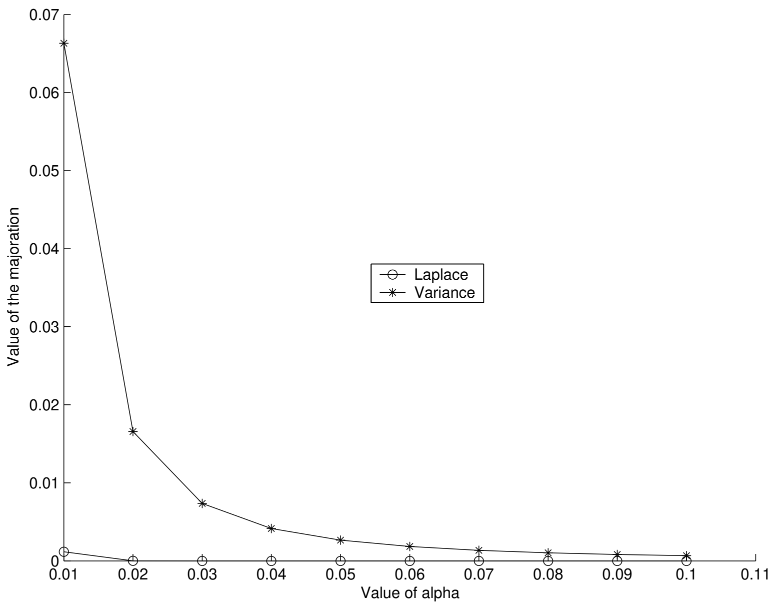


FIG. 7. Bornes supérieures

donnée par la transformée de Laplace est meilleure que celle donnée par la variance. Nous obtenons $\mathbb{P}\left(\frac{|\widehat{P}-P}{P}| \geq \alpha\right) \leq L$. Dans l'exemple précédent où $P = 10^{-9}$, si nous fixons $\alpha = 0.05$ et L proche de 0.01, alors le coût nécessaire correspondant est $3 * 10^7$ pour la variance et $3 * 10^6$ pour la transformée de Laplace.

CHAPITRE 5

MODÈLES D'APPRENTISSAGE

Rappelons qu'en pratique, les probabilités de transition sont généralement inconnues ; à défaut, nous savons faire évoluer les particules du niveau B_i vers le niveau suivant B_{i+1} (par ex. comportement markovien). Il devient donc impossible de déduire les nombres de retirage optimaux. Dans ce chapitre, nous présentons quatre modèles visant à résoudre ce problème inhérent au modèle de branchement avec duplication des trajectoires. Les trois premiers se placent dans un contexte similaire : les niveaux ont été fixés à l'avance, on ne connaît pas les probabilités de transition mais on sait faire évoluer le processus selon ces lois et déterminer s'il atteint ou non le seuil supérieur. Dans le quatrième modèle par contre, on suppose que l'on peut placer les niveaux où l'on veut, ce qui est fait au cours de la simulation de façon adaptative.

5.1. Modèle de branchement avec duplication des trajectoires à deux étapes

5.1.1. Cadre et introduction. — Dans cette section, les probabilités de transition sont inconnues mais appartiennent à $[a,b]$. Choisissons arbitrairement $M+1$ nombres P_i^0 dans $[a,b]$, posons

$$R_i^0 = \frac{1}{\sqrt{P_i^0 P_{i+1}^0}} \sqrt{\frac{1-P_{i+1}^0}{1-P_i^0}} \quad \text{pour} \quad i = 1 \cdots M$$

et introduisons les notations suivantes

$$r_0^0 = 1, \qquad r_i^0 = R_1^0 \cdots R_i^0, \quad i = 1 \cdots M.$$

Remarque 5.1.1. — *Supposer que les probabilités de transition sont dans un compact est essentiel : sinon rien ne peut être fait algorithmiquement. En pratique cette hypothèse est généralement implicite mais rien n'en est dit.*
De plus, la complexité des calculs nous empêche de mener une analyse précise et nous sommes conduits à faire une analyse asymptotique lorsque le coût C tend vers l'infini. Mais le fait de supposer les probabilités conditionnelles bornées implique que le ratio C/N est borné inférieurement et supérieurement ce qui nous permet de faire l'étude asymptotique lorsque N tend vers l'infini. Notons aussi que le nombre total de particules est en réalité une variable aléatoire : en pratique, un budget est fixé en début de simulation et nous génèrerons des particules jusqu'à ce que ce budget soit épuisé ; d'où le caractère aléatoire de N.

Nous proposons ici un algorithme en deux phases : la première est une phase d'apprentissage dans laquelle les nombres de retirage sont $(R_i^0)_{i=1\cdots M}$ tandis que dans la seconde nous utiliserons des estimateurs des $(R_i)_{i=1\cdots M}$ optimaux, estimateurs obtenus pendant la première phase. Plus précisément, nous disposons de N particules que nous scindons en deux groupes de ρ_N et $N-\rho_N$ particules respectivement. Le modèle de branchement avec duplication des trajectoires à deux étapes procède comme suit :

Phase 1 : Nous générons ρ_N particules. Les particules ayant atteint B_1 sont dupliquées en R_1^0 que nous faisons évoluer depuis B_1. Les particules ayant atteint B_2 sont dupliquées en R_2^0 sous-particules que nous faisons évoluer depuis $B_2\cdots$ Et ainsi de suite jusqu'à ce que A soit atteint. De l'étape i ($i=1\cdots M{+}1$), nous obtenons un estimateur $\widehat{P}_i$ de P_i (fraction des particules ayant réussi depuis B_{i-1}). Pour utiliser toute l'information dont nous disposons, nous améliorons l'algorithme en remplaçant $\widehat{P}_i$ par $\widetilde{P}_i := a\vee\widehat{P}_i\wedge b$ pendant la simulation, puisque $P_i\in[a,b]$. Cette substitution a l'avantageuse conséquence de nous prémunir de tout arrêt prématuré de l'algorithme. Maintenant, pour tout $i=1\cdots M$, posons

$$\widetilde{R}_i=\frac{1}{\sqrt{\widetilde{P}_i\widetilde{P}_{i+1}}}\sqrt{\frac{1-\widetilde{P}_{i+1}}{1-\widetilde{P}_i}},$$

ce sont les nombres de retirage de la deuxième phase.

Phase 2 : Nous générons $N-\rho_N$ particules. Les particules ayant atteint B_1 sont dupliquées en $\widetilde{R}_1$ que nous faisons évoluer depuis B_1. Les particules ayant atteint B_2 sont dupliquées en $\widetilde{R}_2$ sous-particules que nous faisons évoluer depuis $B_2 \cdots$ Et ainsi de suite jusqu'à ce que A soit atteint. De l'étape i $(i = 1 \cdots M+1)$, nous obtenons un estimateur $\widehat{P}_i'$ de P_i que nous remplacerons par $\widetilde{P}_i' := a \vee \widehat{P}_i' \wedge b$ pendant la simulation.

Dans la suite, les estimateurs tronqués seront notés avec une tilde $\sim$ et les

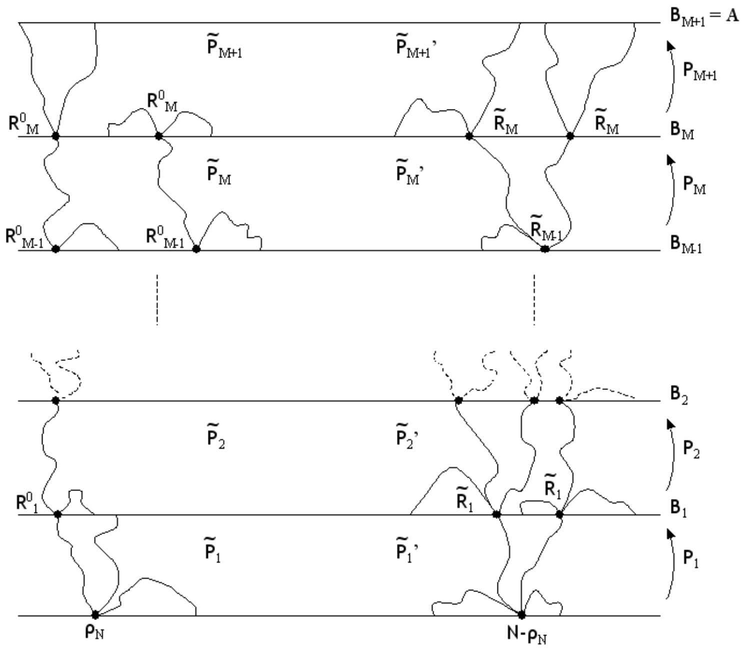


Fig. 1. Le modèle de branchement avec duplication des trajectoires à deux étapes

estimateurs non tronqués avec un chapeau $\wedge$.

<u>Hypothèses</u> :

1. Un étude rapide nous conduit à nous restreindre au cas où
   - $\rho_N \underset{N\to+\infty}{\longrightarrow} +\infty$,
   - $N - \rho_N \underset{N\to+\infty}{\longrightarrow} +\infty$,
   - $\rho_N = o(N - \rho_N)$.

2. De plus, comme expliqué en Remarque 5.1.1, les calculs étant trop complexes pour faire une analyse précise, nous mènerons une étude asymptotique lorsque le coût C (c-à-d quand N) tend vers l'infini. Nous souhaitons déterminer l'algorithme optimal. Après une analyse plus approfondie, nous prouvons que l'optimum est atteint pour
$$\left(\frac{\rho_N}{N}\right)^2 \approx \frac{1}{\rho_N}.$$
Plus précisément, pour comprendre mieux ce qui se passe et afin d'alléger la rédaction et rendre plus claire l'analyse, nous nous restreignons au cas où ρ_N est asymptotiquement de la forme $\lambda_\alpha(N)N^{1-\alpha}$ et nous cherchons à optimiser l'algorithme en α, N et $\lambda_\alpha(N)$. Alors nous prouvons que l'optimum est atteint pour $\alpha = 1/3$: en effet, pour $\alpha < 1/3$, $\lambda_\alpha(N) \underset{N\to\infty}{\longrightarrow} 0$, pour $\alpha = 1/3$, $\lambda_\alpha(N) \underset{N\to\infty}{\longrightarrow}$ cst et pour $\alpha > 1/3$, $\lambda_\alpha(N) \underset{N\to\infty}{\longrightarrow} \infty$.
Nous supposons donc à partir de maintenant que ρ_N a la forme asymptotique $\rho_N = \lambda N^{1-\alpha}$, où λ dénote simplement $\lambda_\alpha(N)$.

<u>Notations</u> :

Posons
$$\mathcal{F}_{2i} = \sigma\left(\widetilde{P}_1, \cdots, \widetilde{P}_i, \widetilde{P}'_1, \cdots, \widetilde{P}'_i\right) \qquad \text{et} \qquad \mathcal{F}_{2i+1} = \mathcal{F}_{2i} \vee \sigma\left(\widetilde{P}_{i+1}\right)$$
Dans la suite, nous noterons $\mathbb{E}_k$ pour $\mathbb{E}\left(\bullet|\mathcal{F}_k\right)$ et var_k pour $\mathrm{var}\left(\bullet|\mathcal{F}_k\right)$.

<u>Estimateurs</u> :

Afin d'exploiter toute l'information fournie par la simulation, nous utiliserons les résultats des deux phases pour estimer les probabilités de transition : P_{i+1} est estimé par le nombre de succès renormalisé de la génération i dans les phases 1 et 2. Plus précisément, pour $i = 0 \cdots M$, P_{i+1} est estimé par
$$\widetilde{P}''_{i+1} = \frac{\text{nb total de succès en } B_{i+1}}{\text{nb total de particules générées depuis } B_i} = \frac{\frac{\lambda}{N^\alpha} r_i^0 \widetilde{p}_{i+1} + (1 - \frac{\lambda}{N^\alpha})\widetilde{r}_i \widetilde{p}'_{i+1}}{\frac{\lambda}{N^\alpha} r_i^0 \widetilde{p}_i + (1 - \frac{\lambda}{N^\alpha})\widetilde{r}_i \widetilde{p}'_i}$$
avec
$$\begin{cases} \widetilde{p}_0 = 1, & \widetilde{p}_i = \widetilde{P}_1 \cdots \widetilde{P}_i \ \text{ pour } \ i = 1 \cdots M+1 \\ \widetilde{p}'_0 = 1, & \widetilde{p}'_i = \widetilde{P}'_1 \cdots \widetilde{P}'_i \ \text{ pour } \ i = 1 \cdots M+1 \\ \widetilde{r}_0 = 1, & \widetilde{r}_i = \widetilde{R}_1 \cdots \widetilde{R}_i \ \text{ pour } \ i = 1 \cdots M \end{cases}$$
généralisations des notations introduites en Section 3.2.

C'est l'estimateur de variance minimale parmi les estimateurs combinaison linéaire de $\widetilde{P}_i$ et $\widetilde{P}'_i$.

Finalement, $\mathbb{P}(A)$ est estimé par le produit de ces fractions :

$$\widetilde{P} = \widetilde{P}''_1 \cdots \widetilde{P}''_{M+1} := \widetilde{p}''_{M+1}$$

et nous obtenons

Proposition 5.1.1. — $\mathbb{P}(A)$ *est estimé par*

$$\widetilde{P} = \widetilde{p}'_{M+1} \frac{\prod_{i=1}^{M+1} \left\{1 + \frac{\lambda}{N^\alpha}\left(\frac{\widetilde{p}_i r^0_{i-1}}{\widetilde{p}'_i \widetilde{r}_{i-1}} - 1\right)\right\}}{\prod_{i=1}^{M} \left\{1 + \frac{\lambda}{N^\alpha}\left(\frac{\widetilde{p}_i r^0_{i}}{\widetilde{p}'_i \widetilde{r}_{i}} - 1\right)\right\}} \tag{77}$$

5.1.2. Outils mathématiques. —

5.1.2.1. Probabilités de troncature. — Dans cette section, nous donnons une borne précise des probabilités de troncature d'une somme de variables aléatoires de Bernoulli i.i.d., outil technique utilisé dans la suite de l'analyse (voir par exemple la preuve du Lemme 5.1.1).

Proposition 5.1.2. — *Pour* $\widehat{P}$ *de la forme*

$$\widehat{P} = \frac{1}{N}\sum_{i=1}^{N} B_i$$

où les B_i *sont des variables aléatoires i.i.d. de même loi que* B *Bernoulli de paramètre* P*, nous avons*

$$\begin{array}{rcl} \mathbb{P}(\widehat{P} \le a) & \le & \exp\{-Ng(a)\} \\ \mathbb{P}(\widehat{P} \ge b) & \le & \exp\{-Ng(b)\} \end{array}$$

avec $g(x) = \log\left[\left(\frac{1-x}{1-P}\right)^{1-x}\left(\frac{x}{P}\right)^{x}\right]$.

Démonstration. — Appliquons la méthode de Chernoff pour obtenir

$$\begin{aligned}
\mathbb{P}(\widehat{P} \leq a) &= \mathbb{P}(\frac{1}{N}\sum_{i=1}^{N} B_i \leq a) = \mathbb{P}(\sum_{i=1}^{N} B_i \leq Na) \\
&= \mathbb{P}(\exp\left\{\lambda \sum_{i=1}^{N} B_i\right\} \geq \exp\{\lambda Na\}) \quad \text{for} \quad \lambda < 0 \\
&\leq \exp\{-\lambda Na\}\, \mathbb{E}(\exp\left\{\lambda \sum_{i=1}^{N} B_i\right\}) \leq \exp\{-\lambda Na\}\, \mathbb{E}(\exp\{\lambda B\})^N
\end{aligned}$$

Mais $\mathbb{E}(\exp\{\lambda B\}) = Pe^{\lambda} + 1 - P$; soit alors f la fonction convexe définie par $f(\lambda) = \log(Pe^{\lambda} + 1 - P) - \lambda a$, son minimum est donné par

$$\lambda_{\min} = \log(\frac{a}{P}\frac{1-P}{1-a}) \leq 0.$$

Donc

$$\mathbb{P}(\widehat{P} \leq a) \leq \exp\left\{-N \log\left[\left(\frac{1-a}{1-P}\right)^{1-a}\left(\frac{a}{P}\right)^{a}\right]\right\}$$

où $\log\left[\left(\frac{1-a}{1-P}\right)^{1-a}\left(\frac{a}{P}\right)^{a}\right] \geq 0$.

En utilisant le même type d'argument mais avec $\lambda > 0$, on déduit une majoration exponentielle de $\mathbb{P}(\widehat{P} \geq b)$, qui conduit au résultat. □

5.1.2.2. Un lemme fondamental. — Dans cette section, nous établissons un résultat sur l'espérance de fonctions des estimateurs tronqués dont nous nous servirons à maintes reprises lors de l'étude analytique du modèle.

Lemme 5.1.1. — *Soient f, g deux fonctions définies sur $]0,1[$, $\mathcal{C}^2$ sur $[a,b]$ et $\widehat{P}$ une variable aléatoire de la forme*

$$\widehat{P} = \frac{1}{N}\sum_{i=1}^{N} B_i$$

où les B_i sont des variables aléatoires i.i.d. de même loi que B Bernoulli de paramètre P, Enfin, soit $\widetilde{P} = a \vee \widehat{P} \wedge b$. Alors

$$\mathbb{E}\left(f(\widetilde{P})\right) = f(P) + \frac{P(1-P)}{2N} f''(P) + o(\frac{1}{N})$$

Par conséquent,

$$\begin{aligned}
var\left(f(\widetilde{P})\right) &= \frac{P(1-P)}{N}f'(P)^2 + o(\frac{1}{N}) \\
cov\left(f(\widetilde{P}), g(\widetilde{P})\right) &= \frac{P(1-P)}{N}f'(P)g'(P) + o(\frac{1}{N})
\end{aligned}$$

Démonstration. — Tout d'abord,

$$\mathbb{E}\left(f(\widetilde{P})\right) = \mathbb{E}\left(f(\widehat{P})\mathbf{1}_{\widehat{P}\in[a,b]}\right) + \mathbb{E}\left(f(\widetilde{P})\mathbf{1}_{\widehat{P}\in[a,b]^c}\right)$$

Mais, puisque $\widetilde{P} \in [a, b]$ et f est bornée sur $[a, b]$ par une constante M_f,

$$\mathbb{E}\left(f(\widetilde{P})\mathbf{1}_{\widehat{P}\in[a,b]^c}\right) \leq M_f\mathbb{P}\left(\widehat{P} \notin [a,b]\right) \leq 2M_f \exp\left\{-Nh(a,b)\right\}$$

par la Proposition 5.1.2 et en posant $h(a, b) = \min\{g(a), g(b)\}$.

Ensuite, notons que

$$\begin{cases}
\mathbb{E}\left(\widehat{P} - P\right) = 0, \qquad \mathbb{E}\left(\left(\widehat{P} - P\right)^2\right) = \frac{P(1-P)}{N}, \\
\mathbb{E}\left(\left(\widehat{P} - P\right)^3\right) = \frac{P(1-P)}{N^2}(1-2P), \\
\mathbb{E}\left(\left(\widehat{P} - P\right)^4\right) = \frac{P(1-P)}{N^2}\left[(P^3 + (1-P)^3)\frac{1}{N} + (1-\frac{1}{N})P(1-P)\right].
\end{cases}$$

et puisque f est $\mathcal{C}^2$ sur $[a, b]$, par un développement de Taylor,

$$f(\widehat{P})\mathbf{1}_{\widehat{P}\in[a,b]} = \left[f(P) + \left(\widehat{P} - P\right)f'(P) + \frac{\left(\widehat{P} - P\right)^2}{2}f''(P) + o\left(\left(\widehat{P} - P\right)^2\right)\right]\mathbf{1}_{\widehat{P}\in[a,b]}$$

et donc

$$\begin{aligned}
\mathbb{E}\left(f(\widehat{P})\mathbf{1}_{\widehat{P}\in[a,b]}\right) = \mathbb{E}&\left(f(P) + \left(\widehat{P} - P\right)f'(P) + \frac{\left(\widehat{P} - P\right)^2}{2}f''(P) + o\left(\left(\widehat{P} - P\right)^2\right)\right) \\
&-\mathbb{E}\left(\left[f(P) + \left(\widehat{P} - P\right)f'(P) + \frac{\left(\widehat{P} - P\right)^2}{2}f''(P) + o\left(\left(\widehat{P} - P\right)^2\right)\right]\mathbf{1}_{\widehat{P}\notin[a,b]}\right)
\end{aligned}$$

Le premier terme dans le côté droit de l'équation est simplement

$$f(P) + \frac{P(1-P)}{2N} f''(P) + o(\frac{1}{N}).$$

Puisque $P \in [a,b]$ et f est bornée sur $[a,b]$, le second terme est borné par une constante fois $\mathbb{P}\left(\widehat{P} \notin [a,b]\right)$, elle-même bornée par $\exp\{-Nh(a,b)\}$, ce qui nous permet de conclure. □

Il est aisé de généraliser ces résultats aux fonctions de plusieurs variables par des formules de duplication.

5.1.3. Les résultats. — Pour déduire l'algorithme optimal, nous avons besoin de développements asymptotiques du coût de la simulation et de la variance de l'estimateur. Une analyse précise du problème d'optimisation montre qu'il est nécessaire de faire un développement asymptotique à l'ordre 2 pour la phase d'apprentissage (c-à-d à l'ordre 2 en ρ_N) et au premier ordre seulement pour la seconde phase (c-à-d à l'ordre 1 en $N - \rho_N$ c-à-d en N).

5.1.3.1. Etude du coût. — Le coût moyen est donné par

$$C = \lambda N^{1-\alpha} \sum_{i=0}^{M} R_0^i \mathbb{E}(\widetilde{p}_i) + (N - \lambda N^{1-\alpha}) \sum_{i=0}^{M} \mathbb{E}(\widetilde{r}_i \widetilde{p}'_i)$$

Remarquons que dans cette section ainsi que dans le reste du chapitre, nous avons supposé que le coût de génération de toute particule était unitaire. Ceci est conforté par l'étude de ce coût dans quelques exemples, comme notamment le processus d'Ornstein-Uhlenbeck (voir Chapitre 6 et plus précisément Section 6.1.3).

Comme dit précédemment, nous souhaitons donner une expression asymptotique du coût lorsque le budget tend vers l'infini. Remarquons que

$$\frac{C}{N} = \frac{\lambda}{N^\alpha} \sum_{i=0}^{M} R_0^i \mathbb{E}(\widetilde{p}_i) + \left(1 - \frac{\lambda}{N^\alpha}\right) \sum_{i=0}^{M} \mathbb{E}(\widetilde{r}_i \widetilde{p}'_i)$$

et puisque $P_i \in [a,b]$ pour tout $i = 1 \cdots M+1$, le côté droit est borné inférieurement et supérieurement ; par conséquent, nous pouvons vérifier que comme annoncé en Introduction,

$$N \to \infty \text{ lorsque } C \to \infty$$

et C est linéaire en N. Maintenant, en appliquant le Lemme 5.1.1 et en utilisant les notations introduites en Sections 3.2 et 5.1.1, il vient

Théorème 5.1.1. —

$$C = N\left[C_{opt} + \frac{B_1}{\lambda N^{1-\alpha}} + B_2\frac{\lambda}{N^\alpha} + o\left(\frac{1}{N^\alpha}\right) + o\left(\frac{1}{N^{1-\alpha}}\right)\right]$$

avec

$$\begin{cases} B_1 = \frac{1}{2}\sum_{i=1}^{M} r_i p_i \left[\frac{3/4}{P_1(1-P_1)} - 2 + 2\sum_{k=1}^{i-1}\frac{1-P_{k+1}}{r_k^0 p_{k+1}} + \frac{3/4-P_{i+1}}{r_i^0 p_{i+1}(1-P_{i+1})}\right] \\ B_2 = \sum_{i=1}^{M} p_i(r_i^0 - r_i) \end{cases}$$

et C_{opt} le coût par particule moyen généré dans le modèle optimal, donné par

$$\sum_{i=0}^{M} r_i p_i$$

Démonstration. — D'abord, nous avons $\mathbb{E}(\widetilde{p_i}) = p_i$. Définissons

$$\varphi(x) = \sqrt{\frac{1}{x} - 1} \quad \text{et} \quad \psi(x) = \sqrt{x(1-x)}$$

et conditionnons par $\mathcal{F}_{2i}$, alors

$$\mathbb{E}\left(\widetilde{r_i}\widetilde{p_i}'\right) \ = \ \mathbb{E}\left(\frac{\widetilde{r}_{i-1}}{\psi\widetilde{P}_i}\widetilde{p_i}\mathbb{E}_{2i}\left(\varphi(\widetilde{P}_{i+1})\right)\right)$$

Nous avons

$$\widetilde{P}_{i+1} = P_{i+1} + \frac{1}{\lambda N^{1-\alpha} r_i^0 \widetilde{p_i}} \sum_{j=1}^{\lambda N^{1-\alpha} r_i^0 \widetilde{p_i}} B_j$$

avec $B_j \underset{\mathcal{L}}{=} \mathrm{Ber}\,(P_{i+1}) - P_{i+1}$. Un application de la Proposition 5.1.1 à φ au voisinage de P_{i+1} et avec pour nombre de particules générées $\lambda N^{1-\alpha} r_i^0 \widetilde{p_i}$ conduit à

$$\begin{aligned}
\mathbb{E}\left(\widetilde{r_i}\widetilde{p_i}'\right) &= \varphi(P_{i+1})\mathbb{E}\left(\frac{\widetilde{r}_{i-1}}{\psi(\widetilde{P}_i)}\widetilde{p_i}'\right) + \frac{P_{i+1}(1-P_{i+1})}{2\lambda N^{1-\alpha} r_i^0}\varphi''(P_{i+1})\mathbb{E}\left(\frac{\widetilde{r}_{i-1}}{\psi(\widetilde{P}_i)}\frac{\widetilde{p_i}'}{\widetilde{p_i}}\right) + o\left(\frac{1}{N^{1-\alpha}}\right) \\
&= \varphi(P_{i+1})\mathbb{E}\left(\frac{\widetilde{r}_{i-1}}{\psi(\widetilde{P}_i)}\widetilde{p_i}'\right) + \frac{P_{i+1}(1-P_{i+1})}{2\lambda N^{1-\alpha} r_i^0}\varphi''(P_{i+1})\frac{r_{i-1}}{\psi(P_i)} + o\left(\frac{1}{N^{1-\alpha}}\right)
\end{aligned}$$

Conditionnons par $\mathcal{F}_{2i-1}$ et appliquons le Lemme 5.1.1 à la fonction identité Id en gardant seulement le premier terme du développement,

$$\mathbb{E}\left(\widetilde{r}_i\widetilde{p}'_i\right) = \varphi(P_{i+1})P_i\mathbb{E}\left(\frac{\widetilde{r}_{i-1}}{\psi(\widetilde{P}_i)}\widetilde{p}'_{i-1}\right) + \frac{P_{i+1}(1-P_{i+1})}{2\lambda N^{1-\alpha}r_i^0}\varphi''(P_{i+1})\frac{r_{i-1}}{\psi(P_i)} + o\left(\frac{1}{N^{1-\alpha}}\right)$$

Pour k de $i-1$ à 2, conditionnons d'abord par $\mathcal{F}_{2k}$, appliquons le Lemme 5.1.1 à $\frac{\varphi}{\psi}$ puis conditionnons par $\mathcal{F}_{2k-1}$ et appliquons le Lemme 5.1.1 à Id. Un dernier conditionnement par $\mathcal{F}_2$ et une application du Lemme 5.1.1 à $\frac{\varphi}{\psi}$ conduisent successivement à

$$\begin{aligned}
\mathbb{E}\left(\widetilde{r}_i\widetilde{p}'_i\right) &= \varphi(P_{i+1})P_i\mathbb{E}\left(\frac{\widetilde{r}_{i-2}\widetilde{p}'_{i-1}}{\psi(\widetilde{P}_{i-1})}\mathbb{E}_{2(i-1)}\left(\frac{\varphi(\widetilde{P}_i)}{\psi(\widetilde{P}_i)}\right)\right)\\
&\quad + \frac{P_{i+1}(1-P_{i+1})}{2\lambda N^{1-\alpha}r_i^0}\frac{r_{i-1}\varphi''(P_{i+1})}{\psi(P_i)} + o\left(\frac{1}{N^{1-\alpha}}\right)\\
&= R_iP_i\varphi(P_i)\mathbb{E}\left(\frac{\widetilde{r}_{i-2}}{\psi(\widetilde{P}_{i-1})}\right) + \frac{P_i(1-P_i)}{2\lambda N^{1-\alpha}r_{i-1}^0}\varphi(P_{i+1})\left(\frac{\varphi}{\psi}\right)''(P_i)\frac{r_{i-2}}{\psi(P_{i-1})}\frac{p_i}{p_{i-1}}\\
&\quad + \frac{P_{i+1}(1-P_{i+1})}{2\lambda N^{1-\alpha}r_i^0}\varphi''(P_{i+1})\frac{r_{i-1}}{p_i\psi(P_i)} + o\left(\frac{1}{N^{1-\alpha}}\right)\\
&= \cdots\\
&= \frac{r_i}{R_1}\frac{p_i}{P_1}\varphi(P_2)\mathbb{E}\left(\frac{\widetilde{P}'_1}{\psi(\widetilde{P}_1)}\right) + \frac{r_ip_i}{2\lambda N^{1-\alpha}}\sum_{k=1}^{i-1}\frac{P_{k+1}(1-P_{k+1})}{r_k^0R_kR_{k+1}}\frac{\varphi(P_{k+2})}{p_k\psi(P_k)}\left(\frac{\varphi}{\psi}\right)''(P_{k+1})\\
&\quad + \frac{P_{i+1}(1-P_{i+1})}{2\lambda N^{1-\alpha}r_i^0}\varphi''(P_{i+1})\frac{r_{i-1}}{\psi(P_i)} + o\left(\frac{1}{N^{1-\alpha}}\right)
\end{aligned}$$

Enfin une dernière application du Lemme 5.1.1 à $(x,y)\mapsto\frac{y}{\psi(x)}$ amène

$$\begin{aligned}
\mathbb{E}\left(\widetilde{r}_i\widetilde{p}_i\right) &= r_ip_i\left[1+\frac{1}{2\lambda N^{1-\alpha}}\left(\frac{3/4}{P_1(1-P_1)}-2\right)+\frac{1}{\lambda N^{1-\alpha}}\sum_{k=1}^{i-1}(1-P_{k+1})\frac{1}{r_k^0p_{k+1}}\right.\\
&\quad\left.+\frac{1}{2\lambda N^{1-\alpha}r_i^0}\frac{1}{p_i}\frac{3/4-P_{i+1}}{P_{i+1}(1-P_{i+1})}\right]+o\left(\frac{1}{N^{1-\alpha}}\right)
\end{aligned}$$

D'où le résultat. □

5.1.3.2. Etude de la variance. — En appliquant le Lemme 5.1.1 et en utilisant les notations introduites en Sections 3.2 et 5.1.1, la variance de l'estimateur est donnée par

Théorème 5.1.2. —
$$var(\widetilde{P}) = \frac{P^2}{N}\left[V_{opt} + \frac{1}{\lambda N^{1-\alpha}}A_1 + \frac{\lambda}{N^\alpha}A_2 + \frac{\lambda^2}{N^{2\alpha}}A_3 + o\left(\frac{1}{N^{1-\alpha}}\right) + o\left(\frac{1}{N^{2\alpha}}\right)\right]$$
avec
$$\begin{cases} A_1 = \frac{1}{2}\sum_{i=1}^M \left[\frac{P_{i+1}-1/4}{r_i^0 r_i p_{i+1}^2} - \frac{1}{4}\frac{r_i p_i}{P_1^2}\right] \\ A_2 = \left(\frac{1}{P_1}-1\right)\sum_{i=1}^M p_i(r_i - r_i^0) \\ A_3 = \left(\frac{1}{P_1}-1\right)\sum_{i=1}^M \frac{p_i}{r_i}(r_i^0 - r_i)^2 \end{cases}$$
et V_{opt} la variance par particule du modèle optimal, donnée par
$$\sum_{i=0}^M \left(\frac{1}{p_{i+1}} - \frac{1}{p_i}\right)\frac{1}{r_i} = \left(\frac{1}{P_1} - 1\right)C_{opt}$$
où la dernière égalité découle du Lemme 5.1.2 suivant.

Avant de prouver ce théorème, énonçons et prouvons le lemme

Lemme 5.1.2. — *En prenant les nombres de retirage optimaux, la variance du modèle de branchement avec duplication des trajectoires classique devient simplement*
$$var(\widehat{P}) = \frac{\mathbb{P}(A)^2}{N}\left(\frac{1}{P_1} - 1\right)\sum_{i=0}^M r_i p_i = \frac{\mathbb{P}(A)^2}{N^2}\left(\frac{1}{P_1} - 1\right)C_{opt}$$

Démonstration. — Tout d'abord, la première égalité provient directement de l'équation (27). Ensuite, nous avons alors
$$\left(\frac{1}{P_{i+1}} - 1\right)\frac{1}{R_i} = R_i P_i(1-P_i)$$
et par une utilisation successive de cette relation, nous obtenons
$$\begin{aligned}\left(\frac{1}{p_{i+1}} - \frac{1}{p_i}\right)\frac{1}{r_i} &= \left(\frac{1}{P_{i+1}} - 1\right)\frac{1}{R_i}\frac{1}{r_{i-1}p_i} = R_i P_i\left(\frac{1}{P_i} - 1\right)\frac{1}{R_{i-1}}\frac{1}{r_{i-2}p_{i-1}} \\ &= \cdots = \left(\frac{1}{P_1} - 1\right) r_i p_i\end{aligned}$$
Finalement nous obtenons la relation souhaitée. □

Démonstration. — Pendant la preuve, afin d'alléger la rédaction, nous utiliserons le symbole $\approx$ pour tout développement asymptotique à l'ordre $1/N^{(1+\alpha)\wedge(2-\alpha)}$.

Soient pour $k = 1..M$,

$$Z_k = \frac{\widetilde{p}_k r_k^0}{\widetilde{p}'_k \widetilde{r}_k} \qquad \text{et} \qquad \bar{Z}_k = \frac{\widetilde{p}_k r_k^0}{\widetilde{p}'_k \widetilde{r}_{k-1}},$$

et remarquons que clairement

$$Z_k = \frac{\bar{Z}_k}{\widetilde{R}_k} = \bar{Z}_k \frac{\psi(\widetilde{P}_k)}{\varphi(\widetilde{P}_{k+1})}.$$

et

$$\widetilde{P}''_k = \widetilde{P}'_k \frac{1 + \frac{\lambda}{N^\alpha}\left(Z_{k-1}\frac{\widetilde{P}_k}{\widetilde{P}'_k} - 1\right)}{1 + \frac{\lambda}{N^\alpha}\left(Z_{k-1} - 1\right)}$$

Conditionnons d'abord par $\mathcal{F}_{2M+1}$:

$$\begin{aligned}
\operatorname{var}\left(\widetilde{P}\right) &= \operatorname{var}(\widetilde{p}''_{M+1}) \\
&= \operatorname{var}\left(\frac{\widetilde{p}''_M}{\widetilde{p}'_M \widetilde{r}_M \left\{1 + \frac{\lambda}{N^\alpha}\left(Z_M - 1\right)\right\}}\left\{\frac{\lambda}{N^\alpha}\widetilde{p}_{M+1} r_M^0 + \left(1 - \frac{\lambda}{N^\alpha}\right)\widetilde{p}'_M \widetilde{r}_M \mathbb{E}_{2M+1}(\widetilde{P}'_{M+1})\right\}\right) \\
&\quad + \mathbb{E}\left(\left[\frac{\widetilde{p}''_M}{\widetilde{p}'_M \widetilde{r}_M \left\{1 + \frac{\lambda}{N^\alpha}\left(Z_M - 1\right)\right\}}\right]^2 \left\{\left(1 - \frac{\lambda}{N^\alpha}\right)\widetilde{p}'_M \widetilde{r}_M\right\}^2 \operatorname{var}_{2M+1}\left(\widetilde{P}'_{M+1}\right)\right)
\end{aligned}$$

Mais par le Lemme 5.1.1,

$$\begin{cases}
\mathbb{E}_{2M+1}(\widetilde{P}'_{M+1}) = P_{M+1} + o\left(\frac{1}{N}\right) \\[2ex]
\operatorname{var}_{2M+1}\left(\widetilde{P}'_{M+1}\right) = \frac{P_{M+1}(1-P_{M+1})}{N\left(1-\frac{\lambda}{N^\alpha}\right)\widetilde{p}'_M \widetilde{r}_M} + o\left(\frac{1}{N}\right)
\end{cases}$$

$$\begin{aligned}
\mathrm{var}\quad \left(\widetilde{P}\right) &= P_{M+1}^2 \mathrm{var}\left(\widetilde{p}''_M \left\{1+\frac{\lambda}{N^\alpha}\left(Z_M-1\right)\right\}^{-1}\left\{1+\frac{\lambda}{N^\alpha}\left(Z_M\frac{\widetilde{P}_{M+1}}{P_{M+1}}-1\right)\right\}\right) \\
&+\frac{P_{M+1}(1-P_{M+1})}{N}\left(1-\frac{\lambda}{N^\alpha}\right)\mathbb{E}\left(\frac{\widetilde{p}''^2_M}{\widetilde{p}'_M\widetilde{r}_M}\left\{1+\frac{\lambda}{N^\alpha}\left(Z_M-1\right)\right\}^{-2}\right) \\
&= P_{M+1}^2 \mathrm{var}\left(\widetilde{p}''_M\left\{1+\frac{\lambda}{N^\alpha}Z_M\left(\frac{\widetilde{P}_{M+1}}{\widetilde{P}_{M+1}}-1\right)\left[1-\frac{\lambda}{N^\alpha}\left(Z_M-1\right)\right]\right\}\right) \\
&+\frac{P_{M+1}(1-P_{M+1})}{N}\left(1-\frac{\lambda}{N^\alpha}\right)\mathbb{E}\left(\frac{\widetilde{p}''^2_M}{\widetilde{p}'_M\widetilde{r}_M}\left\{1-2\frac{\lambda}{N^\alpha}\left(Z_M-1\right)+3\frac{\lambda^2}{N^{2\alpha}}\left(Z_M-1\right)^2\right\}\right) \\
&= P_{M+1}^2 \mathrm{var}\left(\widetilde{p}''_M\left\{1+\frac{\lambda}{N^\alpha}\bar{Z}_M\psi(\widetilde{P}_M)\mathbb{E}_{2M}\left(f(\widetilde{P}_{M+1})\right)\right\}\right) \\
&+P_{M+1}^2\mathbb{E}\left(\widetilde{p}''^2_M\frac{\lambda^2}{N^{2\alpha}}\left[\bar{Z}_M\psi(\widetilde{P}_M)\right]^2\mathrm{var}_{2M}\left(f(\widetilde{P}_{M+1})\right)\right) \\
&+\frac{P_{M+1}(1-P_{M+1})}{N}\left(1-\frac{\lambda}{N^\alpha}\right)\mathbb{E}\left(\frac{\widetilde{p}''^2_M}{\widetilde{p}'_M\widetilde{r}_{M-1}}\psi(\widetilde{P}_M)\mathbb{E}_{2M}\left(g(\widetilde{P}_{M+1})\right)\right) \\
&:= T_1+T_2+T_3
\end{aligned}$$

où la dernière égalité est le résultat d'un conditionnement par $\mathcal{F}_{2M}$ avec

$$f(x)=\frac{1}{\varphi(x)}\left(\frac{x}{P_{M+1}}-1\right)\left\{1-\frac{\lambda}{N^\alpha}\left(\bar{Z}_M\frac{\psi(\widetilde{P}_M)}{\varphi(x)}-1\right)\right\}$$

et

$$g(x)=\frac{1}{\varphi(x)}\left\{1-2\frac{\lambda}{N^\alpha}\left(\bar{Z}_M\frac{\psi(\widetilde{P}_M)}{\varphi(x)}-1\right)+3\frac{\lambda^2}{N^{2\alpha}}\left(\bar{Z}_M\frac{\psi(\widetilde{P}_M)}{\varphi(x)}-1\right)^2\right\}$$

1. <u>Calcul de T_1</u>

Appliquons le Lemme 5.1.1 à f, il vient

$$\begin{aligned}
\mathbb{E}_{2M}\left(f(\widetilde{P}_{M+1})\right) &= f(P_{M+1}) + \frac{P_{M+1}(1-P_{M+1})}{2\lambda N^{1-\alpha}\widetilde{p}_M r_M^0} f''(P_{M+1}) + o\left(\frac{1}{N^{1-\alpha}}\right) \\
&= \frac{1}{2\lambda N^{1-\alpha}\widetilde{p}_M r_M^0}\frac{1}{P_{M+1}\varphi(P_{M+1})}\left\{1-\frac{\lambda}{N^\alpha}\left(2\bar{Z}_M\frac{\psi(\widetilde{P}_M)}{\varphi(P_{M+1})}-1\right)\right\} \\
&\quad + o\left(\frac{1}{N^{1-\alpha}}\right)
\end{aligned}$$

Alors

$$\begin{aligned}
T_1 &\approx P_{M+1}^2 \mathrm{var}\left(\widetilde{p}''_M\left\{1+\frac{1}{N}\frac{\psi(\widetilde{P}_M)\varphi(P_{M+1})^{-1}}{\widetilde{p}'_M\widetilde{r}_{M-1}P_{M+1}}\left\{1-\frac{\lambda}{N^\alpha}\left(2\bar{Z}_M\frac{\psi(\widetilde{P}_M)}{\varphi(P_{M+1})}-1\right)\right\}\right\}\right) \\
&\approx P_{M+1}^2\mathrm{var}\left(\widetilde{p}''_M\right) \\
&\quad + \frac{1}{4N^2\varphi(P_{M+1})^2}\mathrm{var}\left(\frac{\widetilde{p}''_M\psi(\widetilde{P}_M)}{\widetilde{p}'_M\widetilde{r}_{M-1}}\left\{1-\frac{\lambda}{N^\alpha}\left(2\bar{Z}_M\frac{\psi(\widetilde{P}_M)}{\varphi(P_{M+1})}-1\right)\right\}\right) \\
&\quad + \frac{P_{M+1}}{N\varphi(P_{M+1})}\mathrm{cov}\left(\widetilde{p}''_M, \frac{\widetilde{p}''_M\psi(\widetilde{P}_M)}{\widetilde{p}'_M\widetilde{r}_{M-1}}\left\{1-\frac{\lambda}{N^\alpha}\left(2\bar{Z}_M\frac{\psi(\widetilde{P}_M)}{\varphi(P_{M+1})}-1\right)\right\}\right)
\end{aligned}$$

Le second terme étant négligeable (puisqu'en $\frac{1}{N^2}$), nous étudions seulement le terme de covariance et notons

$$\begin{aligned}
T_{13} &:= \mathrm{cov}\left(\widetilde{p}''_M, \widetilde{p}''_M\frac{\psi(\widetilde{P}_M)}{\widetilde{p}'_M\widetilde{r}_{M-1}}\left\{1-\frac{\lambda}{N^\alpha}\left(2\bar{Z}_M\frac{\psi(\widetilde{P}_M)}{\varphi(P_{M+1})}-1\right)\right\}\right) \\
&= \mathbb{E}\left(\widetilde{p}''^2_M\frac{\psi(\widetilde{P}_M)}{\widetilde{p}'_M\widetilde{r}_{M-1}}\left\{1-\frac{\lambda}{N^\alpha}\left(2\bar{Z}_M\frac{\psi(\widetilde{P}_M)}{\varphi(P_{M+1})}-1\right)\right\}\right) \\
&\quad -\mathbb{E}\left(\widetilde{p}''_M\right)\mathbb{E}\left(\widetilde{p}''_M\frac{\psi(\widetilde{P}_M)}{\widetilde{p}'_M\widetilde{r}_{M-1}}\left\{1-\frac{\lambda}{N^\alpha}\left(2\bar{Z}_M\frac{\psi(\widetilde{P}_M)}{\varphi(P_{M+1})}-1\right)\right\}\right) \\
&:= T_{131} - T_{132}T_{133}
\end{aligned}$$

La procédure d'intégration étant la même pour toutes les espérances apparaissant dans cette preuve et afin d'alléger la rédaction, nous détaillerons les calculs seulement pour le terme T_{132}. En conditionnant par $\mathcal{F}_{2M-1}$, nous avons

$$\begin{aligned}
T_{132} = \mathbb{E}\left(\widetilde{p}''_M\right) &= \mathbb{E}\left(\widetilde{p}''_{M-1}\widetilde{P}'_M\left\{\frac{1+\frac{\lambda}{N^\alpha}\left(Z_{M-1}\frac{\widetilde{P}_M}{\widetilde{P}'_M}-1\right)}{1+\frac{\lambda}{N^\alpha}\left(Z_{M-1}-1\right)}\right\}\right)\\
&= \mathbb{E}\left(\widetilde{p}''_{M-1}\left\{1+\frac{\lambda}{N^\alpha}\left(Z_{M-1}-1\right)\right\}^{-1}\mathbb{E}_{2M-1}\left(u_M(\widetilde{P}'_M)\right)\right)
\end{aligned}$$

avec

$$u_M(x) = x\left\{1+\frac{\lambda}{N^\alpha}\left(Z_{M-1}\frac{\widetilde{P}_M}{x}-1\right)\right\}$$

Mais par le Lemme 5.1.1, le second terme étant négligeable,

$$\begin{aligned}
\mathbb{E}_{2M-1}\left(u_M(\widetilde{P}'_M)\right) &= u_M(P_M)+\frac{P_M(1-P_M)}{2N\left(1-\frac{\lambda}{N^\alpha}\right)}u''_M(P_M)+o\left(\frac{1}{N}\right)\\
&= P_M\left\{1+\frac{\lambda}{N^\alpha}\left(Z_{M-1}\frac{\widetilde{P}_M}{P_M}-1\right)\right\}+o\left(\frac{1}{N^\alpha}\right)
\end{aligned}$$

Soit

$$\begin{aligned}
v_M(x) = &\left\{1-\frac{\lambda}{N^\alpha}\left(\bar{Z}_{M-1}\frac{\psi(\widetilde{P}_{M-1})}{\varphi(x)}-1\right)+\frac{\lambda^2}{2N^{2\alpha}}\left(\bar{Z}_{M-1}\frac{\psi(\widetilde{P}_{M-1})}{\varphi(x)}-1\right)^2\right\}\\
&\left\{1+\frac{\lambda}{N^\alpha}\left(\bar{Z}_{M-1}\frac{\psi(\widetilde{P}_{M-1})}{\varphi(x)}\frac{x}{P_M}-1\right)\right\}
\end{aligned}$$

Appliquons le Lemme 5.1.1,

$$\begin{aligned}
\mathbb{E}_{2(M-1)}\left(v_M(\widetilde{P}_M)\right) &= v_M(P_M)+\frac{P_M(1-P_M)}{2\lambda N^{1-\alpha}}v''_M(P_M)+o\left(\frac{1}{N^{1-\alpha}}\right)\\
&= v_M(P_M)+o\left(\frac{1}{N}\right) \quad \text{puisque } v''_M \text{ se comporte comme } \frac{\lambda}{N^\alpha}\\
&= 1+o\left(\frac{1}{N}\right) \quad \text{en développant le produit}
\end{aligned}$$

Alors en conditionnant par $\mathcal{F}_{2(M-1)}$ et en appliquant l'égalité précédente,

$$T_{132} \approx P_M\mathbb{E}\left(\widetilde{p}''_{M-1}\mathbb{E}_{2(M-1)}\left(v_M(\widetilde{P}_M)\right)\right) \approx P_M\mathbb{E}\left(\widetilde{p}''_{M-1}\right) \approx \cdots \approx p_M$$

Par des conditionnements et des applications du Lemme 5.1.1 successifs, il vient

$$\begin{aligned}
T_{131} &\approx \frac{P_M \cdots P_2}{R_{M-1}\cdots R_2}\frac{\psi(P_M)}{\varphi(P_2)}\mathbb{E}\left(\widetilde{P}_1^{\prime\prime 2}\frac{\psi(\widetilde{P}_1)}{\widetilde{P}_1^{\prime}}\left\{1-\frac{\lambda}{N^\alpha}\left(2\frac{\widetilde{P}_1 r_M^0}{\widetilde{P}_1^{\prime}}\frac{\psi(\widetilde{P}_1)}{\varphi(P_2)}-1\right)\right\}\right)\\
&\approx \frac{P_M \cdots P_2}{R_{M-1}\cdots R_2}\frac{\psi(P_M)}{\varphi(P_2)}\mathbb{E}\left(\Phi\left(\widetilde{P}_1,\widetilde{P}_1^{\prime}\right)\right)
\end{aligned}$$

où Φ est défini par

$$\Phi(x,y) = y\psi(x)\left\{1+2\frac{\lambda}{N^\alpha}\left(\frac{x}{y}-1\right)+\frac{\lambda^2}{N^{2\alpha}}\left(\frac{x}{y}-1\right)^2\right\}\left\{1-\frac{\lambda}{N^\alpha}\left(\frac{2r_M^0}{\varphi(P_2)}\frac{x\psi(x)}{y}-1\right)\right\}$$

Mais par le Lemme 5.1.1 appliqué à Φ,

$$\begin{aligned}
\mathbb{E}\left(\Phi\left(\widetilde{P}_1,\widetilde{P}_1^{\prime}\right)\right) &= \Phi(P_1,P_1)+P_1(1-P_1)\left\{\frac{\Phi_x^{\prime\prime}(P_1,P_1)}{2\lambda N^{1-\alpha}}+\frac{\Phi_y^{\prime\prime}(P_1,P_1)}{2N(1-\lambda/N^\alpha)}\right\}+o\left(\frac{1}{N^{1-\alpha}}\right)\\
&= P_1\psi(P_1)\left\{1-\frac{\lambda}{N^\alpha}\left(2\frac{r_M^0}{r_M}-1\right)\right\}+P_1(1-P_1)\frac{\Phi_x^{\prime\prime}(P_1,P_1)}{2\lambda N^{1-\alpha}}+o\left(\frac{1}{N^{1-\alpha}}\right)
\end{aligned}$$

Finalement,

$$T_{131} = \frac{p_M\psi(P_M)}{r_{M-1}}\left[\left\{1-\frac{\lambda}{N^\alpha}\left(2\frac{r_M^0}{r_M}-1\right)\right\}-\frac{1/8}{\lambda N^{1-\alpha}}\frac{1}{P_1(1-P_1)}\right]+o\left(\frac{1}{N^{1-\alpha}}\right)$$

De la même façon, nous avons

$$T_{133} = \frac{\psi(P_M)}{r_{M-1}}\left[\left\{1-\frac{\lambda}{N^\alpha}\left(2\frac{r_M^0}{r_M}-1\right)\right\}-\frac{1/8}{\lambda N^{1-\alpha}}\frac{1}{P_1(1-P_1)}\right]+o\left(\frac{1}{N^{1-\alpha}}\right)$$

de quoi on déduit la nullité du terme de covariance. Finalement,

$$T_1 \approx P_{M+1}^2 \mathrm{var}\left(\widetilde{p}_M^{\prime\prime}\right)$$

2. <u>Calcul de T_2</u>

Par des conditionnements et des applications du Lemme 5.1.1 successifs, il vient

$$
\begin{aligned}
T_2 &\approx \frac{\lambda}{N^{1+\alpha}}(1-P_{M+1})P_{M+1}\cdots P_2\frac{r_M^0}{(R_M\cdots R_2)^2\varphi(P_2)^2}\\
&\quad\mathbb{E}\left(\widetilde{P}_1^{\prime\prime 2}\frac{\widetilde{P}_1\psi(\widetilde{P}_1)^2}{\widetilde{P}_1^{\prime 2}}\left\{1-2\frac{\lambda}{N^\alpha}\left(\frac{r_M^0}{R_M\cdots R_2}\frac{\psi(\widetilde{P}_1)}{\varphi(P_2)}\frac{\widetilde{P}_1}{\widetilde{P}'_1}-1\right)\right\}\right)\\
&\approx \frac{\lambda}{N^{1+\alpha}}(1-P_{M+1})P_{M+1}\cdots P_2\frac{r_M^0}{(R_M\cdots R_2)^2\varphi(P_2)^2}\mathbb{E}\left(\Psi\left(\widetilde{P}_1,\widetilde{P}'_1\right)\right)
\end{aligned}
$$

en posant

$$
\Psi(x,y)=x\psi(x)^2\left\{1+\frac{\lambda}{N^\alpha}\left(\frac{x}{y}-1\right)\right\}\left\{1-2\frac{\lambda}{N^\alpha}\left(\frac{r_M^0}{R_M\cdots R_2}\frac{\psi(x)}{\varphi(P_2)}\frac{x}{y}-1\right)\right\}
$$

Appliquons le Lemme 5.1.1 à Ψ et en ne conservant que le premier terme,

$$
T_2=\frac{\lambda}{N^{1+\alpha}}(1-P_{M+1})p_{M+1}\frac{r_M^0}{r_M^2}\left\{1-2\frac{\lambda}{N^\alpha}\left(\frac{r_M^0}{r_M}-1\right)\right\}+o\left(\frac{1}{N^{1+\alpha}}\right)
$$

3. <u>Calcul de T_3</u>

Appliquons le Lemme 5.1.1 à g définie précédemment,

$$
\begin{aligned}
T_3 &\approx \frac{1}{2\lambda N^{2-\alpha}}\frac{P_{M+1}-1/4}{r_M r_M^0}+\frac{P_{M+1}(1-P_{M+1})}{N}\left(1-\frac{\lambda}{N^\alpha}\right)\mathbb{E}\left(\frac{\widetilde{p}_M^{\prime\prime 2}}{\widetilde{p}'_M\widetilde{r}_{M-1}}\frac{\psi(\widetilde{P}_M)}{\varphi(P_{M+1})}\right.\\
&\quad\left.\left\{1-2\frac{\lambda}{N^\alpha}\left(\frac{r_M^0}{R_M\cdots R_2}\frac{\psi(\widetilde{P}_1)}{\varphi(P_2)}\frac{\widetilde{P}_1}{\widetilde{P}'_1}-1\right)+3\frac{\lambda^2}{N^{2\alpha}}\left(\frac{r_M^0}{R_M\cdots R_2}\frac{\psi(\widetilde{P}_1)}{\varphi(P_2)}\frac{\widetilde{P}_1}{\widetilde{P}'_1}-1\right)^2\right\}\right)\\
&:= T_{31}+T_{32}
\end{aligned}
$$

Par des conditionnements et des applications du Lemme 5.1.1 successifs, il vient

$$
\begin{aligned}
T_{32} &\approx \frac{1-P_{M+1}}{N}\left(1-\frac{\lambda}{N^\alpha}\right)\frac{P_{M+1}\cdots P_2}{R_M\cdots R_2\varphi(P_2)}\mathbb{E}\left(\widetilde{P}_1''^2\frac{\psi(\widetilde{P}_1)}{\widetilde{P}_1'}\left\{1-2\frac{\lambda}{N^\alpha}\left(\frac{r_M^0}{R_M\cdots R_2}\frac{\psi(\widetilde{P}_1)}{\varphi(P_2)}\frac{\widetilde{P}_1}{\widetilde{P}_1'}-1\right)\right.\right.\\
&\quad\left.\left.+3\frac{\lambda^2}{N^{2\alpha}}\left(\frac{r_M^0}{R_M\cdots R_2}\frac{\psi(\widetilde{P}_1)}{\varphi(P_2)}\frac{\widetilde{P}_1}{\widetilde{P}_1'}-1\right)^2\right\}\right)\\
&\approx \frac{1-P_{M+1}}{N}\left(1-\frac{\lambda}{N^\alpha}\right)\frac{P_{M+1}\cdots P_2}{R_M\cdots R_2\varphi(P_2)}\mathbb{E}\left(\Gamma\left(\widetilde{P}_1,\widetilde{P}_1'\right)\right)
\end{aligned}
$$

en posant

$$
\begin{aligned}
\Gamma(x,y) = y\psi(x)&\left\{1+2\frac{\lambda}{N^\alpha}\left(\frac{x}{y}-1\right)+\frac{\lambda^2}{N^{2\alpha}}\left(\frac{x}{y}-1\right)^2\right\}\\
&\left\{1-2\frac{\lambda}{N^\alpha}\left(\frac{r_M^0}{R_M\cdots R_2}\frac{\psi(x)}{\varphi(P_2)}\frac{x}{y}-1\right)+3\frac{\lambda^2}{N^{2\alpha}}\left(\frac{r_M^0}{R_M\cdots R_2}\frac{\psi(x)}{\varphi(P_2)}\frac{x}{y}-1\right)^2\right\}
\end{aligned}
$$

Appliquons le Lemme 5.1.1 à Γ et en ne conservant que les deux premiers termes,

$$
\begin{aligned}
T_{32} &\approx \frac{p_{M+1}(1-P_{M+1})}{Nr_M}\left(1-\frac{\lambda}{N^\alpha}\right)\left\{1-2\frac{\lambda}{N^\alpha}\left(\frac{r_M^0}{r_M}-1\right)+3\frac{\lambda^2}{N^{2\alpha}}\left(\frac{r_M^0}{r_M}-1\right)^2\right\}\\
&\quad-\frac{1-P_{M+1}}{8\lambda N^{2-\alpha}}\frac{p_{M+1}}{r_M}\frac{1}{P_1(1-P_1)}
\end{aligned}
$$

Finalement,

$$
\begin{aligned}
T_3 &\approx \frac{p_{M+1}(1-P_{M+1})}{Nr_M}\left(1-\frac{\lambda}{N^\alpha}\right)\left\{1-2\frac{\lambda}{N^\alpha}\left(\frac{r_M^0}{r_M}-1\right)+3\frac{\lambda^2}{N^{2\alpha}}\left(\frac{r_M^0}{r_M}-1\right)^2\right\}\\
&\quad+\frac{1}{2\lambda N^{2-\alpha}}\frac{1}{r_M}\left[\frac{P_{M+1}-1/4}{r_M^0}-\frac{p_{M+1}}{4}\frac{1-P_{M+1}}{P_1(1-P_1)}\right]
\end{aligned}
$$

alors

$$
\begin{aligned}
\operatorname{var}\left(\widetilde{P}\right) \approx\ & P_{M+1}^2 \operatorname{var}\left(\widetilde{p}_M''\right) + \frac{\lambda}{N^{1+\alpha}}(1-P_{M+1})p_{M+1}\frac{r_M^0}{r_M^2}\left\{1-2\frac{\lambda}{N^\alpha}\left(\frac{r_M^0}{r_M}-1\right)\right\} \\
& + \frac{p_{M+1}(1-P_{M+1})}{Nr_M}\left(1-\frac{\lambda}{N^\alpha}\right)\left\{1-2\frac{\lambda}{N^\alpha}\left(\frac{r_M^0}{r_M}-1\right)+3\frac{\lambda^2}{N^{2\alpha}}\left(\frac{r_M^0}{r_M}-1\right)^2\right\} \\
& + \frac{1}{2\lambda N^{2-\alpha}}\frac{1}{r_M}\left[\frac{P_{M+1}-1/4}{r_M^0}-\frac{p_{M+1}}{4}\frac{1-P_{M+1}}{P_1(1-P_1)}\right]
\end{aligned}
$$

Mais en appliquant le Lemme 5.1.1,

$$
\begin{aligned}
\operatorname{var}\left(\widetilde{P}_1''\right) &= \operatorname{var}\left(\widetilde{P}_1'\left\{1+\frac{\lambda}{N^\alpha}\left(\frac{\widetilde{P}_1}{\widetilde{P}_1'}-1\right)\right\}\right)+o\left(\frac{1}{N^{1-\alpha}}\right) \\
&= P_1(1-P_1)\left[\frac{\Delta_x'^2(P_1,P_1)}{\lambda N^{1-\alpha}}+\frac{\Delta_y'^2(P_1,P_1)}{N\left(1-\frac{\lambda}{N^\alpha}\right)}\right]+o\left(\frac{1}{N^{1-\alpha}}\right) \\
&= P_1(1-P_1)\left[\frac{1}{\lambda N^{1-\alpha}}\frac{\lambda^2}{N^{2\alpha}}+\frac{\left(1-\frac{\lambda}{N^\alpha}\right)^2}{N\left(1-\frac{\lambda}{N^\alpha}\right)}\right]+o\left(\frac{1}{N^{1-\alpha}}\right) \\
&= \frac{P_1(1-P_1)}{N}+o\left(\frac{1}{N^{1-\alpha}}\right)
\end{aligned}
$$

avec

$$
\Delta(x,y) = y\left(1+\frac{\lambda}{N^\alpha}\left(\frac{x}{y}-1\right)\right)
$$

et l'on conclut par récurrence. □

5.1.3.3. Optimisation de l'algorithme en les paramètres. — Les équivalents de la variance et du coût sont donnés par

$$
\begin{cases}
\operatorname{var}\left(\widetilde{P}\right) \underset{N\to\infty}{\sim} \frac{V_{opt}}{N}+\frac{1}{\lambda N^{2-\alpha}}A_1+\frac{\lambda}{N^{1+\alpha}}A_2+\frac{\lambda^2}{N^{1+2\alpha}}A_3 \\
C \underset{N\to\infty}{\sim} NC_{opt}+\lambda N^{1-\alpha}B_2+\frac{N^\alpha}{\lambda}B_1
\end{cases}
$$

Rappelons que nous souhaitons optimiser l'algorithme en les paramètres α, N et λ. Nous prouvons d'abord le lemme suivant :

Lemme 5.1.3. — *Soient F et G deux fonctions à valeurs réelles. Supposons que*

$$\begin{aligned} F(\lambda, N) &\underset{N\to\infty}{\sim} \frac{1}{N} + \frac{f_1(\lambda)}{N^{1+\alpha}} + \frac{f_2(\lambda)}{N^{1+2\alpha}} + \frac{f_3(\lambda)}{N^{2-\alpha}}, \\ G(\lambda, N) &\underset{N\to\infty}{\sim} N + N^{1-\alpha} g_1(\lambda) + N^{\alpha} g_2(\lambda). \end{aligned}$$

et

$$f_1 + g_1 = 0. \tag{78}$$

Alors minimiser F à $G = G_f$ fixé revient asymptotiquement à minimiser la fonction de λ

- $f_3 + g_2$ *pour* $\alpha > \frac{1}{3}$,
- $f_2 + \alpha f_1 g_1$ *pour* $\alpha < \frac{1}{3}$,
- $f_2 + f_3 + g_2 + f_1 g_1$ *pour* $\alpha = \frac{1}{3}$,

N étant donné par l'équation $G = G_f$.

Démonstration. — **Cas 1 - $\alpha > \frac{1}{3}$.** D'abord, N satisfait

$$N + N^{1-\alpha} g_1(\lambda) + N^{\alpha} g_2(\lambda) = G_f \tag{79}$$

et nous pouvons considérer que

$$N \sim G_f + \mu G_f^{1-\alpha} + \nu G_f^{\alpha} \sim G_f \left[1 + \frac{\mu}{G_f^{\alpha}} + \frac{\nu}{G_f^{1-\alpha}}\right]$$

En injectant cette valeur dans (79), on obtient

$$\begin{aligned} G_f &\sim G_f + \mu G_f^{1-\alpha} + \nu G_f^{\alpha} + g_1(\lambda) G_f^{1-\alpha} \left[1 + \frac{\mu}{G_f^{\alpha}} + \frac{\nu}{G_f^{1-\alpha}}\right]^{1-\alpha} + g_2(\lambda) G_f^{\alpha} \left[1 + \frac{\mu}{G_f^{\alpha}} + \frac{\nu}{G_f^{1-\alpha}}\right]^{\alpha} \\ &\sim G_f + \mu G_f^{1-\alpha} + \nu G_f^{\alpha} + g_1(\lambda) G_f^{1-\alpha} + g_2(\lambda) G_f^{\alpha} \\ &\sim G_f + [\mu + g_1(\lambda)]\, G_f^{1-\alpha} + [\nu + g_2(\lambda)]\, G_f^{\alpha} \end{aligned}$$

et par conséquent $\begin{cases} \mu = -g_1(\lambda) \\ \nu = -g_2(\lambda) \end{cases}$. Puisque $\alpha > \frac{1}{3}$, $\frac{1}{N^{1+2\alpha}} = o\left(\frac{1}{N^{2-\alpha}}\right)$ et donc

$$\begin{aligned} F(\lambda, N) &\sim \frac{1}{G_f}\left[1 - \frac{\mu}{G_f^{\alpha}} - \frac{\nu}{G_f^{1-\alpha}}\right] + \frac{f_1(\lambda)}{G_f^{1+\alpha}} + \frac{f_3(\lambda)}{G_f^{2-\alpha}} \\ &\sim \frac{1}{G_f} + \frac{1}{G_f^{1+\alpha}}\left[f_1(\lambda) - \mu\right] + \frac{1}{G_f^{2-\alpha}}\left[f_3(\lambda) - \nu\right] \\ &\sim \frac{1}{G_f} + \frac{1}{G_f^{2-\alpha}}\left[f_3(\lambda) + g_2(\lambda)\right] \end{aligned}$$

Et nous obtenons finalement le résultat pour $\alpha > \frac{1}{3}$.

Cas 2 - $\alpha < \frac{1}{3}$**.** D'abord, N satisfait

(80) $$N + N^{1-\alpha} g_1(\lambda) = G_f$$

et nous pouvons considérer que

$$N \sim G_f + \mu G_f^{1-\alpha} \sim G_f\left[1 + \frac{\mu}{G_f^{\alpha}}\right]$$

En injectant cette valeur dans (80),

$$G_f \sim G_f + \mu G_f^{1-\alpha} + g_1(\lambda) G_f^{1-\alpha}\left[1 + \frac{\mu}{G_f^{\alpha}}\right]^{1-\alpha} \sim G_f + \left[\mu + g_1(\lambda)\right] G_f^{1-\alpha}$$

et par conséquent $\mu = -g_1(\lambda)$. Puisque $\alpha < \frac{1}{3}$, $\frac{1}{N^{2-\alpha}} = o\left(\frac{1}{N^{1+2\alpha}}\right)$ et donc

$$\begin{aligned} F(\lambda, N) &\sim \frac{1}{G_f}\left[1 - \frac{\mu}{G_f^{\alpha}} + \mu^2 \frac{1}{G_f^{1-\alpha}}\right] + \frac{f_1(\lambda)}{G_f^{1+\alpha}}\left[1 - (1+\alpha)\frac{\mu}{G_f^{\alpha}}\right] + \frac{f_2(\lambda)}{G_f^{1+2\alpha}} \\ &\sim \frac{1}{G_f} + \frac{1}{G_f^{1+\alpha}}\left[f_1(\lambda) - \mu\right] + \frac{1}{G_f^{1+2\alpha}}\left[f_2(\lambda) - (1+\alpha)\mu f_1(\lambda) + \mu^2\right] \\ &\sim \frac{1}{G_f} + \frac{1}{G_f^{1+2\alpha}}\left[f_2(\lambda) + \alpha f_1(\lambda) g_1(\lambda)\right] \end{aligned}$$

Et nous obtenons finalement le résultat pour $\alpha < \frac{1}{3}$.

Cas 3 - $\alpha = \frac{1}{3}$**.** D'abord, N satisfait

(81) $$N + N^{2/3} g_1(\lambda) + N^{1/3} g_2(\lambda) = G_f$$

et nous pouvons considérer que

$$N \sim G_f + \mu G_f^{2/3} + \nu G_f^{1/3} = G_f \left[1 + \frac{\mu}{G_f^{1/3}} + \frac{\nu}{G_f^{2/3}}\right]$$

En injectant cette valeur dans (81),

$$G_f \sim G_f + [\mu + g_1(\lambda)]\, G_f^{2/3} + \left[\nu + g_2(\lambda) + \frac{2}{3} g_1(\lambda)\mu\right] G_f^{1/3}$$

et par conséquent $\begin{cases} \mu = -g_1(\lambda) \\ \nu = \frac{2}{3} g_1(\lambda)^2 - g_2(\lambda) \end{cases}$. D'où avec $\hat{f}_2 = f_2 + f_3$,

$$\begin{aligned} F(\lambda, N) &\sim \frac{1}{G_f}\left[1 - \frac{\mu}{G_f^{1/3}} - \frac{\nu}{G_f^{2/3}} + \mu^2 \frac{1}{G_f^{2/3}}\right] + \frac{f_1(\lambda)}{G_f^{4/3}}\left[1 - \frac{4}{3}\frac{\mu}{G_f^{1/3}}\right] + \frac{\hat{f}_2(\lambda)}{G_f^{5/3}} \\ &\sim \frac{1}{G_f} + \frac{1}{G_f^{4/3}}[f_1(\lambda) - \mu] + \frac{1}{G_f^{5/3}}\left[\hat{f}_2(\lambda) - \frac{4}{3}\mu f_1(\lambda) - \nu + \mu^2\right] \\ &\sim \frac{1}{G_f} + \frac{1}{G_f^{5/3}}\left[\hat{f}_2(\lambda) + g_2(\lambda) + f_1(\lambda) g_1(\lambda)\right] \end{aligned}$$

Et nous obtenons finalement le résultat pour $\alpha = \frac{1}{3}$.

□

Par conséquent, nous pouvons déduire la proposition suivante :

Proposition 5.1.3. — *Minimiser la variance pour un coût donné conduit à prendre $\alpha = \frac{1}{3}$ et à la solution*

$$\lambda_s = \left[\frac{C_{opt}\left(A_1 + B_1\left(\frac{1}{P_1} - 1\right)\right)}{2\left(A_2 B_2 + A_3 C_{opt}\right)}\right]^{\frac{1}{3}},$$

N étant donné par le coût.

Démonstration. — En renormalisant par V_{opt} et C_{opt} et en utilisant le lemme précédent : si

$$A_2 C_{opt} + B_2 V_{opt} = 0 \tag{82}$$

est vérifié, minimiser la variance à coût C fixé revient à minimiser les fonctions

- pour $\alpha > \frac{1}{3}$, $\frac{1}{\lambda}\left[\frac{A_1}{V_{opt}} + \frac{B_1}{\lambda C_{opt}}\right]$ qui conduit à $\lambda_s = +\infty$ qui est absurde.
- for $\alpha < \frac{1}{3}$, $\lambda^2\left[\frac{A_3}{V_{opt}} + \alpha \frac{A_2 B_2}{V_{opt} C_{opt}}\right]$ qui conduit à $\lambda_s = 0$ qui est absurde.

– for $\alpha = \frac{1}{3}$, $\frac{\frac{A_1}{\lambda}+A_3\lambda^2}{V_{opt}}+\frac{B_1}{\lambda C_{opt}}+\lambda^2\frac{A_2B_2}{V_{opt}C_{opt}}$ qui conduit à $\lambda_s = \left[\frac{C_{opt}\left(A_1+B_1\left(\frac{1}{P_1}-1\right)\right)}{2(A_2B_2+A_3C_{opt})}\right]^{\frac{1}{3}}$.

N est donné par le coût. Mais par le Lemme 5.1.2 et les expressions des nombres optimaux de retirage, on peut facilement vérifier que (82) est satisfaite. D'après l'étude de cas précédente, on doit donc prendre

$$\alpha = \frac{1}{3}$$

et la solution du problème d'optimisation est alors donnée par

$$\lambda_s = \left[\frac{C_{opt}\left(A_1 + B_1\left(\frac{1}{P_1} - 1\right)\right)}{2\left(A_2B_2 + A_3C_{opt}\right)}\right]^{\frac{1}{3}}$$

et N est donné par le coût. □

Notons que si l'on prend pour la phase d'apprentissage les R_i optimaux, nous obtenons

$$\lambda_s = +\infty$$

Remarque 5.1.2. — *Insistons sur le fait que l'annulation observée dans la preuve ($A_2C_{opt} + B_2V_{opt} = 0$) provient directement du choix particulier des R_i, solutions d'un problème d'optimisation et conduit à prendre $\alpha = \frac{1}{3}$. Si nous n'avions pas pris cette fonction particulière des probabilités de transition pour les R_i, il n'y aurait pas eu d'annulation des coefficients et le α optimal aurait simplement été $\frac{1}{2}$, comme l'on peut s'y attendre a priori. Ceci souligne le fait qu'il est vraiment important de choisir les probabilités de transition les plus proches des valeurs optimales.*

Remarque 5.1.3. — *En notant que N est linéaire en C, nous remarquons que consacrer $\lambda_s N^{2/3}$ particules à la première phase revient à lui consacrer $\mu_s C^{2/3}$ particules, pour un certain μ_s dépendant de λ_s.*

5.1.4. Procédure à suivre en pratique. — En pratique, un budget est fixé à l'avance, ce qui revient à fixer le nombre total N de particules générées lors de la simulation. On génèrera $\rho_N = \lambda_s N^{1-\alpha}$ particules durant la phase d'apprentissage et $N - \rho_N$ durant la seconde. L'optimisation nous suggère de prendre

$$\alpha = \frac{1}{3} \quad \text{et} \quad \lambda_s = \left[\frac{C_{opt}\left(A_1 + B_1\left(\frac{1}{P_1} - 1\right)\right)}{2\left(A_2B_2 + A_3C_{opt}\right)} \right]^{\frac{1}{3}}.$$

Rappelons que les probabilités de transition P_i sont inconnues et donc que l'on ne peut pas évaluer explicitement λ_s. Cependant, ces probabilités conditionnelles étant bornées inférieurement et supérieurement, nous pouvons aisément déterminer une borne inférieure $\lambda_s(\min)$ (et resp. supérieure $\lambda_s(\max)$) de λ_s. Par conséquent, nous procèderons comme suit :

– Première phase-*phase d'apprentissage* :

  Nous générons d'abord $\lambda_s(\min)N^{2/3}$ particules, en procédant comme dans le modèle classique de branchement avec duplication de trajectoires avec $(R_i^0)_{i=1\cdots M}$ comme nombres de retirage choisis arbitrairement en début de simulation. Nous estimons ensuite les probabilités de transition, les nombres optimaux de retirage ainsi que le λ optimal par λ_{est}. Nous continuons à générer des particules jusqu'à ce que le nombre total de particules générées pendant la phase d'apprentissage soit $\lambda_{est}N^{2/3}$.

– Seconde phase :

  Nous générons $N - \rho_N$ particules que nous faisons évoluer comme lors de la première phase mais nous utilisons les estimateurs des $(R_i)_{i=1\cdots M}$ obtenus pendant la première phase et selon le schéma optimal donné au Chapitre 3 comme expliqué précédemment.

5.1.5. Probabilité d'extinction de l'algorithme sans troncature. — La probabilité d'arrêt de l'algorithme sans troncature est donnée par la probabilité d'arrêt lors de l'étape 1 et celle d'arrêt lors de l'étape 2 sachant qu'il n'y a pas eu d'arrêt lors de l'étape 1. Plus précisément,

Proposition 5.1.4. — *Si $P_{extinci}$ désigne la probabilité d'arrêt lors de l'étape i, la probabilité totale d'arrêt de l'algorithme est donnée par*

$$P_{extinc} = P_{extinc1} + (1 - P_{extinc1})P_{extinc2}$$

où

$$\begin{aligned} P_{extinc1} &= [P_1 g_1 \circ \cdots \circ g_M(0) + 1 - P_1]^{\rho_N} \\ P_{extinc2} &= \mathbb{E}\left[P_1 h_1 \circ \cdots \circ h_M(0) + 1 - P_1\right]^{N-\rho_N} \end{aligned}$$

$$\text{avec } \begin{cases} g_i(s) = (P_{i+1}s + 1 - P_{i+1})^{R_i^0} \\ h_i(s) = (P_{i+1}s + 1 - P_{i+1})^{\widehat{R}_i} \end{cases}$$

En prenant $s = 0$ dans la Remarque 3.6.5, nous obtenons directement

Corollaire 5.1.1. —

$$\left[1 - \left\{\frac{1}{\mathbb{P}(A) r_M^0 (1-s)} + \frac{1}{2}\sum_{j=1}^{M} \frac{g_j''(0)}{P_{j+2|0} r_{j+1}^0 P_{j+1} R_j^0}\right\}^{-1}\right]^{\rho_N}$$

$$\leq P_{extinc1} \leq \left[1 - \left\{\frac{1}{\mathbb{P}(A) r_M^0 (1-s)} + \sum_{j=1}^{M} \frac{g_j''(1)}{P_{j+2|0} r_{j+1}^0 P_{j+1} R_j^0}\right\}^{-1}\right]^{\rho_N}$$

avec

$$\begin{cases} g_j''(0) = R_j^0 (R_j^0 - 1) P_{j+1}^2 \left(1 - P_{j+1}\right)^{R_j^0 - 2} \\ g_j''(1) = R_j^0 (R_j^0 - 1) P_{j+1}^2 \end{cases}$$

Nous allons généraliser les encadrements obtenus au chapitre 3, section 3.6.4 afin d'estimer $P_{extinc2}$. Nous obtenons les encadrements suivants pour la fonction génératrice π_n de Z_n, processus de Galton-Watson en milieu aléatoire défini par les fonctions génératrices successives $\{h_j, j \geq 0\}$:

Proposition 5.1.5. —

$$1 - \left\{\left[\prod_{j=0}^{n-1} m_j (1-s)\right]^{-1} + \frac{1}{2}\sum_{j=0}^{n-1}\left(\mathbb{E}\left(h_j''(0)\right)/m_j \prod_{i=0}^{j} m_i\right)\right\}^{-1}$$

$$\leq \pi_n(s) \leq 1 - \left\{\left[\prod_{j=0}^{n-1} \mu_j (1-s)\right]^{-1} + \sum_{j=0}^{n-1} \mathbb{E}\left[\frac{h_j''(1)}{h_j'^2(1)}\right] / \prod_{i=0}^{j-1} \mu_i\right\}^{-1}$$

avec

$$\begin{cases} \mu_j = \mathbb{E}\left(h_j'(1)^{-1}\right)^{-1} \\ m_j = \mathbb{E}\left(h_j'(1)\right) \end{cases}$$

Démonstration. — Tout d'abord, notons r l'environnement aléatoire : $r = \{\widehat{R}_1, \cdots, \widehat{R}_M\}$ et $h_j(\bullet) = h_j(\widehat{R}_j, \bullet)$ les fonctions génératrices (aléatoires) des lois de reproduction. Il vient par application de l'inégalité de Jensen, les fonctions h_j étant convexes,

$$\begin{aligned} \pi_n(s) &= \mathbb{E}\left(\pi_n(r,s)\right) = \mathbb{E}\left(h_1 \circ \cdots \circ h_M(s)\right) \\ &\geq \mathbb{E}\left(h_1\right) \circ \mathbb{E}\left(h_2 \circ \cdots \circ h_M(s)\right) \geq \cdots \\ &\geq \mathbb{E}\left(h_1\right) \circ \mathbb{E}\left(h_2\right) \circ \cdots \circ \mathbb{E}\left(h_M(s)\right) \end{aligned}$$

Ainsi la fonction génératrice d'un processus de Galton-Watson en milieu aléatoire défini par les fonctions génératrices successives $\{h_j, j \geq 0\}$ est minorée par celle d'un processus de Galton-Watson en milieu variable défini par les espérances des fonctions génératrices successives précédentes. En appliquant le résultat de la Remarque 3.6.5, nous obtenons la minoration.

En utilisant le résultat de la Remarque 3.6.5, nous obtenons, en posant

$$Q_n = \prod_{j=0}^{n-1} h_j'(1),$$

$$\pi_n(r;s) \quad \leq \quad 1 - \left\{ \frac{1}{Q_n(1-s)} + \sum_{j=0}^{n-1} \frac{h_j''(1)}{h_j'(1) Q_{j+1}} \right\}^{-1}$$

et par une application de l'inégalité de Jensen à la fonction $\phi(x) = \frac{1}{x}$, il vient

$$\pi_n(s) \quad \leq \quad 1 - \left\{ \mathbb{E}\left[\frac{1}{Q_n(1-s)} + \sum_{j=0}^{n-1} \frac{h_j''(1)}{h_j'(1) Q_{j+1}} \right] \right\}^{-1}$$

D'où le résultat.

Ainsi la fonction génératrice d'un processus de Galton-Watson en milieu aléatoire défini par les fonctions génératrices successives $\{h_j, j \geq 0\}$ est majorée par celle d'un processus de Galton-Watson en milieu variable défini par les fonctions génératrices successives l_j ayant pour dérivée première en 1 μ_j et pour dérivée seconde en 1 $2\mu_j^2\mathbb{E}\left[h''_j(1)/h'^2_j(1)\right]$. □

En utilisant la définition des fonctions génératrices et en prenant $s = 0$ dans la proposition précédente,

Corollaire 5.1.2. —

$$\left[1 - \left\{\left[\prod_{j=1}^{M} m_j(1-s)\right]^{-1} + \frac{1}{2}\sum_{j=1}^{M}\left(\mathbb{E}\left(h''_j(0)\right)/P_1 m_j \prod_{i=1}^{j} m_i\right)\right\}^{-1}\right]^{N-\rho_N}$$

$$\leq P_{extinc2} \leq 1 - \left[\left\{\left[P_1\prod_{j=1}^{M}\mu_j(1-s)\right]^{-1} + \sum_{j=1}^{M}\mathbb{E}\left[\frac{h''_j(1)}{h'^2_j(1)}\right]/P_1\prod_{i=1}^{j-1}\mu_i\right\}^{-1}\right]^{N-\rho_N}$$

avec

$$\begin{cases} h''_j(0) = \widehat{R}_j(\widehat{R}_j - 1)P_{i+1}^2\left(1 - P_{i+1}\right)^{\widehat{R}_j - 2} \\ h''_j(1) = \widehat{R}_j(\widehat{R}_j - 1)P_{i+1}^2 \end{cases} \quad et \quad \begin{cases} \mu_j = P_{j+1}^{-1}\mathbb{E}\left(\widehat{R}_j^{-1}\right)^{-1} \\ m_j = P_{j+1}\mathbb{E}\left(\widehat{R}_j\right) \end{cases}$$

Ainsi l'extinction d'un processus de Galton-Watson en milieu aléatoire défini par les fonctions génératrices successives $\{h_j, j \geq 0\}$ s'éteint plus rapidement qu'un processus de Galton-Watson en milieu variable défini par les espérances de ces fonctions génératrices mais moins rapidement qu'un processus de Galton-Watson en milieu variable défini par les fonctions génératrices successives l_j ayant pour dérivée première en 1 μ_j et pour dérivée seconde en 1.

On obtient ainsi la proposition suivante :

Proposition 5.1.6. — *La probabilité d'extinction de l'algorithme est donnée par*

$$\begin{aligned}
&\left[1-\left\{\frac{1}{\mathbb{P}(A)r_M^0(1-s)}+\frac{1}{2}\sum_{j=1}^{M}\frac{g_j''(0)}{P_{j+2|0}r_{j+1}^0P_{j+1}R_j^0}\right\}^{-1}\right]^{\rho_N}\\
+&\left[1-\left\{\left[\prod_{j=1}^{M}m_j(1-s)\right]^{-1}+\frac{1}{2}\sum_{j=1}^{M}\left(\mathbb{E}\left(h_j''(0)\right)/P_1m_j\prod_{i=1}^{j}m_i\right)\right\}^{-1}\right]^{N-\rho_N}\\
\leq\ & P_{extinc}\leq\left[1-\left\{\frac{1}{\mathbb{P}(A)r_M^0(1-s)}+\sum_{j=1}^{M}\frac{g_j''(1)}{P_{j+2|0}r_{j+1}^0P_{j+1}R_j^0}\right\}^{-1}\right]^{\theta_N N}\\
&+1-\left[\left\{\left[P_1\prod_{j=1}^{M}\mu_j(1-s)\right]^{-1}+\sum_{j=1}^{M}\mathbb{E}\left[\frac{h_j''(1)}{h_j'^2(1)}\right]/P_1\prod_{i=1}^{j-1}\mu_i\right\}^{-1}\right]^{N-\rho_N}
\end{aligned}$$

avec

$$\begin{cases}
g_j''(0)=R_j^0(R_j^0-1)P_{j+1}^2\left(1-P_{j+1}\right)^{R_j^0-2}\\
g_j''(1)=R_j^0(R_j^0-1)P_{j+1}^2\\
h_j''(0)=\widehat{R}_j(\widehat{R}_j-1)P_{j+1}^2\left(1-P_{j+1}\right)^{\widehat{R}_j-2}\\
h_j''(1)=\widehat{R}_j(\widehat{R}_j-1)P_{j+1}^2
\end{cases}\quad et\quad\begin{cases}
\mu_j=P_{j+1}^{-1}\mathbb{E}\left(\widehat{R}_j^{-1}\right)^{-1}\\
m_j=P_{j+1}\mathbb{E}\left(\widehat{R}_j\right)
\end{cases}$$

5.2. Modèle Aldous-Vazirani

Nous proposons dans cette section une simulation en $M+1$ étapes pour palier le fait que l'on ne connaisse pas les niveaux et donc les nombres de retirage, basée sur l'heuristique proposée par Aldous et Vazirani dans [3]. Nous supposons que dans ce modèle aussi, les niveaux ont été fixés à l'avance.

5.2.1. Déroulement de la simulation. — De la même façon que précédemment, choisissons arbitrairement $M+1$ nombres $P_i^{(0)}$ dans $[a,b]$ et posons

$$R_i^{(0)}=\frac{1}{\sqrt{P_i^{(0)}P_{i+1}^{(0)}}}\sqrt{\frac{1-P_{i+1}^{(0)}}{1-P_i^{(0)}}}\quad\text{pour}\quad i=1\cdots M$$

On procède ensuite comme suit : on dispose de $(M+1)$ groupes de particules, le groupe i contenant $\theta_i N$ particules de telle sorte que $\sum_{i=1}^{M+1} \theta_i = 1$ c'est à dire que l'on envoie N particules au total.

Etape 1 : On envoie les N particules. Les particules du groupe 1 ayant at-

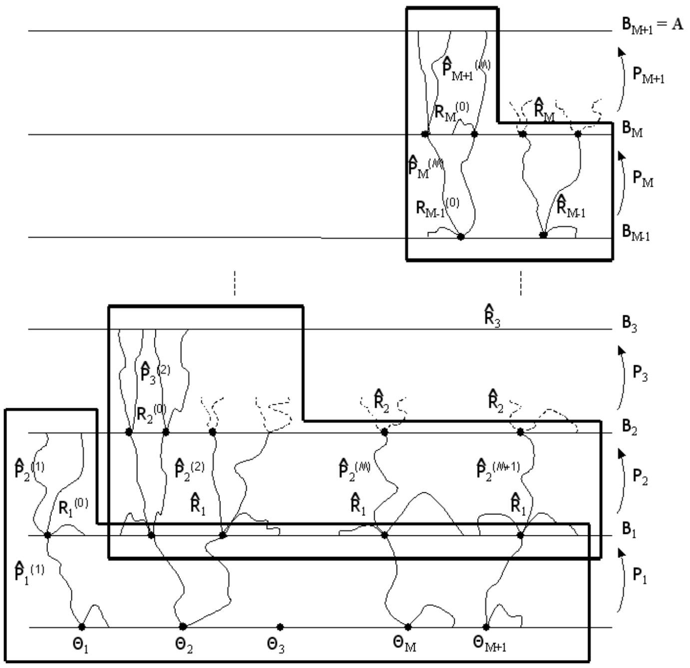


FIG. 2. Modèle de branchement avec duplication des trajectoires version Aldous-Vazirani

teint B_1 sont dupliquées en $R_1^{(0)}$ sous-particules. On compte celles arrivées en B_2. L'étude du premier groupe fournit ainsi un estimateur $\widehat{P}_1^{(1)}$ de P_1 et un estimateur $\widehat{P}_2^{(1)}$ de P_2. Posons alors

$$\widehat{R}_1 := \frac{1}{\sqrt{\widehat{P}_1^{(1)} \widehat{P}_2^{(1)}}} \sqrt{\frac{1-\widehat{P}_2^{(1)}}{1-\widehat{P}_1^{(1)}}}$$

et nous prendrons $\widehat{R}_1$ comme nombre de retirage depuis B_1 pour la prochaine étape de la simulation. Fin de l'étape 1.

Notation : Dans un souci de clarté et de lisibilité, nous noterons par la suite $\widehat{P}_i^j$ l'estimateur de P_i obtenu avec le j-ième groupe.

Etape i $(2 \leq i \leq M+1)$: Les groupes 1 à $i-1$ ont été écartés car ils ont servi à estimer $R_1, R_2, \cdots, R_{i-1}$. Les groupes restant contiennent chacun un nombre aléatoire de particules ayant atteint le niveau B_{i-1}. On compte le nombre de particules du groupe i ayant atteint le niveau B_{i-1}. Ces dernières sont dupliquées en $\widehat{R}_{i-1}$ sous-particules qui évoluent. Celles arrivées en B_i sont à leur tour dupliquées en $R_i^{(0)}$ sous-particules et on compte celles arrivées en B_{i+1}. L'étude de ce groupe fournit ainsi un estimateur $\widehat{P}_i^{(i)}$ de P_i et un estimateur $\widehat{P}_{i+1}^{(i)}$ de P_{i+1}. Posons alors de la même façon que précédemment

$$\widehat{R}_i := \frac{1}{\sqrt{\widehat{P}_i^{(i)}\widehat{P}_{i+1}^{(i)}}}\sqrt{\frac{1-\widehat{P}_{i+1}^{(i)}}{1-\widehat{P}_i^{(i)}}}$$

et nous prendrons $\widehat{R}_i$ comme nombre de retirage depuis B_i pour la prochaine étape de la simulation.

On compte ensuite les particules des groupes $i+1$ à $M+1$ qui ont atteint B_{i-1} et on les fait évoluer après les avoir dupliquées en $\widehat{R}_{i-1}$ sous-particules. Fin de l'étape i.

Hypothèses

Guidés par la démarche adoptée pour le modèle à deux étapes présenté en Section 5.1, nous supposons que

1. $\theta_i N \underset{N\to+\infty}{\longrightarrow} +\infty$ pour tout $i = 1\cdots M+1$,
2. $\theta_i N = o(\theta_{M+1}N)$ pour tout $i = 1\cdots M$.

De plus, nous supposons aussi que θ_i a la forme suivante $\lambda/N^\alpha M$ pour tout $i = 1\cdots M$ (avec $0<\alpha<1$ afin que les hypothèses soient vérifiées) et donc, puisque $\sum_{i=1}^{M+1}\theta_i = 1$,

$$\theta_{M+1} = 1 - \frac{\lambda}{N^\alpha}.$$

L'estimateur de $\mathbb{P}(A)$ est alors donné par

$$\widehat{P} = \frac{\mathbb{P}(A)}{\theta_{M+1}N}\sum_{i=1}^{\theta_{M+1}N}\frac{Z_{M+1}^{(i)}}{\mathbb{E}(Z_{M+1})}$$

5.2.2. Analyse du modèle. —

5.2.2.1. Coût. — Le coût moyen est donné par

$$C = N\left[\sum_{i=1}^{M}\theta_i R_i^{(0)}\mathbb{E}\left(\widehat{p}_i^{(i)}\right)\mathbb{E}\left(\prod_{j=1}^{i-1}\widehat{R}_j\right)+\sum_{i=1}^{M+1}\mathbb{E}\left(\prod_{j=1}^{i-1}\widehat{R}_j\right)\sum_{j=i}^{M+1}\theta_j\mathbb{E}\left(\widehat{p}_{j-1}^{(j)}\right)\right]$$

en notant

$$\widehat{p}_j^{(i)} = \widehat{P}_1^{(i)}\cdots\widehat{P}_j^{(i)}, \qquad i,j = 1\cdots M+1$$

En procédant comme dans la Section 5.1.3.1 et en utilisant les notations précédemment introduites, on peut montrer que

Proposition 5.2.1. — *Le coût moyen est donné par*

$$\begin{aligned}C = N\Bigg[&p_M\sum_{i=0}^{M}r_i+\frac{p_M}{2\lambda N^{1-\alpha}M}\sum_{i=0}^{M}r_i\Sigma_i\\&+\frac{\lambda}{N^\alpha}\left[\frac{1}{M}\left\{\sum_{i=0}^{M-1}R_i^{(0)}r_ip_{i+1}+\sum_{i=0}^{M}r_i\sum_{j=i}^{M-1}p_j\right\}-p_M\sum_{i=0}^{M}r_i\right]\Bigg]\end{aligned}$$

avec

$$\Sigma_i = r_i\sum_{j=1}^{i}\frac{1}{r_{j-1}p_j}\left[P_j+\frac{3}{4}\frac{(1-2P_j)^2}{1-P_j}+\frac{1}{R_j^{(0)}}\frac{3/4-P_{j+1}}{P_{j+1}(1-P_{j+1})}\right]$$

5.2.2.2. Variance. — En utilisant les notations précédemment introduites, il vient

Proposition 5.2.2. — *La variance est donnée par*

$$\begin{aligned}var(\widehat{P}) \;=\; \frac{\mathbb{P}(A)^2}{\theta_{M+1}N}\Bigg[&\sum_{i=0}^{M}\left(\frac{1}{p_{i+1}}-\frac{1}{p_i}\right)\frac{1}{\prod_{k=1}^{i}\mathbb{E}\left(\widehat{R}_k\right)}\\&+\sum_{i=1}^{M}\frac{var\left(\widehat{R}_i\right)}{p_k\prod_{k=1}^{i}\mathbb{E}\left(\widehat{R}_k\right)\mathbb{E}\left(\widehat{R}_i\right)}\Bigg]\end{aligned}$$

Démonstration. — En effet, tout d'abord

$$
\begin{aligned}
\mathbb{E}(Z_{M+1}) &= \mathbb{E}\left(\sum_{i=1}^{Z_M} X_i^M\right) = \mathbb{E}\left(Z_M \mathbb{E}\left(X_i^M | Z_M\right)\right) = \mathbb{E}\left(Z_M \mathbb{E}\left(X_i^M\right)\right) \\
&= \mathbb{E}\left(Z_M\right) \mathbb{E}\left(\mathbb{E}\left(X_i^M | \widehat{R}_M\right)\right) = \mathbb{E}\left(Z_M\right) \mathbb{E}\left(\widehat{R}_M\right) P_{M+1} = \cdots \\
&= p_{M+1} \prod_{i=1}^{M} \mathbb{E}\left(\widehat{R}_i\right)
\end{aligned}
$$

Ensuite

$$
\begin{aligned}
\operatorname{var}(Z_{M+1}) &= \operatorname{var}\left(\sum_{i=1}^{Z_M} X_i^M\right) = \operatorname{var}\left(\sum_{i=1}^{Z_M} \mathbb{E}\left(X_i^M\right)\right) + \mathbb{E}\left(\sum_{i=1}^{Z_M} \operatorname{var}\left(X_i^M\right)\right) \\
&= \operatorname{var}\left(Z_M\right)\left(\mathbb{E}\left(\widehat{R}_M\right) P_{M+1}\right)^2 + \mathbb{E}\left(Z_M\right) \operatorname{var}\left(\mathbb{E}\left(X^M | \widehat{R}_M\right)\right) \\
&\quad + \mathbb{E}\left(Z_M\right) \mathbb{E}\left(\operatorname{var}\left(X^M | \widehat{R}_M\right)\right) \\
&= \operatorname{var}\left(Z_M\right)\left(\mathbb{E}\left(\widehat{R}_M\right) P_{M+1}\right)^2 + \mathbb{E}\left(Z_M\right) \operatorname{var}\left(\widehat{R}_M\right) P_{M+1}^2 \\
&\quad + \mathbb{E}\left(Z_M\right) \mathbb{E}\left(\widehat{R}_M\right) P_{M+1}(1 - P_{M+1}) \\
&= \cdots \\
&= \operatorname{var}\left(Z_1\right) \prod_{i=1}^{M}\left(\mathbb{E}\left(\widehat{R}_i\right) P_{i+1}\right)^2 + \sum_{i=1}^{M} \mathbb{E}\left(Z_i\right) \operatorname{var}\left(\widehat{R}_i\right) P_{i+1}^2 \prod_{k=i+1}^{M}\left(\mathbb{E}\left(\widehat{R}_k\right) P_{k+1}\right)^2 \\
&\quad + \sum_{i=1}^{M} \mathbb{E}\left(Z_i\right) \mathbb{E}\left(\widehat{R}_i\right) P_{i+1}(1 - P_{i+1}) \prod_{k=i+1}^{M}\left(\mathbb{E}\left(\widehat{R}_k\right) P_{k+1}\right)^2 \\
&= \sum_{i=1}^{M} \mathbb{E}\left(Z_i\right) \operatorname{var}\left(\widehat{R}_i\right) P_{i+1}^2 \prod_{k=i+1}^{M}\left(\mathbb{E}\left(\widehat{R}_k\right) P_{k+1}\right)^2 \\
&\quad + \sum_{i=0}^{M} \mathbb{E}\left(Z_i\right) \mathbb{E}\left(\widehat{R}_i\right) P_{i+1}(1 - P_{i+1}) \prod_{k=i+1}^{M}\left(\mathbb{E}\left(\widehat{R}_k\right) P_{k+1}\right)^2
\end{aligned}
$$

D'où l'on déduit le résultat recherché. □

En procédant comme dans la Section 5.1.3.2, on peut montrer que

Proposition 5.2.3. — *Le variance est donnée par*

$$var(\widehat{P}) = \frac{\mathbb{P}(A)^2}{\theta_{M+1}N}\left[V_{opt}\left(1-\frac{\lambda}{N^\alpha}\right)+\frac{1}{4\lambda N^{1-\alpha}M}\sum_{i=1}^{M}\frac{A_i}{p_i^2 r_{i-1}^2}\right]$$

avec

$$A_i = P_i(1-2P_{i+1})^2 + \frac{1}{R_i^{(0)}P_{i+1}(1-P_{i+1})}$$

5.2.2.3. Optimisation en $\theta_1, \theta_2, \cdots, \theta_{M+1}$. — Notre but est de minimiser la variance à coût fixé. En procédant comme en Section 5.1.3.3, on déduit qu'il faut égaler $1-\alpha$ et α c-à-d prendre $\alpha = 1/2$.

En procédant comme suit, on retrouve les paramètres proposés dans [3]. Mais notons bien que la démarche d'optimisation est alors totalement différente de celle proposée dans cette thèse. Considérons maintenant l'algorithme global comme $M+1$ sous-algorithmes, chacun de ces sous-algorithmes représentant l'évolution des différents groupes de particules. Notons que l'expression du coût est globale et ne reflète pas le coût de chacun de ces sous-algorithmes. Remarquons aussi que les particules du groupe i c-à-d de l'algorithme i $(1 \le i \le M)$ franchissent $i+1$ niveaux alors que celles du dernier groupe c-à-d de l'algorithme $M+1$ franchissent $M+1$ niveaux. Allouons donc, à chaque algorithme i $(1 \le i \le M)$ et $M+1$, les coûts $(i+1) \times c_{niv}$ et $(M+1) \times c_{niv}$ pour un certain «coût par niveau» c_{niv}. Ainsi

$$C = \sum_{i=2}^{M+1} i c_{niv} + (M+1)c_{niv} = \frac{M^2+5M+2}{2}c_{niv}$$

D'où la valeur de c_{niv}.

Mais le coût du groupe 1 est $2c_{niv} = \theta_1 N\left(1+\mathbb{E}(\widehat{P}_1^{(1)})R_1^{(0)}\right)$ et ainsi

$$\theta_1 = \frac{1}{N}\frac{2c_{niv}}{1+\mathbb{E}(\widehat{P}_1^{(1)})R_1^{(0)}}$$

De même, le coût du groupe 2 est

$$3c_{niv} = \theta_2 N\left(1+\mathbb{E}(\widehat{P}_1^{(2)})\mathbb{E}(\widehat{R}_1)+\mathbb{E}(\widehat{P}_1^{(2)}\widehat{P}_2^{(2)})R_2^{(0)}\mathbb{E}(\widehat{R}_1)\right)$$

et donc

$$\theta_2 = \frac{1}{N}\frac{3c_{niv}}{1+\mathbb{E}(\widehat{P}_1^{(2)})\mathbb{E}(\widehat{R}_1)+\mathbb{E}(\widehat{P}_1^{(2)}\widehat{P}_2^{(2)})R_2^{(0)}\mathbb{E}(\widehat{R}_1)}$$

et ainsi de suite... Enfin, le coût du groupe $(M+1)$ est

$$(M+1)c_{niv} = \theta_{M+1} N \left(1 + \sum_{i=1}^{M} \mathbb{E}\left(\prod_{j=1}^{i} \left(\widehat{P}_j^{(M+1)} \widehat{R}_j\right)\right)\right)$$

et ainsi

$$\theta_{M+1} = \frac{1}{N} \frac{(M+1)c_{niv}}{1 + \sum_{i=1}^{M} \mathbb{E}\left(\prod_{j=1}^{i} \left(\widehat{P}_j^{(M+1)} \widehat{R}_j\right)\right)}$$

On remarque que lorsque l'on connaît les paramètres et que l'on prend les valeurs optimales, on obtient

$$\theta_i = \frac{c_{niv}}{N} \quad \text{pour} \quad i = 1 \cdots M+1$$

ce qui signifie que les coûts sont uniformément répartis sur chacun des sous-algorithmes. On prendra donc dans la pratique

$$\theta_i = \frac{1}{M+1}$$

puisqu'on veut aussi $\sum_{i=1}^{M+1} \theta_i = 1$.

5.2.3. Précisions. — Comme longuement évoqué précédemment, les valeurs des nombres de retirage ne sont pas toujours entières. Afin de remédier à ce problème, nous avons vu au Chapitre 4 qu'il faut considérer les nombres de retirage comme des variables aléatoires et si R_i est le nombre optimal au niveau i, il s'agit de générer à chaque succès en i une variable aléatoire, notée par la suite $N(R_i)$, sur $\{\lfloor R_i \rfloor, \lceil R_i \rceil\}$ de moyenne R_i. C'est à dire que

$$\mathbb{P}(N(R_i) = \lfloor R_i \rfloor) = \lceil R_i \rceil - R_i \quad \text{et} \quad \mathbb{P}(N(R_i) = \lceil R_i \rceil) = R_i - \lfloor R_i \rfloor$$

En Section 5.1 ainsi que dans cette section, cette manipulation a été omise car source de complexité dans l'analyse précise et l'optimisation de l'algorithme. En pratique, nous prendrons l'entier le plus proche de la valeur optimale, comme expliqué en Section 4.6.

Remarquons cependant que nous obtenons alors les deux lemmes suivants :

Lemme 5.2.1. — *Soit Z une loi de Bernoulli sur $\{a, b\}$ telle que $\mathbb{P}(Z = a) = p$ et $\mathbb{P}(Z = b) = q$. Alors*

$$\begin{aligned} \mathbb{E}(Z) &= ap + bq = b + (a-b)p \\ var(Z) &= (b-a)^2 p(1-p) \end{aligned}$$

et

$$\mathbb{E}\left(\frac{1}{Z}\right) = \frac{1}{\mathbb{E}(Z)} + \frac{1}{ab}\frac{var(Z)}{\mathbb{E}(Z)}$$

Dans le cas particulier où $b = a + 1$ avec $a \geq 1$,

$$\mathbb{E}\left(\frac{1}{Z}\right) = \frac{1}{\mathbb{E}(Z)} + \frac{1}{a(a+1)}\frac{var(Z)}{\mathbb{E}(Z)}$$

et

$$\frac{1}{\mathbb{E}(Z)} \leq \mathbb{E}\left(\frac{1}{Z}\right) \leq \frac{1}{\mathbb{E}(Z)} + \frac{9}{8}\frac{1}{\mathbb{E}(Z)}$$

Démonstration. — Les deux premières égalités sont évidentes. Pour la troisième,

$$\begin{aligned} \mathbb{E}\left(\frac{1}{Z}\right) - \frac{1}{\mathbb{E}(Z)} &= \frac{p}{a} + \frac{q}{b} - \frac{1}{ap+bq} \\ &= \frac{1}{ab\mathbb{E}(Z)}\{pb(ap+bq) + aq(ap+bq) - ab\} \\ &= \frac{1}{ab\mathbb{E}(Z)}pq\mathrm{var}(Z) \end{aligned}$$

L'encadrement, dans le cas particulier, s'obtient d'une part grâce à l'inégalité de Jensen et d'autre part par décroissance de la fonction $f(x) = 1/[x(1+x)]$ et par l'inégalité $\mathrm{var}(Z) \leq 1/4$. □

On en déduit le lemme suivant qui donne la valeur de $\mathbb{E}\left(N(R)\right)$ et des encadrements de $\mathbb{E}\left(\frac{1}{N(R)}\right)$ lorsque R est déterministe ou aléatoire :

Lemme 5.2.2. — *(i) Lorsque R est déterministe, on a par définition $\mathbb{E}(N(R)) = R$ et l'encadrement*

$$\frac{1}{R} - \leq \mathbb{E}\left(\frac{1}{N(R)}\right) \leq \frac{9}{8R}$$

(ii) Lorsque R est aléatoire, on a par définition $\mathbb{E}(N(R)) = \mathbb{E}(R)$ et l'encadrement

$$\mathbb{E}\left(\frac{1}{R}\right) \leq \mathbb{E}\left(\frac{1}{N(R)}\right) \leq \frac{9}{8}\mathbb{E}\left(\frac{1}{R}\right)$$

5.2.4. Conclusion. — Nous remarquons que le coût semble plus élevé et que la variance sera plus grande puisque elle est en $1/\theta_{M+1}N$ alors que celle du modèle adaptatif précédemment présenté est en $1/N$. Ici, $N^{1/2}$ particules servent à la phase d'apprentissage (comprenant les algorithmes 1 à M) contre $N^{2/3}$ dans le premier modèle mais toute l'information acquise lors de cette phase d'apprentissage est perdue car elle ne peut être prise en compte dans l'estimateur.

5.3. Un modèle qui garde le système particulaire en vie

Le but de ce modèle introduit par François LeGland et Nadia Oudjane et présenté par exemple dans [47, 48] est ici aussi d'estimer la probabilité d'un événement rare en utilisant des niveaux de branchement. Ici les niveaux de transition sont fixés à l'avance mais le nombre de particules générées à chaque niveau de transition est aléatoire et on génère jusqu'à ce qu'un nombre, fixé à l'avance, de particules atteigne le niveau supérieur. Ainsi l'algorithme ne meurt jamais.

5.3.1. Déroulement de la simulation. — Nous nous plaçons dans les mêmes conditions que pour les systèmes particulaires présentés en Section 2.4. Le but de cet algorithme est de garder automatiquement le système particulaire en vie ce qui assure sa non-extinction.

Pour tout $H > 0$ entier, et tout $k = 0, \cdots, M+1$, définissons le nombre aléatoire de particules (temps d'arrêt)

$$N_k^H := \inf \left\{ N \geq 1 \ : \ \sum_{i=1}^{N} g_k(\zeta_k^i) \geq H \sup_{x \in E} g_k(x) \right\}$$

où les variables aléatoires $\zeta_0^1, \cdots, \zeta_0^i, \cdots$ sont i.i.d. de loi μ_0 et où les variables aléatoires $\zeta_k^1, \cdots, \zeta_k^i, \cdots$, conditionnées au passé, sont i.i.d. de loi $\mu_{k-1}^H Q_k$. L'approximation particulaire $\{\mu_k^H, \ k = 0, 1, \cdots M+1\}$ est maintenant paramétrée par le seuil $H > 0$ et son évolution est décrite par le diagramme suivant :

$$\mu_{k-1}^H \underset{\text{sélection}}{\longrightarrow} \widehat{\mu}_k^H = S^{N_k^H}(\mu_{k-1}^H Q_k) \underset{\text{mutation}}{\longrightarrow} \mu_k^H = g_k \widehat{\mu}_k^H \tag{83}$$

avec la condition initiale $\mu_0^H = g_0 \widehat{\mu}_0^H$ et $\widehat{\mu}_0^H = S^{N_0^H}(\mu_0)$ et où la notation $S^N(\mu)$ désigne la distribution empirique associée à un N-échantillon de loi commune μ.

On peut définir l'approximation du flot non renormalisé et on vérifie que les équations satisfaites dans le cadre des systèmes particulaires le sont aussi pour cet algorithme. Clairement, $N_k^H \geq H$ et si $\mu_{k-1}^H Q_k(g_k) > 0$, on peut montrer que le nombre aléatoire de particules N_k^H est fini presque sûrement. De plus,

$$\widehat{\mu}_0^H(g_0) = S^{N_0^H}(\mu_0)(g_0) = \frac{1}{N_0^H}\sum_{i=1}^{N_0^H} g_0(\zeta_0^i) \geq \frac{H}{N_0^H}\sup_{x\in E} g_0(x) > 0,$$

et

$$\widehat{\mu}_k^H(g_k) = S^{N_k^H}(\mu_k)(g_k) = \frac{1}{N_k^H}\sum_{i=1}^{N_k^H} g_k(\zeta_k^i) \geq \frac{H}{N_k^H}\sup_{x\in E} g_k(x) > 0,$$

pour tout $k = 0, \cdots, M+1$ c-à-d le système particulaire ne s'éteint pas et l'algorithme peut toujours continuer, par construction.

Dans notre cas particulier, où nous étudions la probabilité d'un événement (rare), les fonctions de sélection g_k sont des indicatrices et on obtient

$$N_k^H := \inf\left\{N \geq 1 \ : \ \sum_{i=1}^{N} g_k(\zeta_k^i) = H\right\}$$

et ainsi on génère un nombre aléatoire N_k^H de particules jusqu'à ce qu'exactement H trajectoires atteignent le niveau supérieur.

Concrètement, de même que pour les systèmes particulaires, l'algorithme consiste en la répétition de trois étapes :

1. **sélection** des particules de poids non nul : indépendamment pour tout $i = 1 \cdots N_k^H$,
$$t_{k-1}^i \sim \left(\omega_{k-1}^1, \cdots, \omega_{k-1}^{N_{k-1}^H}\right)$$
et
$$\widehat{\zeta}_{k-1}^i = \zeta_{k-1}^{t_{k-1}^i} \ \text{ et } \ \widehat{T}_{k-1}^i = T_{k-1}^{t_{k-1}^i}$$
2. **mutation** : indépendamment pour tout $i = 1 \cdots N_k^H$,
$$\zeta_k^i = \left((T_k^i, X^i(t)),\ \widehat{T}_{k-1}^i \leq t \leq T_k^i\right)$$
selon le modèle markovien à temps continu sous-jacent, partant de $\widehat{\zeta}_{k-1}^i$
$$T_k^i = \inf\left\{t \geq \widehat{T}_{k-1}^i \ : \ (t, X^i(t)) \in A_k\right\}$$

La taille de la population N_k^H est choisie de telle sorte que H trajectoires exactement réussissent à atteindre le niveau suivant, c-à-d

$$N_k^H := \inf\left\{N \geq 1 \ : \ \sum_{i=1}^{N} g_k(\zeta_k^i) = H\right\} = \inf\left\{N \geq 1 \ : \ |I_k^N| = H\right\}$$

3. **pondération** selon les succès d'atteinte du niveau supérieur : pour tout $i = 1 \cdots N_k^H$,
$$\omega_k^i \propto g_k(\zeta_k^i) = \mathbf{1}_{\zeta_k^i \in A_k}$$

Dans ce cas, les probabilités de transition sont approchées par

$$P_k = \mathbb{P}\left(T_k \leq T | T_{k-1} \leq T\right) = \widehat{\mu}_k(g_k) \approx \widehat{\mu}_k^H(g_k) = \frac{|I_k^{N_k^H}|}{N_k^H} = \frac{H}{N_k^H}$$

et la probabilité de l'événement rare est approchée par

$$\mathbb{P}(V(X) \geq a) = \gamma_{M+2}(1) \approx \gamma_{M+2}^H(1) = \prod_{k=0}^{M+1} \widehat{\mu}_k^H(g_k) = \prod_{k=0}^{M+1} \frac{H}{N_k^H}$$

5.3.2. Propriétés asymptotiques. — Enonçons maintenant le théorème central limite :

Théorème 5.3.1. —
$$\sqrt{H}\left(\gamma_{M+2}^H(1) - \mathbb{P}(V(X) \geq a)\right) \underset{\mathcal{L}, H \to +\infty}{\longrightarrow} G_{M+2}$$
où G_{M+2} est un processus gaussien centré de variance

$$g_{M+2}^2 = \mathbb{P}(V(X) \geq a)^2 \sum_{i=0}^{M} (1 - P_{i+1})$$

La variance asymptotique obtenue ci-dessus est à comparer à

$$v_{M+1}^2 = \mathbb{P}(V(X) \geq a)^2 \sum_{i=0}^{M} \left(\frac{1}{P_{i+1}} - 1\right),$$

variance asymptotique obtenue pour les systèmes particulaires.

Une autre normalisation (taille moyenne des systèmes particulaires) conduit au théorème central limite suivant :

Théorème 5.3.2. —
$$\left[\frac{1}{M+1}\sum_{i=1}^{M+1} N_k^H\right]^{1/2} \left(\gamma_{M+2}^H(1) - \mathbb{P}(V(X) \geq a)\right) \underset{\mathcal{L}, H\to+\infty}{\longrightarrow} G'_{M+2}$$
où G'_{M+2} est un processus gaussien centré de variance
$$g'^2_{M+2} = \mathbb{P}(V(X) \geq a)^2 \left[\frac{1}{M+1}\sum_{i=1}^{M+1} \frac{1}{P_i}\right]^2 \sum_{i=0}^{M} (1 - P_{i+1})$$

où
$$\frac{1}{H}\left[\frac{1}{M+1}\sum_{i=1}^{M+1} N_k^H\right] \underset{\mathbb{P},\ H\to\infty}{\longrightarrow} \frac{1}{M+1}\sum_{i=1}^{M+1}\frac{1}{P_i}$$

Pour plus de détails et les preuves de ces résultats, nous renvoyons le lecteur à [48].

5.4. Un modèle où les niveaux sont placés durant la simulation

Le but de ce modèle introduit par Arnaud Guyader et Frédéric Cérou et présenté dans [19] est ici aussi d'estimer la probabilité d'un événement rare en utilisant des niveaux de branchement mais le procédé est sensiblement différent de ceux présentés dans les deux paragraphes précédents. Afin de résoudre le problème du choix des niveaux, les niveaux de transition sont placés pas à pas durant la simulation de façon à avoir un processus de branchement aussi critique que possible et afin d'éviter une extinction prématurée de l'algorithme.

5.4.1. Déroulement de la simulation. — Nous considèrons ici un processus de Markov fort à valeurs dans $\mathbb{R}$ et à trajectoires continues. Supposons que $X_0 = x_0 > 0$, que l'origine 0 est un point attractif. Nous souhaitons estimer la probabilité que le processus atteigne la hauteur $M+1$ avant d'atteindre 0 :
$$\mathbb{P}(A) := \mathbb{P}(T_M < T_0)$$
où T_i représente le premier temps d'atteinte de i : $T_i := \inf\{t \geq 0\ :\ X_t = i\}$.

On procède en plusieurs étapes équivalentes, exceptées la première et la dernière. Le principe est le suivant : on travaille à taille de population n constante et on place les niveaux au fur et à mesure de la simulation afin d'avoir exactement une proportion $p = \frac{k}{N}$ de trajectoires qui atteint chaque niveau supérieur. On se prémunit ainsi de tout arrêt prématuré de l'algorithme. Plus précisément, on procède comme suit

• Etape 1 : On génère N trajectoires i.i.d. (X_t^j) selon la loi du processus (X_t) partant de $X_0^j = x_0 > 0$. On attend que les trajectoires soient revenues en 0 (pour la trajectoire j, 0 étant attractif, il faut attendre un temps T_0^j de moyenne finie). Notons

$$S_{N,j}^1 = \sup_{0 \le t \le T_0^j} X_t^j$$

et on réordonne l'échantillon $\left(S_{N,1}^1, S_{N,2}^1, \cdots, S_{N,N}^1\right)$ par ordre croissant :

$$S_{N,(1)}^1 \le \cdots \le S_{N,(N-k)}^1 \le \cdots \le S_{N,(N)}^1$$

On garde en mémoire la quantité

$$\widehat{q}_1 = S_{N,(N-k)}^1.$$

• Etape 2 : On garde $(S_{N,(N-k+1)}^1, \cdots, S_{N,(N)}^1)$ inchangé, que l'on renomme $(S_{N,N-k+1}^2, \cdots, S_{N,N}^2)$. On génère $(N-k)$ trajectoires (X_t^j) partant de $\widehat{q}_1$. On attend que ces $(N-k)$ trajectoires aient atteint 0. Pour chaque $j \in \{1, \cdots, N-k\}$, notons

$$S_{N,j}^2 = \sup_{0 \le t \le T_0^j} X_t^j$$

et on réordonne l'échantillon $\left(S_{N,1}^2, S_{N,2}^2, \cdots, S_{N,N}^2\right)$ par ordre croissant :

$$S_{N,(1)}^2 \le \cdots \le S_{N,(N-k)}^2 \le \cdots \le S_{N,(N)}^2$$

On garde en mémoire la quantité

$$\widehat{q}_2 = S_{N,(N-k)}^2.$$

• ...

• On répète la procédure jusqu'à ce que $\widehat{q}_{n+1} \ge M + 1$. Parmi l'échantillon $(S_{N,1}^n, \cdots, S_{N,N}^n)$, une proportion $r > 0$ est supérieure à $M + 1$.

• On estime ensuite $\mathbb{P}(A)$ par

$$\widehat{P}_N = rp^n$$

5.4.2. Propriétés asymptotiques. — Hypothèses $(\mathcal{H})$: Le processus de Markov fort $(X_t)_t$ part de $x_0 > 0$, avec 0 point attractif. $(X_t)_t$ a des trajectoires continues et la fonction de répartition F de la variable aléatoire $S = \sup_{0 \le t \le T_0} X_t$ est continue.

Par l'étude des processus empiriques, nous pouvons établir la convergence presque sûre de l'estimateur

Proposition 5.4.1. — *Sous les hypothèses* $(\mathcal{H})$,

$$\widehat{P}_N \underset{N\to\infty,\ p.s.}{\longrightarrow} \mathbb{P}(A)$$

Notons $\rho = \mathbb{P}(A)p^{-n}$. En utilisant le même type de techniques, la normalité asymptotique de l'estimateur est ensuite établie :

Théorème 5.4.1. — *Sous les hypothèses* $(\mathcal{H})$,

$$\sqrt{N}(\mathbb{P}(A) - \widehat{P}_N) \underset{N\to\infty,\ \mathcal{L}}{\longrightarrow} \mathcal{N}(0, \sigma^2)$$

avec $\sigma^2 = \mathbb{P}(A)^2 \left(n\frac{1-p}{p} + \frac{1-\rho}{\rho} \right)$.

Pour plus de détails et les preuves de ces résultats, nous renvoyons le lecteur à [19].

5.5. Comparaison des différents modèles

Dans cet exemple, on se place dans $\mathbb{R}^n$. Soient $a \in (\mathbb{R}^+)^n$, $b > 0$ et

$$B = \left\{ (x_1, \cdots, x_n) \ : \ x_i \ge 0, \ \ \sum_{i=1}^n x_i \le 1 \right\}$$

Définissons maintenant

$$D = \{(x_1, \cdots, x_n) \in B \ : \ x.a \le b\}$$

Nous cherchons à estimer $\mathrm{vol}(D)/\mathrm{vol}(B)$, ce qui est fait en utilisant différentes méthodes :

- par la simulation simple,
- par un algorithme de Metropolis,
- par les systèmes particulaires,

– par le modèle de branchement avec duplication des trajectoires,
– par le modèle de branchement avec duplication des trajectoires en deux étapes,
– par un algorithme séquentiel,
– par un algorithme où l'on place les niveaux au cours de la simulation.

Décrivons maintenant le principe de ces différents algorithmes de simulation.

Introduisons tout d'abord des seuils intermédiaires, fixés de façon arbitraire.

La simulation simple- On génère N vecteurs aléatoirement et uniformément dans B et on calcule la proportion de ces vecteurs dans D.

L'algorithme de Metropolis- Voir Annexe 3. On génère N vecteurs aléatoirement et uniformément dans B. On fait «évoluer», entre chaque seuil, chacun de ces vecteurs (étape de mutation) selon l'échantillonneur de Gibbs : pour chaque vecteur, on tire au hasard une coordonnée que l'on modifie en tirant une variable uniforme sur [0,borne], où la quantité «borne» est telle que le vecteur reste, après cette mutation, dans la région désirée B. On procède ainsi sur toutes les coordonnées.

Les systèmes particulaires [20]- Voir section 2.4. Le principe de base est le même. On génère N vecteurs aléatoirement et uniformément dans B. On fait «évoluer», entre chaque seuil, chacun de ces vecteurs (étape de mutation) selon l'échantillonneur de Gibbs. On dénombre les vecteurs ayant atteint le seuil supérieur et on rééchantillonne les autres vecteurs sur ces derniers (étape de sélection), afin d'avoir une taille de population constante.

Le modèle de branchement avec duplication des trajectoires- Le principe de base est le même. On génère N vecteurs aléatoirement et uniformément dans B. On fait «évoluer», entre chaque seuil, chacun de ces vecteurs (étape de mutation) selon l'échantillonneur de Gibbs. On dénombre les vecteurs ayant atteint le seuil supérieur et on les duplique en R vecteurs (étape de sélection).

Le modèle de branchement avec duplication des trajectoires en deux étapes- Le principe de base est le même que pour le modèle précédent. Mais on ne connaît pas les niveaux et on procède en deux étapes comme expliqué en Section 5.1.

Un algorithme séquentiel [47, 48]- Voir Section 5.3. On applique le principe de l'algorithme particulaire mais en fixant le nombre de succès H.

Un algorithme adaptatif [19]- Voir section 5.4. Ici, les niveaux ne sont pas fixés à l'avance de façon arbitraire mais au fur et à mesure de la simulation de façon adaptative : on déroule la simulation selon le principe précédent et on fixe le premier seuil de telle sorte qu'une proportion p, fixée arbitrairement, de vecteurs aient atteint ce seuil. On continue jusqu'à atteindre D.

Pour l'application, nous avons pris les valeurs numériques de [54] :
- $n = 10$,
- $a = [0.513, 0.944, 0.960, 0.116, 0.032, 0.944, 0.691, 0.489, 0.020, 0.710]$,
- $b = a.\mathbf{1}/20 = 0.271$.

Dans ce cas,
$$\mathrm{vol}(D)/\mathrm{vol}(B) \approx 0.02.$$
Fixons les seuils de la façon suivante :
$$a.\mathbf{1}/10, \quad a.\mathbf{1}/13, \quad a.\mathbf{1}/15, \quad a.\mathbf{1}/17.$$
Le tableau ci-dessous contient les résultats des différentes simulations pour $N = 10^4$ vecteurs générés au départ :

	Estimation	Coût simulation
Simulation simple	0.018101	1.0e+04
Simulation Metropolis	0.029322	4.3e+04
Algorithmes particulaires	0.027469	6.0e+04
Branch. ac dupli. des traj. $R = 2$	0.021825	7.6e+04
Branch. ac dupli. des traj. $R = 5$	0.020001	1.2e+06
Branch. ac dupli. des traj.-2 étapes $R = 2$	0.022432	1.1e+05
Branch. ac dupli. des traj.-2 étapes $R = 5$	0.022234	2.1e+05
Algorithme séquentiel $H = 10^2$	0.017674	1.1e+03
Algorithme séquentiel $H = 10^2$	0.017674	1.1e+03
Algorithme séquentiel $H = 10^3$	0.022210	1.0e+04
Algorithme séquentiel $H = 10^4$	0.020420	1.1e+05
Algorithme séquentiel $H = 10^5$	0.020200	1.1e+06
Algorithme adaptatif $p = 1/3$	0.030313	4.0e+04
Algorithme adaptatif $p = 1/2$	0.027984	6.0e+04
Algorithme adaptatif $p = 2/3$	0.023403	1.0e+05
Algorithme adaptatif $p = 3/4$	0.021883	1.4e+05

5.6. Conclusion

Nous avons proposé dans ce chapitre un modèle en deux étapes basé sur le modèle de branchement avec duplication des trajectoires afin de contourner le problème des probabilités de transition inconnues. Contrairement à ce à quoi on s'attendait a priori, des développements asymptotiques à l'ordre 2 étaient nécessaires pour la première phase alors que l'ordre 1 était suffisant pour la seconde phase. Une analyse précise a finalement montré que nous devions consacrer asymptotiquement $\mu_s C^{2/3}$ particules à la phase d'apprentissage contre $C/C_{opt} - \mu_s C^{2/3}$ à la seconde phase, où μ_s se déduit de l'optimisation de l'algorithme et peut être exprimé explicitement ; c-à-d que si on suppose que le nombre ρ_N de particules générées dans la phase d'apprentissage a asymptotiquement la forme $\mu_\alpha(C)C^{1-\alpha}$, on doit prendre $\alpha = 1/3$.

Ce résultat provient directement de l'annulation observée pendant l'optimisation qui est elle-même une conséquence du choix particulier des R_i. Si nous n'avions pas choisi les R_i selon le critère optimal proposé au Chapitre 3, il n'y aurait pas eu l'annulation mentionnée précédemment et le α optimal aurait simplement été égal à 1/2, comme l'on peut s'y attendre a priori. Ceci souligne l'importance du choix des probabilités de transition proches des valeurs optimales et des nombres de retirage selon le critère optimal déjà mentionné. Ce résultat insiste aussi sur l'intérêt de mettre en place un algorithme adaptatif qui évaluerait à chaque étape les probabilités de transition et réajusterait les niveaux avant chaque phase.

CHAPITRE 6

APPLICATION AU PROCESSUS D'ORNSTEIN-UHLENBECK «MODIFIÉ»

Nous souhaitons étudier le processus appelé *processus d'Ornstein-Uhlenbeck «modifié»* et gouverné par l'équation différentielle stochastique suivante

$$dX_t = -\mu(X_t)X_t dt + \sigma^2 dB_t \tag{84}$$

où le drift inconnu est tel que

$$0 < \mu_1 \leq \mu(x) \leq \mu_2 < \infty.$$

Dans l'application numérique, nous prendrons un drift inconnu constant par morceaux. Rappelons que le processus d'Ornstein-Uhlenbeck général est gouverné par l'équation différentielle stochastique suivante

$$dX_t = -\mu X_t dt + \sigma^2 dB_t \tag{85}$$

Le processus d'Ornstein-Uhlenbeck est récurrent et ses excursions au dessus de grands niveaux sont rares. Il en est d'ailleurs de même du processus d'Ornstein-Uhlenbeck «modifié». Ici nous souhaitons seulement estimer les probabilités de franchissement de hauts niveaux partant d'un certain $x > 0$ et avant de retourner en 0. Lorsque le drift est connu, il est facile de calculer ces probabilités. Dans le cas contraire, on estime le drift mais cette estimation est très onéreuse en termes de simulation. Puisque ici nous cherchons seulement à estimer les probabilités de franchissement de hauts niveaux, il n'est pas nécessaire d'estimer μ et le modèle de branchement avec duplication des trajectoires semble être très prometteur pour résoudre le problème.

Avant de faire l'application numérique, rappelons les résultats généraux concernant le processus d'Ornstein-Uhlenbeck général.

6.1. Etude analytique du processus d'Ornstein-Uhlenbeck général

6.1.1. Introduction et premiers résultats. — Etudions le processus d'Ornstein-Uhlenbeck général gouverné par l'équation stochastique différentielle suivante :

$$dX_t = -\mu X_t dt + \sigma^2 dB_t \tag{86}$$

La mesure de vitesse m est donnée par

$$m(dx) = \rho(x)dx = e^{-\mu x^2/\sigma^2} dx \tag{87}$$

et le générateur infinitésimal $\mathcal{L}$ par

$$\mathcal{L}(f)(x) = f''(x) - \mu x f'(x) = f''(x) + \frac{\sigma^2}{2}\frac{\rho'(x)}{\rho(x)} f'(x) \tag{88}$$

Le processus d'Ornstein-Uhlenbeck standard est gouverné par l'équation stochastique différentielle suivante :

$$dX_t = -X_t dt + \sqrt{2} dB_t \tag{89}$$

(c-à-d le drift μ est égal à 1 et la variance σ^2 à 2).

Alors la mesure de vitesse m est donnée par

$$m(dx) = \rho(x)dx = e^{-x^2/2} dx \tag{90}$$

et le générateur infinitésimal $\mathcal{L}$ par

$$\mathcal{L}(f)(x) = f''(x) - x f'(x) = f''(x) + \frac{\rho'(x)}{\rho(x)} f'(x) \tag{91}$$

Dans la suite, nous étudierons le processus d'Ornstein-Uhlenbeck standard (drift $\mu = 1$ et variance $\sigma^2 = 2$). Les résultats pour $\mu \neq 1$ et $\sigma^2 \neq 2$ s'obtiennent aisément par changement de variables.

Soient a, x et b tels que $a < x < b$. Supposons que $X_0 = x$ et notons $H(x) = \int_a^x \frac{dy}{\rho(y)}$ et $T_{a,b}$ le premier temps de sortie de $[a, b]$:

$$T_{a,b} = \inf\{t \geq 0 : X_t \notin [a, b]\} \tag{92}$$

Il est immédiat que $T_{a,b}$ est un temps d'arrêt.

Maintenant, en utilisant la propriété de martingale de H, nous déduisons les probabilités d'atteindre b avant a partant de x et d'atteindre a avant b partant de x :

Proposition 6.1.1. — *Nous avons*

$$\begin{cases} \mathbb{P}_x(X_{T_{a,b}} = b) = \frac{H(x)}{H(b)} \\ \mathbb{P}_x(X_{T_{a,b}} = a) = \frac{H(b)-H(x)}{H(b)} \end{cases}$$

Démonstration. — D'abord, H est une solution de

$$\mathcal{L}(f) = 0 \tag{93}$$

En effet, nous avons $H'(x) = \frac{1}{\rho(x)}$ et $H''(x) = -\frac{\rho'(x)}{\rho(x)^2}$. Donc

$$H''(x) + \frac{\rho'(x)}{\rho(x)}H'(x) = -\frac{\rho'(x)}{\rho(x)^2} + \frac{\rho'(x)}{\rho(x)}\frac{1}{\rho(x)} = 0$$

Par conséquent, puisque $f(X_t) - \int_a^x \mathcal{L}f(X_s)ds$ est une martingale, $H(X_t)$ en est une aussi et par le théorème d'arrêt de Doob

$$\mathbb{E}(H(X_{T_{a,b}})) = \mathbb{E}(H(X_0)) = H(x)$$

Mais par définition de $T_{a,b}$,

$$\begin{cases} H(a)\mathbb{P}_x(X_{T_{a,b}} = a) + H(b)\mathbb{P}_x(X_{T_{a,b}} = b) = H(x) \\ \mathbb{P}_x(X_{T_{a,b}} = a) + \mathbb{P}_x(X_{T_{a,b}} = b) = 1 \end{cases}$$

d'où l'on déduit

$$\begin{cases} \mathbb{P}_x(X_{T_{a,b}} = b) = \frac{H(x)-H(a)}{H(b)-H(a)} \\ \mathbb{P}_x(X_{T_{a,b}} = a) = \frac{H(b)-H(x)}{H(b)-H(a)} \end{cases}$$

En notant que $H(a) = 0$, on obtient le résultat recherché. □

Soit R la primitive de ρ : $R(t) = \int_a^t \rho(s)ds$ et

$$M_{a,b}(x) = -\int_a^x \frac{R(t)}{\rho(t)}dt + \frac{\int_a^b \frac{R(t)}{\rho(t)}dt}{\int_a^b \frac{1}{\rho(t)}dt}\int_a^x \frac{1}{\rho(t)}dt$$

En suivant la même démarche, on déduit le temps moyen mis par une particule partant de x pour atteindre a ou b :

Proposition 6.1.2. — *Nous avons*

$$\mathbb{E}_x(T_{a,b}) = M_{a,b}(x) = -\int_a^x \frac{R(t)}{\rho(t)}dt + \frac{\int_a^b \frac{R(t)}{\rho(t)}dt}{\int_a^b \frac{1}{\rho(t)}dt}\int_a^x \frac{1}{\rho(t)}dt \tag{94}$$

Démonstration. — Tout d'abord, $M_{a,b}$ vérifie l'équation

(95) $$\mathcal{L}(f) = -1$$

En effet,

- $M_{a,b}(x) = -\int_a^x \frac{R(t)}{\rho(t)}dt + A\int_a^x \frac{1}{\rho(t)}dt,$
- $M'_{a,b}(x) = -\frac{R(x)}{\rho(x)} + A\frac{1}{\rho(x)},$
- $M''_{a,b}(x) = -1 + \frac{\rho'(x)}{\rho(x)^2}[R(x) + A].$

D'où

$$M''_{a,b}(x)+\frac{\rho'(x)}{\rho(x)}M'_{a,b}(x) = -1+\frac{\rho'(x)}{\rho(x)^2}R(x)+A\frac{\rho'(x)}{\rho(x)^2}-\frac{\rho'(x)}{\rho(x)^2}R(x)-\frac{\rho'(x)}{\rho(x)^2}A = -1$$

Remarquons aussi que $M_{a,b}(a) = M_{a,b}(b) = 0$.

Par conséquent, puisque $f(X_t) - \int_a^x \mathcal{L}f(X_s)ds$ est une martingale, $M_{a,b}(X_t) + (T_{a,b} - a)$ en est une aussi et par le théorème d'arrêt de Doob,

$$\mathbb{E}(M_{a,b}(X_{T_{a,b}}) + T_{a,b} - a) = \mathbb{E}(M_{a,b}(X_0) - a)$$

Mais $M_{a,b}(X_{T_{a,b}}) = M_{a,b}(a)$ ou $M_{a,b}(b)$ et par définition de $T_{a,b}$,

(96) $$\mathbb{E}_x(T_{a,b}) = M_{a,b}(x) = -\int_a^x \frac{R(t)}{\rho(t)}dt + \frac{\int_a^b \frac{R(t)}{\rho(t)}dt}{\int_a^b \frac{1}{\rho(t)}dt}\int_a^x \frac{1}{\rho(t)}dt$$

□

6.1.2. Recherche des seuils optimaux. — L'optimisation de l'algorithme indique que l'on doit prendre les probabilités de transition toutes égales c-à-d si l'on note les seuils b_i,

$$\mathbb{P}_{b_n}(X_{T_{a,b_{n+1}}} = b_{n+1}) = \frac{H(b_n)}{H(b_{n+1})} = \theta$$

pour un certain θ. Mais, pour de grandes valeurs de n,

$$\begin{aligned}
\mathbb{P}_{b_n}(X_{T_{a,b_{n+1}}} = b_{n+1}) = \theta \quad &\Leftrightarrow \quad H(b_n) = \theta H(b_{n+1}) \underset{n\to\infty}{\sim} \theta[H(b_n) + (b_{n+1} - b_n)H'(b_n)] \\
&\Leftrightarrow \quad H(b_n) \underset{n\to\infty}{\sim} \theta[H(b_n) + (b_{n+1} - b_n)e^{b_n^2/2}] \\
&\Leftrightarrow \quad b_{n+1} - b_n \underset{n\to\infty}{\sim} e^{-b_n^2/2}H(b_n)(\frac{1}{\theta} - 1) \\
&\Leftrightarrow \quad b_{n+1} - b_n \underset{n\to\infty}{\sim} \frac{1}{b_n}(\frac{1}{\theta} - 1)
\end{aligned}$$

puisque $\int_a^x e^{y^2/2}dy \underset{x\to\infty}{\sim} \frac{e^{x^2/2}}{x}$. En effet, remarquons que

$$e^{y^2/2} \underset{y\to\infty}{\sim} e^{y^2/2}(1-\frac{1}{y^2}) = \left(\frac{e^{y^2/2}}{y}\right)'$$

Comme $\int_a^{+\infty} e^{y^2/2}dy$ diverge, on obtient que

$$\int_a^x e^{y^2/2}dy \underset{x\to\infty}{\sim} \int_a^x e^{y^2/2}(1-\frac{1}{y^2})dy = \frac{e^{x^2/2}}{x} - \frac{e^{a^2/2}}{a} \underset{x\to\infty}{\sim} \frac{e^{x^2/2}}{x}$$

Si par exemple, on veut que $\theta = 1/2$, le calcul précédent suggère de prendre

$$b_{n+1} = b_n + \frac{1}{b_n}$$

Recherchons maintenant un équivalent de b_n. Nous allons en réalité montrer le résultat plus précis suivant :

Lemme 6.1.1. — *Soit* (u_n) *une suite réelle satisfaisant*

$$u_0 > 0 \quad et \quad \forall n \in \mathbb{N}, \quad u_{n+1} = u_n + \frac{1}{u_n^\alpha}$$

avec $\alpha > -1$. *Alors*

$$u_n \underset{n\to\infty}{\sim} [n(1+\alpha)]^{\frac{1}{\alpha+1}}$$

Démonstration. — Tout d'abord, on remarque que (u_n) diverge pour $n \to \infty$. En effet, comme $u_0 > 0$, (u_n) est croissante et ne peut être majorée, sinon elle convergerait et sa limite vérifierait $l = l + \frac{1}{l^\alpha}$ ce qui est absurde.

Ensuite, on cherche s'il existe un $\beta > 0$ tel que $(u_{n+1}^\beta - u_n^\beta)$ converge. Comme $\alpha > -1$ et (u_n) diverge,

$$\begin{aligned} u_{n+1}^\beta &= \left\{u_n + \frac{1}{u_n^\alpha}\right\}^\beta = u_n^\beta \left(1 + \frac{1}{u_n^{\alpha+1}}\right)^\beta \\ &= u_n^\beta \left(1 + \beta\frac{1}{u_n^{\alpha+1}} + o\left(\frac{1}{u_n^{\alpha+1}}\right)\right) \end{aligned}$$

D'où $\quad u_{n+1}^\beta - u_n^\beta \underset{n\to\infty}{\sim} \beta\, u_n^{\beta-(1+\alpha)}$, on choisit donc

$$\beta = 1 + \alpha$$

En sommant ces équivalents, puisque $\sum(\alpha+1)$ diverge,

$$u_n^{\alpha+1} - u_0^{\alpha+1} = \sum_{k=0}^{n-1}\left(u_{k+1}^{\alpha+1} - u_k^{\alpha+1}\right) \underset{n\to\infty}{\sim} \sum_{k=0}^{n-1}(\alpha+1) = n\,(\alpha+1)$$

Ainsi $u_n^{\alpha+1} \sim n(\alpha+1)$ et $u_n \underset{n\to\infty}{\sim} [n(1+\alpha)]^{\frac{1}{\alpha+1}}$. □

Et on déduit la proposition suivante en prenant $\alpha = 1$:

Proposition 6.1.3. — *Choisir les probabilités de transition constantes et égales à $\theta = \frac{1}{2}$ revient à prendre*

$$b_n \underset{n\to\infty}{\sim} \sqrt{2n}$$

6.1.3. Coût de transition. — La proposition suivante donne un équivalent asymptotique du coût de transition c-à-d le temps moyen mis par une particule partant de x pour atteindre 0 ou le seuil suivant $x + \frac{1}{x}$:

Proposition 6.1.4. — *On a*

$$\mathbb{E}_x(T_{0,x+\frac{1}{x}}) \underset{x\to+\infty}{\sim} (1-e^{-1})\log(x)$$

Remarque 6.1.1. — *La proposition précédente justifie l'hypothèse qu'asymptotiquement les coûts de transition sont égaux.*

Démonstration. — On a tout d'abord

$$\mathbb{E}_x(T_{0,x+\frac{1}{x}}) = -\int_0^x \frac{R(t)}{\rho(t)}dt + \frac{\int_0^{x+\frac{1}{x}} \frac{R(t)}{\rho(t)}dt}{\int_0^{x+\frac{1}{x}} \frac{1}{\rho(t)}dt} \int_0^x \frac{1}{\rho(t)}dt \tag{97}$$

Ecrivons $R(t) = R(\infty) - D(t)$ avec $D(t) = \int_t^\infty \rho(u)du$; en multipliant de chaque côté par $\int_0^{x+\frac{1}{x}} \frac{dt}{\rho(t)}$,

$$\begin{aligned}
\mathbb{E}_x(T_{0,x+\frac{1}{x}}) \int_0^{x+\frac{1}{x}} \frac{dt}{\rho(t)} &= \int_0^x \frac{dt}{\rho(t)} \int_0^{x+\frac{1}{x}} \frac{R(t)}{\rho(t)}dt - \int_0^{x+\frac{1}{x}} \frac{dt}{\rho(t)} \int_0^x \frac{D(t)}{\rho(t)}dt \\
&= -\int_0^x \frac{dt}{\rho(t)} \int_x^{x+\frac{1}{x}} \frac{D(t)}{\rho(t)}dt + \int_x^{x+\frac{1}{x}} \frac{dt}{\rho(t)} \int_0^x \frac{D(t)}{\rho(t)}dt
\end{aligned}$$

Mais
– $\int_0^x e^{t^2/2}dt \underset{x\to+\infty}{\sim} \frac{e^{x^2/2}}{x}$

- $\int_0^{x+\frac{1}{x}} e^{t^2/2}dt \underset{x\to+\infty}{\sim} \frac{e^{(x+\frac{1}{x})^2/2}}{x+\frac{1}{x}} \underset{x\to+\infty}{\sim} e\frac{e^{x^2/2}}{x}$
- $\int_x^{x+\frac{1}{x}} e^{t^2/2}dt \underset{x\to+\infty}{\sim} (e-1)\frac{e^{x^2/2}}{x}$
- $\int_0^x D(t)e^{t^2/2}dt \underset{x\to+\infty}{\sim} \log x$
- $\int_x^{x+\frac{1}{x}} D(t)e^{t^2/2}dt \underset{x\to+\infty}{\sim} \frac{1}{x^2}$

D'où le résultat. Recherchons maintenant les équivalents utilisés.

• $\int_0^x e^{t^2/2}dt \underset{x\to+\infty}{\sim} \frac{e^{x^2/2}}{x}$

Cet équivalent a déjà été démontré dans la partie précédente.

• $\int_0^{x+\frac{1}{x}} e^{t^2/2}dt \underset{x\to+\infty}{\sim} \frac{e^{(x+\frac{1}{x})^2/2}}{x+\frac{1}{x}} \underset{x\to+\infty}{\sim} e\frac{e^{x^2/2}}{x}$

On en déduit que $\int_0^{x+\frac{1}{x}} e^{t^2/2}dt \underset{x\to+\infty}{\sim} \frac{e^{(x+\frac{1}{x})^2/2}}{x+\frac{1}{x}} \underset{x\to+\infty}{\sim} \frac{e^{(x+\frac{1}{x})^2/2}}{x}$.

Mais

$$\lim_{x\to+\infty} \frac{e^{(x+\frac{1}{x})^2/2}}{e^{x^2/2}} = \lim_{x\to+\infty} e^{1+\frac{1}{2x^2}} = e$$

D'où l'équivalent recherché.

• $\int_x^{x+\frac{1}{x}} e^{t^2/2}dt \underset{x\to+\infty}{\sim} (e-1)\frac{e^{x^2/2}}{x}$

On en déduit que

$$\int_x^{x+\frac{1}{x}} e^{t^2/2}dt = \int_0^{x+\frac{1}{x}} e^{t^2/2}dt - \int_0^x e^{t^2/2}dt \underset{x\to+\infty}{\sim} e\frac{e^{x^2/2}}{x} - \frac{e^{x^2/2}}{x}$$

D'où le résultat annoncé.

• $\int_0^x D(t)e^{t^2/2}dt \underset{x\to+\infty}{\sim} \log x$

De la même façon que pour le premier équivalent, on montre que

$$D(t) = \int_t^\infty e^{-u^2/2}du \underset{t\to\infty}{\sim} \frac{e^{-t^2/2}}{t} \quad \text{ainsi} \quad D(t)e^{t^2/2} \underset{t\to\infty}{\sim} \frac{1}{t}$$

et par conséquent,

$$\int_a^x D(t)e^{t^2/2}dt \underset{x\to+\infty}{\sim} \int_a^x \frac{1}{t}dt = \log(x/a) \underset{t\to\infty}{\sim} \log(x)$$

et

$$\int_0^x D(t)e^{t^2/2}dt = \int_a^x D(t)e^{t^2/2}dt + \int_0^a D(t)e^{t^2/2}dt \underset{x\to+\infty}{\sim} \log(x)$$

- $\int_x^{x+\frac{1}{x}} D(t)e^{t^2/2}dt \underset{x\to+\infty}{\sim} \frac{1}{x^2}$

On en déduit que

$$\int_x^{x+\frac{1}{x}} D(t)e^{t^2/2}dt = \int_0^{x+\frac{1}{x}} D(t)e^{t^2/2}dt - \int_0^x D(t)e^{t^2/2}dt \underset{x\to+\infty}{\sim} \log(x+\frac{1}{x}) - \log(x)$$

Ainsi

$$\int_x^{x+\frac{1}{x}} D(t)e^{t^2/2}dt \underset{x\to+\infty}{\sim} \frac{1}{x^2}$$

$\square$

6.1.4. Généralisation au processus d'Ornstein-Uhlenbeck général. — Par changement de variables, on obtient le même type de résultats pour le processus d'Ornstein-Uhlenbeck général (c-à-d $\mu \neq 1$ et $\sigma^2 \neq 2$) régi par l'équation différentielle :

$$dX_t = -\mu X_t dt + \sigma dB_t \tag{98}$$

Les probabilités de transition sont données par

$$\begin{cases} \mathbb{P}_x(X_{T_{a,b}} = b) = \frac{H(x)}{H(b)} \\ \mathbb{P}_x(X_{T_{a,b}} = a) = \frac{H(b)-H(x)}{H(b)} \end{cases}$$

où

$$H(x) \underset{x\to+\infty}{\sim} \frac{e^{\mu x^2/\sigma^2}}{x}\frac{\sigma^2}{2\mu}$$

Les seuils optimaux asymptotiques sont donnés par le même genre de relation

$$b_{i+1} = b_i + \frac{\sigma^2}{2\mu}\frac{1}{b_i}$$

Enfin le coût de transition est asymptotiquement équivalent à

$$\mathbb{E}_x(T_{0,x+\frac{1}{x}}) \underset{x\to+\infty}{\sim} \frac{\sigma^2}{2\mu}(1-e^{-2\mu/\sigma^2})\log(x)$$

6.2. Application au processus d'Ornstein-Uhlenbeck général

On suppose que le drift inconnu est constant et appartient à l'intervalle $[\mu_i, \mu_s]$. On détermine arbitrairement un drift moyen constant $\mu_0 = \frac{\mu_i+\mu_s}{2}$. On recherche ensuite les seuils $B(i)$ tels que les probabilités de transition soient égales à un certain p pour ce drift moyen et les nombres de retirage optimaux pour ce drift moyen. On fait évoluer ρ_N processus selon l'équation

$$dX_t = -\mu X_t dt + \sigma dB_t \tag{99}$$

que l'on discrétise en

$$X_{n+h} = -(\mu h - 1)X_n + \sigma\sqrt{h}N$$

où h est le pas de la discrétisation et N suit une loi $\mathcal{N}(0,1)$.

On obtient ainsi des estimations des probabilités de transition qui nous permettent d'estimer les nombres de retirage optimaux. On fait ensuite évoluer les $N-\rho_N$ processus restants selon la discrétisation de l'équation (99) avec les nombres de retirage estimés.

Nous comparons aussi les estimations obtenues avec le modèle de branchement avec duplication des trajectoires en deux étapes avec les résultats obtenus en utilisant les algorithmes présentés en Sections 5.3 et 5.4. Nous avons tracé en Figures 1, 2 et 3 le principe de ces trois différents algorithmes.

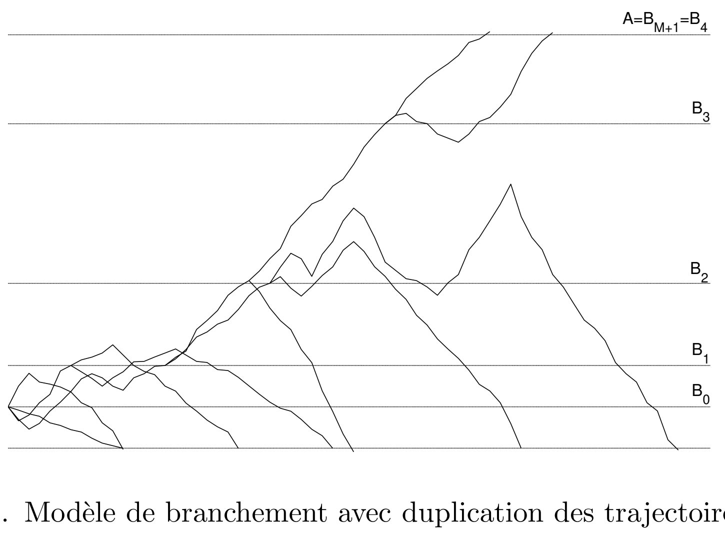


FIG. 1. Modèle de branchement avec duplication des trajectoires R=2

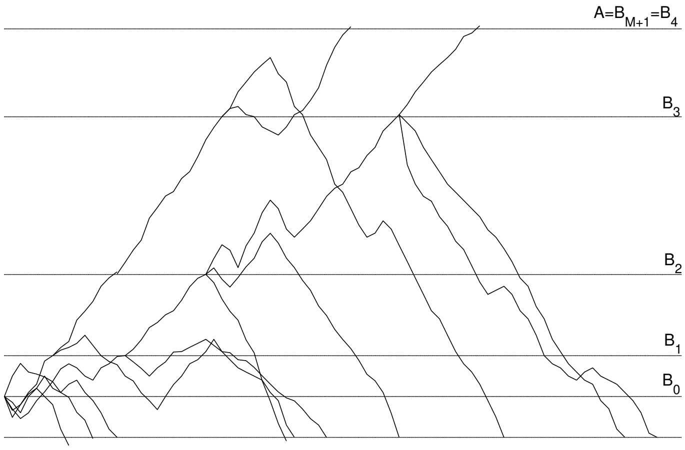


FIG. 2. Algorithme séquentiel de LeGland-Oudjane H=2

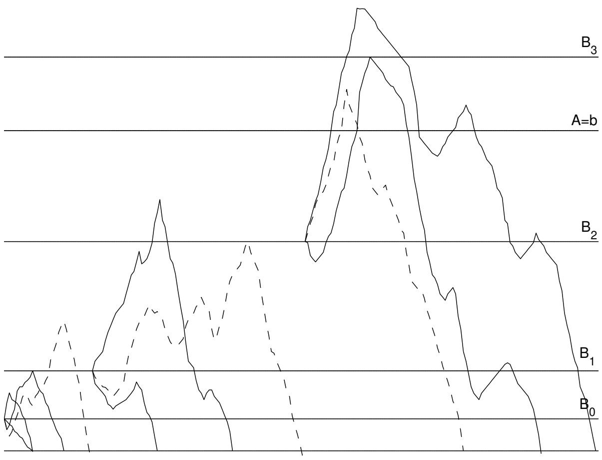


FIG. 3. Algorithme adaptatif de Cérou-Guyader p=1/2

Application : Nous souhaitons étudier le processus d'Ornstein-Uhlenbeck gouverné par l'équation (99) où l'écart-type σ vaut 0.3 et le drift μ appartient à $[0.07, 0.11]$.Plus précisément, nous souhaitons estimer la probablité que ce processus partant de $x = 0.1$ atteigne $b = 4$ avant de retourner en $a = 0$. En prenant un drift μ égal à 0.1, cela correspond à une probabilité cible

$$\mathbb{P}_x(X_{T_{a,b}} = b) = 1.64604e - 08$$

Fixons le drift moyen μ_0 égal à 0.09. Les seuils correspondant à ce drift moyen et à une probabilité de transition $p = 0.2$ sont donnés par

$$\begin{cases} B_1 = 0.465697364; \quad B_2 = 1.278877035; \quad B_3 = 1.915817876; \\ B_4 = 2.358424859; \quad B_5 = 2.711218626; \quad B_6 = 3.014546687; \\ B_7 = 3.285699406; \quad B_8 = 3.533621025; \quad B_9 = 3.763649017; \\ B_{10} = 3.979295907; \quad B_{11} = 4; \end{cases}$$

Le tableau ci-dessous contient les résultats des différentes simulations pour $N = 10^5$ vecteurs générés au départ et 50 itérations. Les abréviations BaDdT2, ASLO, AACG signifient respectivement Branchement avec Duplication des Trajectoires en deux étapes, Algorithme Séquentiel détaillé en Section 5.3 et introduit par F.LeGland et N.Oudjane et Algorithme Adaptatif détaillé en Section 5.4 et introduit par F.Cérou et A.Guyader.

	Estimation	Erreur	Lg IC à 95%	Coût simul.
BaDdT2	1.68446e-08	3.84209e-10	1.80043e-09	6.4e+07
ASLO $H = 10^2$	1.51638e-08	1.29660e-09	2.38012e-09	8.7e+07
ASLO $H = 10^3$	1.68404e-08	3.79944e-10	8.79072e-10	8.5e+08
ASLO $H = 10^4$	1.66723e-08	3.05484e-10	2.11908e-10	9.2e+08
AACG $p = 1/3$	–	–	–	–
AACG $p = 1/2$	–	–	–	–
AACG $p = 2/3$	–	–	–	–
AACG $p = 3/4$	–	–	–	–

Malheureusement, l'algorithme adaptatif de Cérou et Guyader ne fournit pas de résultat : le nombre d'échantillons semble être trop élévé et le programme tourne sans fin...Par contre, pour $N = 10^4$, on obtient

	Estimation	Erreur	Lg IC à 95%	Coût simul.
AACG $p = 1/3$	1.02153e-08	6.24517e-09	2.92397e-10	1.3e+09
AACG $p = 1/2$	1.04803e-08	5.98009e-09	2.86177e-10	1.4e+09
AACG $p = 2/3$	1.10764e-08	5.38396e-09	3.93731e-10	1.9e+09
AACG $p = 3/4$	1.20356e-08	4.42486e-09	3.46029e-10	2.1e+09

6.3. Etude analytique du processus d'Ornstein-Uhlenbeck «modifié»

Rappelons que nous considérons ici un processus appelé processus d'Ornstein-Uhlenbeck «modifié» et gouverné par l'équation

$$dX_t = -\mu X_t dt + \sigma dB_t \tag{100}$$

où le drift μ est constant par morceaux et vaut μ_i sur l'intervalle $[m_{i-1}, m_i]$, $i = 1, 2, \cdots$. Par convention, nous prendrons $m_0 = 0$.

La mesure de vitesse m constante par morceaux est donnée par

$$m(dx) = \rho(x)dx = e^{-\mu_i x^2/\sigma^2} dx := \rho_i(x)dx \quad \text{sur} \quad [m_{i-1}, m_i] \tag{101}$$

et le générateur infinitésimal $\mathcal{L}$ par

$$\mathcal{L}(f)(x) = f''(x) - \mu_i x f'(x) \quad \text{sur} \quad [m_{i-1}, m_i] \tag{102}$$

Soient maintenant a, x et b tels que $a < x < b$. Supposons que $X_0 = x$ et notons comme précédemment $T_{a,b}$ le premier temps de sortie de $[a, b]$:

$$T_{a,b} = \inf\{t \geq 0 : X_t \notin [a, b]\} \tag{103}$$

Il est immédiat que $T_{a,b}$ est un temps d'arrêt.

Nous procédons ensuite comme pour le processus d'Ornstein-Uhlenbeck général et cherchons une solution à l'équation

$$\mathcal{L}(f) = 0 \tag{104}$$

Proposition 6.3.1. — *Définissons H de la façon suivante :*
pour $z \in [m_{i-1}, m_i]$,

$$H(z) := \beta \sum_{k=1}^{i-1} \prod_{j=1}^{k-1} \frac{\rho_j(m_j)}{\rho_{j+1}(m_j)} \int_{m_{k-1}}^{m_k} \rho_k(y)dy + \beta \prod_{j=1}^{i-1} \frac{\rho_j(m_j)}{\rho_{j+1}(m_j)} \int_{m_{i-1}}^{z} \rho_k(y)dy$$

Alors, pour tout β, H est $\mathcal{C}^1$ et solution de (104).

Démonstration. — De façon à avoir la continuité, on définit H par

$$H(x) = \begin{cases} S_1(x) := \beta_1 \int_0^x \rho_1(y)dy & \text{sur} \quad [0, \mu_1] \\ S_2(x) := S_1(\mu_1) + \beta_2 \int_{\mu_1}^x \rho_1(y)dy & \text{sur} \quad [\mu_1, \mu_2] \\ S_3(x) := S_2(\mu_2) + \beta_3 \int_{\mu_2}^x \rho_1(y)dy & \text{sur} \quad [\mu_2, \mu_3] \\ S_4(x) := S_3(\mu_3) + \beta_4 \int_{\mu_3}^x \rho_1(y)dy & \text{sur} \quad [\mu_3, \mu_4] \\ \cdots \end{cases}$$

Puis on ajuste les constantes β_i pour avoir la continuité de la dérivée première et on est conduit immédiatement à la solution de la proposition. □

Maintenant, en utilisant la propriété de martingale de H, nous déduisons de la même façon que pour le processus d'Ornstein-Uhlenbeck général les probabilités d'atteindre b avant a partant de x et d'atteindre a avant b partant de x :

Proposition 6.3.2. — *Nous avons*

$$\begin{cases} \mathbb{P}_x(X_{T_{a,b}} = b) = \frac{H(x)}{H(b)} \\ \mathbb{P}_x(X_{T_{a,b}} = a) = \frac{H(b)-H(x)}{H(b)} \end{cases}$$

Nous obtenons ainsi une expression analytique mais non explicite de la probabilité recherchée.

6.4. Application au processus d'Ornstein-Uhlenbeck «modifié»

On suppose que le drift inconnu est constant par morceaux et que sur chaque intervalle i

$$\mu \in [\mu_1^i, \mu_2^i].$$

De la même façon que précédemment, on détermine arbitrairement un drift moyen constant par morceau en prenant sur chaque intervalle i $\mu_0 = \frac{\mu_1^i+\mu_2^i}{2}$. On recherche ensuite les seuils B_i tels que les probabilités de transition soient égales à un certain p pour ce drift moyen et les nombres de retirage optimaux pour ce drift moyen. On fait évoluer ρ_N processus selon l'équation

$$dX_t = -\mu X_t dt + \sigma dB_t \tag{105}$$

que l'on discrétise en

$$X_{n+h} = -(\mu h - 1)X_n + \sigma\sqrt{h}N$$

où h est le pas de la discrétisation et N suit une loi $\mathcal{N}(0,1)$.

On obtient ainsi des estimations des probabilités de transition qui nous permettent d'estimer les nombres de retirage optimaux. On fait ensuite évoluer les $N - \rho_N$ processus restants selon la discrétisation de l'équation (100) avec les nombres de retirage estimés.

Nous comparons aussi les estimations obtenues avec le modèle de branchement avec duplication des trajectoires en deux étapes avec les résultats obtenus en utilisant les algorithmes présentés en Sections 5.3 et 5.4.

Remarque 6.4.1. — *Procédure pour la simulation*
Comme le drift est constant par morceaux, un léger biais est introduit à chaque changement de régime pour le drift. Nous proposons ci-dessous une procédure afin de diminuer ce biais. Notons $(X_n)_n$ le processus étudié.

Soit X_n tel que $m_{i-1} < X_n < m_i$, alors X_n va évoluer selon l'équation (100) avec le drift μ_i et

$$X_{n+1} = X_n - \mu_i h X_n + \sigma\sqrt{h}Z$$

où Z est une réalisation de la loi $\mathcal{N}(0,1)$.

Si $X_{n+1} < m_i$, on poursuit la simulation. Si par contre, $X_{n+1} > m_i$, on résout l'équation du second degré en X_{n+1}

$$X_{n+1} = X_n - \frac{\mu_i(m_i - X_n) + \mu_{i+1}(X_{n+1} - m_i)}{X_{n+1} - X_n} h X_n + \sigma\sqrt{h}Z$$

Ce qui revient à actualiser le drift a priori. Finalement, on prend pour X_{n+1} la solution obtenue.

Application : Nous souhaitons étudier le processus «approché» d'Ornstein-Uhlenbeck gouverné par l'équation (100) où l'écart-type σ vaut 0.3 et le drift μ est tel que

$$\mu \in \begin{cases} [0.05, 0.065] & \text{sur } [0, 1/2] \\ [0.06, 0.080] & \text{sur } [1/2, 1] \\ [0.07, 0.095] & \text{sur } [1, 3/2] \\ [0.08, 0.110] & \text{sur } [3/2, 2] \\ [0.09, 0.125] & \text{sur } [2, 5/2] \\ [0.10, 0.140] & \text{sur } [5/2, 3] \end{cases}$$

Plus précisément, nous souhaitons estimer la probablité que ce processus partant de $x = 0.1$ atteigne $b = 3$ avant de retourner en $a = 0$. En prenant un drift μ et un drift moyen égaux à

$$\mu = \begin{cases} 0.064 & \text{sur} \quad [0, 1/2] \\ 0.061 & \text{sur} \quad [1/2, 1] \\ 0.071 & \text{sur} \quad [1, 3/2] \\ 0.081 & \text{sur} \quad [3/2, 2] \\ 0.091 & \text{sur} \quad [2, 5/2] \\ 0.101 & \text{sur} \quad [5/2, 3] \end{cases} \quad \text{et} \quad \mu_0 = \begin{cases} 0.0575 & \text{sur} \quad [0, 1/2] \\ 0.0700 & \text{sur} \quad [1/2, 1] \\ 0.8250 & \text{sur} \quad [1, 3/2] \\ 0.0950 & \text{sur} \quad [3/2, 2] \\ 0.1075 & \text{sur} \quad [2, 5/2] \\ 0.1200 & \text{sur} \quad [5/2, 3] \end{cases}$$

cela correspond à une probabilité cible :

$$\mathbb{P}_x(X_{T_{a,b}} = b) = 1.15227e - 04$$

Les seuils correspondant à ce drift moyen et à une probabilité de transition $p = 0.2$ sont donnés par

$$\begin{cases} B_1 = 0.476922045; & B_2 = 1.377658754; \\ B_3 = 2.003754081; & B_4 = 2.383557180; \\ B_5 = 2.670334726; & B_6 = 2.905049847; \\ B_7 = 3; \end{cases}$$

Le tableau ci-dessous contient les résultats des différentes simulations pour $N = 10^4$ vecteurs générés au départ et 50 itérations. Les abréviations BaDdT1, BaDdT2, BaDdT2 am, ASLO, AACG signifient respectivement Branchement avec Duplication des Trajectoires classique, Branchement avec Duplication des Trajectoires en deux étapes, Branchement avec Duplication des Trajectoires en deux étapes amélioré grace à la Remarque 6.4.1, Algorithme Séquentiel détaillé en Section 5.3 et introduit par F.LeGland et N.Oudjane et Algorithme Adaptatif détaillé en Section 5.4 et introduit par F.Cérou et A.Guyader.

	Estimation	Erreur	Lg IC à 95%	Coût simul.
BaDdT2 am	1.12162e-04	3.06565e-06	5.69956e-06	1.4e+08
BaDdT2	1.11484e-04	3.74433e-06	5.35379e-06	1.4e+08
BaDdT1	1.16635e-04	1.40752e-06	2.40927e-06	1.5e+09
ASLO $H = 10^2$	1.29189e-04	1.39611e-05	2.06204e-05	1.0e+09
ASLO $H = 10^3$	1.19470e-04	4.24243e-06	5.33502e-06	1.1e+09
ASLO $H = 10^4$	1.17517e-04	2.28947e-06	1.78598e-06	1.2e+09
AACG $p = 1/3$	1.46799e-04	3.15714e-05	1.78346e-06	8.6e+08
AACG $p = 1/2$	1.47946e-04	3.27184e-05	2.52657e-06	1.0e+09
AACG $p = 2/3$	1.53858e-04	3.86298e-05	2.98097e-06	1.2e+09
AACG $p = 3/4$	1.57391e-04	4.21629e-05	2.95576e-06	1.3e+09

6.5. Conclusion

Ainsi le modèle de branchement avec duplication des trajectoires en deux étapes permet d'obtenir des résultats précis pour un coût de simulation moindre par rapport aux autres algorithmes présentés.

CHAPITRE 7

CONCLUSION

Modèle de branchement avec duplication des trajectoires

Nous avons présenté, dans cette thèse, le modèle de branchement avec duplication des trajectoires, *branching splitting model* en anglais. Nous avons ensuite optimisé l'algorithme en minimisant la variance de l'estimateur à coût de simulation fixé et ainsi dégagé les paramètres optimaux (probabilités de transition, nombres de retirage, de particules générées et enfin de seuils). Il est apparu que les nombres de retirage devaient égaler l'inverse d'une certaine valeur P_0, valeur commune des probabilités de transition optimales. Ce résultat n'est pas surprenant puisqu'il signifie que les processus de branchement étudiés sont des Galton-Watson critiques et traduit un certain équilibre entre un nombre de répliques trop important qui entraînerait l'explosion du coût et un nombre de répliques insuffisant conduisant à un arrêt prématuré de l'algorithme.

La simplicité relative du modèle nous a ensuite permis d'établir des résultats explicites comme des bornes de Chernoff de l'erreur relative entre la probabilité d'intérêt $\mathbb{P}(A)$ et son estimateur $\widehat{P}$. Un des défis essentiels pour l'obtention des bornes de Chernoff était de déterminer l'itérée $M+1$-ième d'une certaine fonction ψ dont l'expression, quoique simple, ne permettait pas de donner une forme explicite de son itérée $M+1$-ième. En pratique, le nombre optimal de niveaux M n'étant pas très grand, des estimées asymptotiques auraient été imprécises et une approche numérique s'avérait inefficace du fait que M n'est connu qu'après optimisation de ces bornes. C'est pourquoi, nous avons cherché des estimées précises de ψ_{M+1} et d'abord de ψ. Mais au lieu d'utiliser une seule fonction pour borner ψ, nous en avons utilisé plusieurs afin d'obtenir des bornes supérieures et inférieures plus précises. Pour cela, nous avons fait appel aux groupes de Lie de faible dimension des fonctions homographiques et des fonctions affines. Rappelons que plus la dimension du groupe de Lie est grande, plus précise est l'approximation, puisque la dimension décrit le nombre de paramètres à ajuster. Malheureusement, nous n'avons pas trouvé de groupe de Lie de plus grande dimension ayant la propriété de monotonicité requise pour itérer directement les inégalités obtenues pour ψ. L'intérêt d'utiliser de telles fonctions réside aussi dans le fait que leurs itérées ont une forme explicite. En utilisant cette technique, nous avons pu dériver des bornes précises de la probabilité de l'erreur relative $\mathbb{P}(A)$ et son estimateur.

Nous avons ensuite étudié la sensibilité des bornes de Chernoff en fonction du choix des nombres de retirage R dans trois algorithmes différents, tous trois basés sur le modèle de branchement avec duplication des trajectoires : il apparaît alors que, lorsqu'on ne peut être exactement dans le cas critique (qui correspond à l'algorithme optimal), la meilleure façon de procéder est de considérer R comme une variable aléatoire que l'on génère à chaque succès au cours de la simulation. Cette procédure est d'ailleurs couramment utilisée en pratique, voir par exemple [3].

Au vu des résultats et des simulations, ce modèle semble finalement être très prometteur, facile à implémenter et applicable à une large classe de systèmes.

En pratique, les probabilités de transition sont généralement inconnues et on ne peut donc pas déterminer les nombres de retirage. L'ajustement des niveaux peut alors se faire lors d'une première phase d'apprentissage, ainsi que le choix des nombres de retirage. Nous avons proposé, dans cette thèse, un modèle en deux étapes basé sur le modèle de branchement avec duplication des trajectoires afin de résoudre ce problème. Une analyse précise a montré que nous devons consacrer, asymptotiquement lorsque C tend vers l'infini, $\mu_s C^{2/3}$ particules à la phase d'apprentissage contre $C/C_{opt} - \mu_s C^{2/3}$ à la seconde phase, pour un certain μ_s déduit de l'optimisation de l'algorithme et ayant une forme explicite ; c-à-d que si on suppose que le nombre de particules générées dans la phase d'apprentissage a asymptotiquement la forme $\mu_\alpha(C)C^{1-\alpha}$, l'analyse conduit à $\alpha = 1/3$ et non à $\alpha = 1/2$, comme l'on pouvait s'attendre a priori.

Ce résultat provient directement de l'annulation observée pendant l'optimisation qui est elle-même une conséquence du choix particulier des R_i. Ceci souligne l'importance du choix des probabilités de transition proches des valeurs optimales et des nombres de retirage selon le critère optimal déjà mentionné. Ce résultat insiste aussi sur l'intérêt de mettre en place un algorithme adaptatif qui évalue à chaque étape les probabilités de transition et réajuste les niveaux avant chaque phase. Plus précisément, on peut imaginer un algorithme en plusieurs phases procédant comme suit : nous considérons comme précédemment que les probabilités de transition sont inconnues mais appartiennent à $[a,b]$. Choisissons arbitrairement $M+1$ nombres $\widetilde{P}_i^{(0)}$ dans $[a,b]$, posons

$$\widetilde{R}_i^{(0)} = \frac{1}{\sqrt{\widetilde{P}_i^{(0)}\widetilde{P}_{i+1}^{(0)}}}\sqrt{\frac{1-\widetilde{P}_{i+1}^{(0)}}{1-\widetilde{P}_i^{(0)}}} \quad \text{pour} \quad i = 1\cdots M,$$

et plaçons les niveaux de telle sorte que les probabilités de transition soient les $\widetilde{P}_i^{(0)}$, $i = 1\cdots M+1$.

Nous générons les particules une à une : générons tout d'abord une particule. Si elle atteint B_1, nous la dupliquons en $\widetilde{R}_1^{(0)}$ sous-particules que nous faisons évoluer depuis B_1. Dans le cas contraire, nous générons une autre particule. Les particules ayant atteint B_2 sont dupliquées en $\widetilde{R}_2^{(0)}$ sous-particules et nous les faisons évoluer depuis B_2... Et ainsi de suite jusqu'à ce que A soit atteint. De l'étape i ($i = 1\cdots M+1$), nous obtenons un estimateur $\widehat{P}_i^{(1)}$ de P_i (fraction des particules ayant réussi à atteindre B_i depuis B_{i-1}). Pour utiliser toute l'information dont nous disposons, nous améliorons l'algorithme en remplaçant $\widehat{P}_i^{(1)}$ par $\widetilde{P}_i^{(1)} := a \vee \widehat{P}_i^{(1)} \wedge b$ pendant la simulation, puisque $P_i \in [a,b]$. Cette substitution nous prémunit de tout arrêt prématuré de l'algorithme. Maintenant, pour tout $i = 1\cdots M$, posons

$$\widetilde{R}_i^{(1)} = \frac{1}{\sqrt{\widetilde{P}_i^{(1)}\widetilde{P}_{i+1}^{(1)}}}\sqrt{\frac{1-\widetilde{P}_{i+1}^{(1)}}{1-\widetilde{P}_i^{(1)}}}.$$

Réajustons les niveaux afin que les probabilités de transition soient les $\widetilde{P}_i^{(j-1)}$, $i = 1\cdots M+1$. Nous continuons ensuite l'algorithme en générant les particules une à une, en réactualisant les estimations des probabilités de transition ainsi que celles des nombres de retirage et en repositionnant les niveaux selon ces nouvelles estimées après l'évolution de chaque particule. On continue jusqu'à ce que le budget soit épuisé ou jusqu'à ce qu'un critère d'arrêt basé par exemple sur la précision des estimées soit atteint.

Au vu des résultats du chapitre précédent, on imagine que cette méthode est nécessairement meilleure. Mais une analyse très fine de tels algorithmes à coût donné est sans doute infiniment plus compliquée à mener. De plus, un tel algorithme suppose qu'il est possible de placer les niveaux comme on le souhaite ; ce qui est le cas par exemple pour les diffusions et en particulier les processus d'Ornstein-Uhlenbeck.

Rappelons que le modèle de branchement avec duplication des particules est un modèle simple pour lequel nous avons pu mener une étude analytique précise et complète, destinée à nous apporter une meilleure compréhension du modèle lui-même mais aussi des modèles plus complexes tels que les algorithmes particulaires, le modèle RESTART auxquels il est intimement lié. Il serait donc intéressant de mener maintenant le même type d'étude pour ces algorithmes plus complexes : optimisation du modèle sous une contrainte de coût, étude des bornes de Chernoff... Dans de tels modèles, l'hypothèse que les processus sont markoviens est généralement faite (garantissant dans notre cas, l'indépendance entre les franchissements de seuils). On pourrait donc aussi s'intéresser aux modèles non-markoviens.

Généralisation à des systèmes multidimensionnels

Nous nous sommes restreints, dans cette thèse, à des systèmes unidimensionnels. Il serait intéressant de généraliser et d'implémenter cet algorithme en dimensions supérieures. Rappelons le contexte général des modèles multiniveaux : on considère un processus de Markov $X := (X_t,\ t \geq 0)$ sur l'espace d'état E et f une fonction mesurable de E dans $\mathbb{R}$, associant à tout état x de E une valeur réelle à laquelle nous faisons référence en l'appelant «niveau». Plus précisément, on définit le processus Z par $Z_t := f(X_t)$, pour tout $t \geq 0$ et l'événement rare A sera de la forme $A := \{x \in E | f(x) \geq L\}$ pour un certain «niveau» ou «seuil» L. La fonction f est appelée *fonction d'importance* et renvoie le niveau de l'état X_t au temps t grâce à $Z_t = f(X_t)$. L'événement A correspond alors aux plus grandes valeurs de la fonction d'importance. Puisque l'espace d'état du processus est complètement caché dans la fonction f, on peut ramener tout processus markovien X au graphe de $Z_t = f(X_t)$ en fonction de t. Ainsi l'évolution du processus est maintenant très similaire à un modèle de dépassement d'échelle ou de digue où l'on veut étudier la hauteur de l'eau contre la digue. L'événement rare est alors représenté par l'inondation c-à-d le dépassement par l'eau de la hauteur de la digue.

Dans les systèmes unidimensionnels comme ceux étudiés dans cette thèse, il est naturel de choisir la fonction d'importance comme une échelle de la variable de l'espace d'état. Dans les systèmes multidimensionnels, par contre, il n'est généralement pas optimal, comme l'on peut s'y attendre, de choisir la fonction d'importance comme une échelle de l'une des variables de l'espace d'état. Cette non-optimalité a été prouvée par exemple dans [30] dans le cadre de files d'attente en tandem. Marnix Garvels dans [26] propose un algorithme optimal pour le choix de la fonction d'importance : comme l'on pouvait s'y attendre, la fonction d'importance doit tenir compte des chemins menant à l'événement rare et de leur importance en terme de probabilité. Plus précisément, il s'agit de choisir la fonction d'importance de telle sorte que la probabilité d'atteinte du niveau supérieur est égale en tout point d'entrée (le résultat exact dépendant de paramètres inconnus du modèle, l'auteur propose afin de l'estimer une méthode de retournement de temps qui fournit de bons résultats). Remarquons qu'alors une bonne connaissance du comportement du système est nécessaire pour déterminer la fonction d'importance optimale et que le même genre d'information est aussi nécessaire à l'implémentation de la méthode d'échantillonnage préférentiel. Dans [46], LeGland obtient la variance asymptotique des algorithmes particulaires multidimensionnels qui redonne le même résultat que dans le modèle de branchement avec duplication des trajectoires lorsque la probabilité de franchissement des seuils est invariante sur la surface délimitant ces derniers.

Dans les systèmes multidimensionnels, se pose aussi le problème des points d'entrée : l'analyse devient plus complexe lorsque le succès d'un chemin (c-à-d l'atteinte du niveau supérieur B_{i+1}) partant du niveau B_i dépend de son point d'entrée dans B_i. Plus précisément, notons la variable aléatoire $S_i := X_{T_i}$ ayant pour distribution la loi d'entrée en B_i, où

$$T_i := \inf\{t \geq 0 | Z_t \in B_i\}.$$

Il est évident que la probabilité de succès d'un chemin dépendra du point d'entrée S_i ; notons

$$P_{i+1}(s) := \mathbb{E}(I_{i+1}(s)), \qquad s \in \mathcal{S}(S_i)$$

où $I_{i+1}(s)$ est une variable de Bernoulli représentant le succès d'un chemin partant de s à atteindre le niveau supérieur B_{i+1} et $\mathcal{S}(S_i)$ l'espace de définition de S_i. Clairement, on aura alors

$$P_{i+1} = \mathbb{E}(P_{i+1}(s))$$

Il serait alors intéressant d'étudier, dans les systèmes multidimensionnels, la sensibilité des bornes de Chernoff en fonction du choix de la fonction d'importance et des points d'entrée.

BIBLIOGRAPHIE

[1] Alan Agresti. Bounds on the extinction time distribution of a branching process. *Advances in Appl. Probability*, 6 :322–335, 1974.

[2] David Aldous. *Probability approximations via the Poisson clumping heuristic*, volume 77 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1989.

[3] Davis Aldous and Umesh V. Vazirani. "go with the winners" algorithms. *IEEE Symposium on Fundations of Computer Science*, (7) :492–501, 1994.

[4] Krishna B. Athreya and Samuel Karlin. Branching processes with random environments. II. Limit theorems. *Ann. Math. Statist.*, 42 :1843–1858, 1971.

[5] Krishna B. Athreya and Samuel Karlin. On branching processes with random environments. I. Extinction probabilities. *Ann. Math. Statist.*, 42 :1499–1520, 1971.

[6] Krishna B. Athreya and Peter E. Ney. *Branching processes*. Springer-Verlag, New York, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196.

[7] Martine Babillot, Philippe Bougerol, and Laure Elie. The random difference equation $X_n = A_n X_{n-1} + B_n$ in the critical case. *Ann. Probab.*, 25(1) :478–493, 1997.

[8] A.J. Bayes. A minimum variance sampling technique for simulation models. *Journal of the Association for Computing Machinery*, 19 :734–741, October 1972.

[9] K. Below. Rare-event simulation in atm networks : Restart/lre method, implementation and improvement. *Digital Communication Systems*, pages 1–128, 1999.

[10] K. Below, L. Battaglia, and U. Killat. Restart/lre simulation : The reliability issue. *Second International Workshop on Rare Event Simulation, RESIM99, Enschede, The Netherlands*, 1999.

[11] G. Bennett. Probability inequalities for the sum of independant random variables. *J. Amer. Statist. Assoc.*, 57 :33–45, 1962.

[12] S.N. Bernstein. *The Theory of Probabilities.* Gastehizdat Publishing House (Moscow). Moscow, 1946.

[13] Hélène Boistard and Agnès Lagnoux. Estimation des probabilités de grandes déviations par la méthode de Monte Carlo. *Mémoire module de DEA : Grandes Déviations*, 2003.

[14] Philippe Bougerol and Nico Picard. Strict stationarity of generalized autoregressive processes. *Ann. Probab.*, 20(4) :1714–1730, 1992.

[15] Nicolas Bouleau. *Processus stochastiques et applications*, volume 1420 of *Actualités Scientifiques et Industrielles [Current Scientific and Industrial Topics]*. Hermann, Paris, 1988.

[16] Andreas Brandt. The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. in Appl. Probab.*, 18(1) :211–220, 1986.

[17] M. Cosnard and J. Demongeot. Théorèmes de point fixe et processus de Galton-Watson. *Ann. Sci. Math. Québec*, 8(1) :5–21, 1984.

[18] Frédéric Cérou, Pierre Del Moral, François LeGland, and Pascal Lezaud. Genetic genealogical model in rare event analysis. *Latin American Journal of Probability And Mathematical Statistics*. 2006.

[19] Frédéric Cérou and Arnaud Guyader. Adaptive multilevel splitting for rare event analysis. *Rapport de recherche de l'INRIA - Rennes , Equipe : ASPI*. 2005.

[20] Pierre Del Moral. *Feynman-Kac formulae*. Probability and its Applications (New York). Springer-Verlag, New York, 2004. Genealogical and interacting particle systems with applications.

[21] Persi Diaconis and David Freedman. Iterated random functions. *SIAM Rev.*, 41(1) :45–76 (electronic), 1999.

[22] Persi Diaconis and Susan Holmes. Three examples of Monte-Carlo Markov chains : at the interface between statistical computing, computer science, and statistical mechanics. In *Discrete probability and algorithms (Minneapolis, MN, 1993)*, volume 72 of *IMA Vol. Math. Appl.*, pages 43–56. Springer, New York, 1995.

[23] Arnaud Doucet, Nando de Freitas, and Neil Gordon. An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo methods in practice*, Stat. Eng. Inf. Sci., pages 3–14. Springer, New York, 2001.

[24] Dean H. Fearn. Probability of extinction of critical generation-dependent Galton-Watson processes. *J. Appl. Probability*, 13(3) :573–577, 1976.

[25] William Feller. *An introduction to probability theory and its applications. Vol. I.* John Wiley and Sons, Inc., New York, 1957. 2nd ed.

[26] Marnix J.J. Garvels. The splitting method in rare event simulation. *PhD thesis*. 2000.

[27] Marnix J.J. Garvels and Dirk P. Kroese. Splitting for rare event simulation : Analysis of simple cases. In *Proceedings of the 1998 Winter Simulation Conference*, pages 601–609, 1998.

[28] Marnix J.J. Garvels and Dirk P. Kroese. On the entrance distribution in restart simulation. pages 65–88. RESIM Workshop 11.-12. March 1999, University of Twente, Enschede, The Netherlands, 1999.

[29] Marnix J.J. Garvels, Dirk P. Kroese, and J.C.W. van Ommeren. On the importance function in splitting simulation. In *Proceedings of the 2000 Symposium on Performance Evaluation of Computer and Telecommunications Systems*, pages 131–138, 2000.

[30] Paul Glasserman, Philip Heidelberger, Perwez Shahabuddin, and Tim Zajic. A large deviations perspective on the efficiency of multilevel splitting. *IEEE Trans. Automat. Control*, 43(12) :1666–1679, 1998.

[31] Paul Glasserman, Philip Heidelberger, Perwez Shahabuddin, and Tim Zajic. Multilevel splitting for estimating rare event probabilities. *Oper. Res.*, 47(4) :585–600, 1999.

[32] Yves Guivarc'h, Emile Le Page, and Quansheng Liu. Normalisation d'un processus de branchement critique dans un environnement aléatoire. *C. R. Math. Acad. Sci. Paris*, 337(9) :603–608, 2003.

[33] Carmelita Görg and Oliver Fuß. Comparison and optimization of restart run time strategies. In *AEU*, pages 52(3) :197–204, 1998.

[34] Carmelita Görg and Oliver Fuß. Simulating rare event details of atm delay time distributions with restart/lre. In *Proceedings of RESIM Workshop*, pages 41–54, 1999.

[35] Carmelita Görg and Friedrich Schreiber. Rare event simulation : a modified restart method using the lre-algorithm. In *Proceedings of the 14th International Teletraffic Congress*, pages 787–796, 1994.

[36] Z. Haratzi and J. Townsend. Rare event simulation of delay in packet switching networks using dpr-based splitting. In *Proceedings of RESIM Workshop*, pages 185–190, 1999.

[37] Theodore E. Harris. *The theory of branching processes.* Dover Phoenix Editions. Dover Publications Inc., Mineola, NY, 2002. Corrected reprint of the 1963 original [Springer, Berlin ; MR **29** #664].

[38] P.E. Heegaard. Speed-up techniques for simulation. *Telektronikk ISSN 0085-7130*, 91(2-3) :195–207, 1995.

[39] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58 :13–30, 1963.

[40] Mark Jerrum and Alistair Sinclair. *The Markov chain Monte Carlo method : an approach to approximate counting and integration.* Approximation Algorithms for NP-hard Problems. Dorit Hochbaum, 1997.

[41] Christian Kelling. A framework for rare event simulation of stochastic petri nets using "restart". pages 317–324, Coronado, California, USA, December 1996.

[42] Harry Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Math.*, 131 :207–248, 1973.

[43] Agnès Lagnoux. Rare event simulation. *Probability in the Engineering and Informational Sciences*, 20(1) :45–66, 2006.

[44] Agnès Lagnoux-Renaudie. Effective branching splitting method under cost contraint. Submitted to *Stochastic Processes and their Applications*, 2006.

[45] Bernard Lapeyre, Étienne Pardoux, and Rémi Sentis. *Méthodes de Monte-Carlo pour les équations de transport et de diffusion*. Mathématiques & Applications [Mathematics & Applications], 29. Springer-Verlag, Berlin, 1998.

[46] François LeGland. *Genetic Genealogical Model in Rare Event Analysis.* RARE : Analysis of Rare Events Using Monte Carlo Methods. Kick-off meeting, Rennes, April 13-14, 2006.

[47] François LeGland and Nadia Oudjane. *A Sequential Particle Algorithm that Keeps the Particle System Alive.* Antalya, September 4-8, 2005. Proceedings of the 13th European Signal Processing Conference.

[48] François LeGland and Nadia Oudjane. *A Sequential Particle Algorithm that Keeps the Particle System Alive.* Stochastic Hybrid Systems : Theory and Safety Critical Applications. Springer-Verlag, Berlin, 2006. Lecture Notes in Control and Information Sciences 337.

[49] Gábor Lugosi. Concentration-of-measure inequalities. *Lecture notes.* In *Winter School on Probabilistic Methods in High Dimension Phenomena, Toulouse, January 10-14, 2005*, 2006.

[50] Russell Lyons. Probability on trees and networks. *Preprint.* http ://mypage.iu.edu/ rdlyons/. 2002.

[51] V. M. Maksimov. A generalized Bernoulli scheme and its limit distributions. *Teor. Verojatnost. i Primenen.*, 18 :547–556, 1973.

[52] J. R. Norris. *Markov chains.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998. Reprint of 1997 original.

[53] John S. Sadowsky. On Monte Carlo estimation of large deviations probabilities. *Ann. Appl. Probab.*, 6(2) :399–422, 1996.

[54] Eric Slud. *Markov chain Monte Carlo : Simulation Techniques in Statistics.* spring minicourse. pages 1–16, 2004.

[55] Walter L. Smith and William E. Wilkinson. On branching processes in random environments. *Ann. Math. Statist.*, 40 :814–827, 1969.

[56] David Tanny. Limit theorems for branching processes in a random environment. *Ann. Probability*, 5(1) :100–116, 1977.

[57] Wim Vervaat. On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. *Adv. in Appl. Probab.*, 11(4) :750–783, 1979.

[58] Manuel Villén-Altamirano and José Villén-Altamirano. On the efficiency of restart for multidimensional systems. *To appear in ACM Transaction on Modelling and Computer Simulation.*

[59] Manuel Villén-Altamirano and José Villén-Altamirano. Accelerated simulation of rare events using restart method with hysteresis. pages 675–686. Elsevier, Amsterdam, Netherlands, 1991.

[60] Manuel Villén-Altamirano and José Villén-Altamirano. Restart : a method for accelerating rare event simulations. pages 71–76. North-Holland, 1991.

[61] Manuel Villén-Altamirano and José Villén-Altamirano. Restart : An efficient and general method for fast simulation of rare event. *Technical report 7.* 1997.

[62] Manuel Villén-Altamirano and José Villén-Altamirano. Analysis of restart simulation : Theoretical basis and sensitivity study. *European Transactions on Telecommunications*, 13 n°4 :373–385, 2002.

TABLE DES FIGURES

1 Simulation classique ... 35

2 Simulation RESTART ... 36

1 Modèle de branchement avec duplication des trajectoires ... 58

2 Behavior of H and L ... 70

3 Comparaison entre les différentes méthodes-Files d'attente ... 72

4 Arbre considéré dans l'exemple ... 78

5 Correspondance entre arbre et problème du recuit simulé ... 79

6 Bornes of $f_M(s)$ ... 83

1 Sensibilité des bornes en m pour différentes valeurs de α ... 101

2 Comportement de ψ pour $m > 1$ ... 105

3 Estimées supérieures pour $m \neq 1$ ... 111

4 Bornes de Hoeffding, Bennett et Bernstein ... 118

5 Graphe p=f(δ) asymptotiquement en M 132

6 Intervalle de confiance 138

7 Bornes supérieures 142

1 Le modèle de branchement avec duplication des trajectoires à deux étapes 145

2 Modèle de branchement avec duplication des trajectoires version Aldous-Vazirani 171

1 Modèle de branchement avec duplication des trajectoires R=2 195

2 Algorithme séquentiel de LeGland-Oudjane H=2 196

3 Algorithme adaptatif de Cérou-Guyader p=1/2 196

ANNEXE 1 : PREUVES-MODÈLE RESTART (CHAP.2)

Preuve de la Proposition 2.3.1 On souhaite montrer que

$$\mathrm{var}(\hat{P}^M) = \frac{K_A \mathbb{P}(A)}{N}\Big[\frac{1}{r_M} + \sum_{i=1}^{M} \frac{s_i P_{A/i}(R_i - 1)}{r_i}\Big],$$

avec

$$r_i = \prod_{j=1}^{i} R_j, \quad P_{A|i} = \mathbb{P}(A|C_i), \quad K_A = \frac{\mathrm{Var}(N_A^0)}{\mathbb{E}[N_A^0]},$$

$$s_i = \frac{1}{K_A P_{A|i}} \frac{\mathrm{Var}\big(\mathbb{E}[N_A^0|\chi_i]\big)}{\mathbb{E}[N_A^0]}, \quad \chi_i = (N_i^0, X_{\tau_i^{(1)}}, \cdots, X_{\tau_i^{(N_i^0)}}).$$

Démonstration. — On montre ce résultat par récurrence. Rappelons, tout d'abord, que la variance conditionnelle est définie par

$$\mathrm{var}(X|\mathcal{F}) = \mathbb{E}\big[(X - \mathbb{E}(X|\mathcal{F}))^2|\mathcal{F}\big].$$

• Pour 0 seuil (simulation simple sans retirage) : on a $\hat{P}^0 = \frac{N_A^0}{N}$ avec N_A^0 le temps passé en A par la simulation principale. Ainsi

$$\mathrm{var}(\hat{P}^0) = \mathrm{var}(\frac{N_A^0}{N}) = \frac{\mathrm{var}(N_A^0)}{N^2} = \frac{\mathbb{P}(A)}{N}\frac{\mathrm{var}(N_A^0)}{\mathbb{E}(N_A^0)} = \frac{\mathbb{P}(A)K_A}{N}, \text{ si } K_A = \frac{\mathrm{var}(N_A^0)}{\mathbb{E}(N_A^0)}.$$

K_A traduit en fait l'autocorrélation du processus. Si le processus est décorrélé, $K_A = 1 - \mathbb{P}(A)$. Plus précisément,

$$K_A = 1 - \mathbb{P}(A) + \sum_{1 \le j \le N,\ j \ne i} [\mathbb{P}(Y_j = 1|Y_i = 1) - \mathbb{P}(A)].$$

En effet, on va montrer que $\mathrm{var}(N_A^0) = N\mathbb{P}(A)K_A$. Or, $\mathrm{var}(N_A^0) = \mathrm{var}(\sum_{i=1}^N Y_i)$ avec $Y_i =_{loi} \mathcal{B}(\mathbb{P}(A))$.

Faisons le calcul pour $N = 2$.

$$\begin{aligned} \mathrm{var}(Y_1 + Y_2) &= \mathrm{var}(Y_1) + \mathrm{var}(Y_2) + 2\,\mathrm{cov}(Y_1, Y_2) \\ &= 2(\mathrm{var}(Y_1) + \mathrm{cov}(Y_1, Y_2)). \end{aligned}$$

Or, $\mathbb{E}(Y_1Y_2) = \mathbb{E}(Y_1Y_2\mathbf{1}_{Y_1=1}) + \mathbb{E}(Y_1Y_2\mathbf{1}_{Y_1=0}) = \mathbb{E}(Y_2\mathbf{1}_{Y_1=1}) = \mathbb{E}(Y_2\mathbb{E}(\mathbf{1}_{Y_1=1}|Y_2))$ et comme

$$\mathbb{E}(\mathbf{1}_{Y_1=1}|Y_2) = \mathbb{P}(Y_1 = 1|Y_2 = 1)\mathbf{1}_{Y_2=1} + \mathbb{P}(Y_1 = 1|Y_2 = 0)\mathbf{1}_{Y_2=0},$$

$$\begin{aligned} \mathbb{E}(Y_1Y_2) &= \mathbb{E}(Y_2\mathbf{1}_{Y_2=1}\mathbb{P}(Y_1 = 1|Y_2 = 1)) + \mathbb{E}(Y_2\mathbf{1}_{Y_2=0}\mathbb{P}(Y_1 = 1|Y_2 = 0)) \\ &= \mathbb{E}(\mathbf{1}_{Y_2=1}\mathbb{P}(Y_1 = 1|Y_2 = 1)) = \mathbb{P}(Y_1 = 1|Y_2 = 1)\mathbb{P}(Y_2 = 1) \\ &= \mathbb{P}(Y_1 = 1|Y_2 = 1)\mathbb{P}(A) \end{aligned}$$

Or, $\mathbb{E}(Y_1) = \mathbb{E}(Y_2) = \mathbb{P}(A)$, d'où

$$\mathrm{cov}(Y_1, Y_2) = \mathbb{P}(Y_1 = 1|Y_2 = 1)\mathbb{P}(A) - \mathbb{P}(A)^2 = \mathbb{P}(A)[\mathbb{P}(Y_1 = 1|Y_2 = 1) - \mathbb{P}(A)],$$

ainsi

$$\mathrm{var}(Y_1 + Y_2) = 2[\mathrm{var}(Y_1) + \mathrm{cov}(Y_1, Y_2)] = 2\mathbb{P}(A)[1 - \mathbb{P}(A) + \mathbb{P}(Y_1 = 1|Y_2 = 1) - \mathbb{P}(A)]$$

Nous obtenons

$$K_A = \frac{\mathrm{var}(N_A^0)}{\mathbb{E}(N_A^0)} = 1 - \mathbb{P}(A) + [\mathbb{P}(Y_1 = 1|Y_2 = 1) - \mathbb{P}(A)].$$

Pour $N \geq 3$, on développe de la même façon le produit, on obtient N fois la variance et la somme des covariances que l'on exprime de la même manière que précédemment.

• Pour M seuils : supposons que, pour $M - 1$ seuils, la variance de $\hat{P}^{M-1}$ soit donnée par la formule précédente adaptée au cas de $M - 1$ seuils numérotés de 2 à M, ie

$$\mathrm{var}(\hat{P}^{M-1}) = \frac{K_A\mathbb{P}(A)}{N}\Big[\frac{1}{\prod_{i=2}^M R_i} + \sum_{i=2}^{M} \frac{s_i P_{A|i}(R_i - 1)}{\prod_{i=2}^M R_i}\Big]$$

Remarquons ensuite que

$$\begin{cases} \mathrm{var}(\hat{P}^M) = \mathrm{var}(\mathbb{E}(\hat{P}^M|\chi_1)) + \mathbb{E}(\mathrm{var}(\hat{P}^M|\chi_1)) \\ \mathrm{var}(\hat{P}^{M-1}) = \mathrm{var}(\mathbb{E}(\hat{P}^{M-1}|\chi_1)) + \mathbb{E}(\mathrm{var}(\hat{P}^{M-1}|\chi_1)) \end{cases}$$

Comme $\mathbb{E}(\alpha_1|\chi_1) = \mathbb{E}(\alpha_2|\chi_1) = \cdots = \mathbb{E}(\alpha_{R_1}|\chi_1)$,

$$\begin{aligned}\mathbb{E}(\hat{P}^M|\chi_1) &= \mathbb{E}(\frac{N_A}{N\prod_{i=1}^M R_i}|\chi_1) = \frac{\sum_{i=1}^{R_1}\mathbb{E}(\alpha_i|\chi_1)}{N\prod_{i=1}^M R_i} = \frac{\mathbb{E}(\alpha_1|\chi_1)}{N\prod_{i=2}^M R_i} \\ &= \mathbb{E}(\hat{P}^{M-1}|\chi_1) = \mathbb{E}(\hat{P}^0|\chi_1).\end{aligned}$$

Ainsi $\mathrm{var}(\mathbb{E}(\hat{P}^M|\chi_1)) = \mathrm{var}(\mathbb{E}(\hat{P}^{M-1}|\chi_1))$. D'autre part, comme les R_1 sont des va $\alpha_i|\chi_1$ sont iid,

$$\begin{aligned}\mathrm{var}(\hat{P}^M|\chi_1) &= \mathrm{var}(\frac{N_A}{N\prod_{i=1}^M R_i}|\chi_1) = \mathrm{var}(\frac{\sum_{i=1}^{R_1}\alpha_i}{N\prod_{i=1}^M R_i}|\chi_1) \\ &= \frac{1}{R_1}\mathrm{var}(\frac{\alpha_1}{N\prod_{i=1}^M R_i}|\chi_1) = \frac{1}{R_1}\mathrm{var}(\hat{P}^{M-1}|\chi_1)\end{aligned}$$

Donc,

$$\mathbb{E}(\mathrm{var}(\hat{P}^M|\chi_1)) = \frac{1}{R_1}\mathbb{E}(\mathrm{var}(\hat{P}^{M-1}|\chi_1)).$$

Par conséquent, $\mathrm{var}(\hat{P}^M) = \mathrm{var}(\mathbb{E}(\hat{P}^{M-1}|\chi_1)) + \frac{1}{R_1}\mathbb{E}(\mathrm{var}(\hat{P}^{M-1}|\chi_1))$. Or,

$$\mathrm{var}(\hat{P}^{M-1}) = \mathrm{var}(\mathbb{E}(\hat{P}^{M-1}|\chi_1)) + \mathbb{E}(\mathrm{var}(\hat{P}^{M-1}|\chi_1))$$

donc

$$\frac{1}{R_1}\mathbb{E}(\mathrm{var}(\hat{P}^{M-1}|\chi_1)) = \frac{1}{R_1}\mathrm{var}(\hat{P}^{M-1}) - \frac{1}{R_1}\mathrm{var}(\mathbb{E}(\hat{P}^{M-1}|\chi_1)).$$

D'où

$$\begin{aligned}\mathrm{var}(\hat{P}^M) &= \frac{\mathrm{var}(\hat{P}^{M-1})}{R_1} + \frac{R_1-1}{R_1}\mathrm{var}(\mathbb{E}(\hat{P}^{M-1}|\chi_1)) = \frac{\mathrm{var}(\hat{P}^{M-1})}{R_1} + \frac{R_1-1}{R_1}\mathrm{var}(\mathbb{E}(\hat{P}^0|\chi_1)) \\ &= \frac{\mathrm{var}(\hat{P}^{M-1})}{R_1} + \frac{R_1-1}{R_1}\mathrm{var}(\frac{\mathbb{E}(N_A^0|\chi_1)}{N}) \\ &= \frac{1}{R_1}\frac{K_A\mathbb{P}(A)}{N}\Big[\frac{1}{\prod_{i=2}^M R_i} + \sum_{i=2}^M \frac{s_i P_{A|i}(R_i-1)}{\prod_{i=2}^M R_i}\Big] + \frac{R_1-1}{R_1 N^2}\mathrm{var}(\mathbb{E}(N_A^0|\chi_1))\end{aligned}$$

par hypothèse de récurrence. En utilisant les définitions des différents paramètres, on obtient le résultat. □

Preuve de la Proposition 2.3.3 On souhaite montrer que

$$G = \frac{1}{\mathbb{P}(A)[\sum_{i=0}^M \frac{s_{i+1}u_i(1-P_{i+1|i})}{P_{i+1|0}r_i}][\sum_{i=0}^M y_i v_{i+1} P_{i|0} r_i (1-Pi+1|i)]},$$

avec

$$\begin{cases} u_i & = \frac{1-(s_i/s_{i+1})P_{i+1|i}}{1-P_{i+1|i}} \\ v_{i+1} & = \frac{1-(y_{i+1}/y_i)P_{i+1|i}}{1-P_{i+1|i}} \\ s_0 & = 0, s_{M+1} = 1 \text{ et } y_{M+1} = 0. \end{cases}$$

Démonstration. — On a $C_s = N, V_s = \frac{\mathbb{P}(A)K_A}{N}$, par conséquent,

$$G = \frac{1}{[y_0 + \sum_{i=1}^M y_i P_{i|0} r_{i-1}(R_i - 1)][\frac{1}{r_M} + \sum_{i=1}^M \frac{s_i P_{A|i}(R_i-1)}{r_i}]}.$$

Donc,

$$\begin{aligned}
\mathbb{P}(A) \sum_{i=0}^{M} \frac{s_{i+1} u_i (1 - P_{i+1|i})}{P_{i+1|0} r_i} &= \mathbb{P}(A) \sum_{i=0}^{M} \frac{s_{i+1}}{P_{i+1|0} r_i} \frac{1 - \frac{s_i}{s_{i+1}} P_{i+1|i}}{1 - P_{i+1|i}} (1 - P_{i+1|i}) \\
&= \sum_{i=0}^{M} \frac{s_{i+1} P_{A|i+1}}{r_i} (1 - \frac{s_i}{s_{i+1}} P_{i+1|i}) \text{ car } \frac{\mathbb{P}(A)}{P_{i+1|0}} = P_{A|i+1} \\
&= \sum_{i=0}^{M} \frac{s_{i+1} P_{A|i+1}}{r_i} - \sum_{i=0}^{M} \frac{s_i P_{A|i}}{r_i} = \sum_{i=1}^{M+1} \frac{s_i P_{A|i}}{r_{i-1}} - \sum_{i=0}^{M} \frac{s_i P_{A|i}}{r_i} \\
&= \sum_{i=1}^{M} \frac{s_i P_{A|i}}{r_{i-1}} - \frac{s_0 P_{A|0}}{r_0} + \frac{s_{M+1} P_{A|M+1}}{r_M} = \frac{1}{r_M} + \sum_{i=1}^{M} \frac{s_i P_{A|i}}{r_{i-1}} \\
&= \frac{1}{r_M} + \sum_{i=1}^{M} \frac{s_i P_{A|i}(R_i - 1)}{r_i}
\end{aligned}$$

et,

$$\begin{aligned}
\sum_{i=0}^{M} y_i v_{i+1} P_{i|0} r_i (1 - P_{i+1|i}) &= \sum_{i=0}^{M} y_i \frac{1 - \frac{y_{i+1}}{y_i} P_{i+1|i}}{1 - P_{i+1|i}} P_{i|0} r_i (1 - P_{i+1|i}) \\
&= \sum_{i=0}^{M} y_i (1 - \frac{y_{i+1}}{y_i} P_{i+1|i}) P_{i|0} r_i = \sum_{i=0}^{M} y_i P_{i|0} r_i - \sum_{i=0}^{M} y_{i+1} P_{i+1|0} r_i \\
&= \sum_{i=0}^{M} y_i P_{i|0} r_i - \sum_{i=1}^{M+1} y_i P_{i|0} r_{i-1} \\
&= \sum_{i=1}^{M} y_i P_{i|0} (r_i - r_{i-1}) - y_{M+1} P_{M+1|0} r_M + y_0 P_{0|0} r_0 \\
&= y_0 + \sum_{i=1}^{M} y_i P_{i|0} r_{i-1} (R_i - 1)
\end{aligned}$$

On obtient ainsi l'expression du gain. □

Calcul de var$(\mathbb{E}(N_A^0|\chi_1))$ On souhaite montrer que

$$\mathrm{var}(\mathbb{E}(N_A^0|\chi_1)) = \mathbb{E}(N_i^0)(P_{A|i})^2 [K_i^{,} + \frac{\mathrm{var}(P^*_{A|X_i})}{(P^*_{A|i})^2} \gamma_i]$$

Démonstration. — D'après le lemme de Wald, si les X_i^l sont indépendants, pour toute fonction f,

$$\mathbb{E}(f(X_i)) = \frac{\mathbb{E}(\sum_{l=1}^{N_i^0} f(X_i^l))}{\mathbb{E}(N_i^0)}. \tag{106}$$

Ainsi,

- pour $f(X_i) = P^*_{A|X_i}$, comme $P^*_{A|i} = \mathbb{E}(P^*_{A|X_i})$ et $\mathbb{E}(\sum_{l=1}^{N_i^0} P^*_{A|X_i^l}) = \mathbb{E}(N_A^0)$ alors $P^*_{A|i} = \mathbb{E}(N_A^0)/\mathbb{E}(N_i^0)$,
- pour $f(X_i) = (P^*_{A|X_i})^2$, on obtient $\mathrm{var}(P^*_{A|X_i}) = \frac{\mathbb{E}(\sum_{l=1}^{N_i^0} (P^*_{A|X_i^l})^2)}{\mathbb{E}(N_i^0)} - (P^*_{A|i})^2$.

On peut étendre la formule (106), par polarisation, pour approximer une fonction en deux événements B_i. Cela permet de définir l'autocovariance de $P^*_{A|X_i}$

comme suit :

$$ACV_m(P^*_{A|X_i}) = \frac{\mathbb{E}(\sum_{l=1}^{N_i^0-m} P^*_{A|X_i^l} P^*_{A|X_i^{l+m}})}{\mathbb{E}(\max(0, N_i^0 - m))} - (P^*_{A|i})^2.$$

Comme $N_A^0 = \sum_{l=1}^{N_i^0} Z_{i,1}^l$, $\mathbb{E}(\sum_{l=1}^{N_i^0} Z_{i,1}^l|\chi_i) = \sum_{l=1}^{N_i^0} \mathbb{E}(Z_{i,1}^l|\chi_i) = \sum_{l=1}^{N_i^0} P^*_{A|X_i^l}$,

$$\begin{aligned}
\mathrm{var}(\mathbb{E}(N_A^0|\chi_1)) &= \mathbb{E}(\mathbb{E}(N_A^0|\chi_i)^2) - \mathbb{E}(N_A^0)^2 = \mathbb{E}(\mathbb{E}(\sum_{l=1}^{N_i^0} Z_{i,1}^l|\chi_i)^2) - \mathbb{E}(N_A^0)^2 \\
&= \mathbb{E}((\sum_{l=1}^{N_i^0} P^*_{A|X_i^l})^2) - \mathbb{E}(N_A^0)^2 \\
&= \mathbb{E}(\sum_{l=1}^{N_i^0} (P^*_{A|X_i^l})^2) + 2\sum_{m=1}^{\infty} \mathbb{E}(\sum_{l=1}^{N_i^0-m} P^*_{A|X_i^l} P^*_{A|X_i^{l+m}}) - \mathbb{E}(N_A^0)^2
\end{aligned}$$

Enfin, puisque $2\sum_{m=1}^{\infty} \mathbb{E}[\max(0, N_i^0 - m)] = \mathbb{E}((N_i^0)^2) - \mathbb{E}(N_i^0)$,

$$\mathrm{var}(\mathbb{E}(N_A^0|\chi_1)) = (P_{A|i})^2 \mathrm{var}(N_i^0) + \mathbb{E}(N_i^0)\mathrm{var}(P^*_{A|X_i}) + 2\sum_{m=1}^{\infty} \mathbb{E}[\max(0, N_i^0 - m)] ACV_m(P^*_{A|X_i}).$$

Définissons pour $i = 1..M$, $K_i^{'} = \frac{\mathrm{var}(N_i^0)}{\mathbb{E}(N_i^0)}$ et

$$\gamma_i = 1 + 2\sum_{m=1}^{\infty} \frac{\mathbb{E}[\max(0, N_i^0 - m)]}{\mathbb{E}(N_i^0)} \frac{ACV_m(P^*_{A|X_i})}{\mathrm{var}(P_{A|X_i})}$$

On obtient alors le résultat. □

Calcul de K_A On souhaite montrer que

$$K_A = a_j P_{A|j} \left(K_j^{'} + \frac{\mathrm{var}(P^*_{A|X_j})}{(P^*_{A|j})^2}\gamma_i + \frac{\mathbb{E}(\ \mathrm{var}(Z_{i,1}|X_j))}{(P^*_{A|j})^2} \right)$$

Démonstration. — Tout d'abord, $K_A = \frac{\mathrm{var}(N_A^0)}{\mathbb{E}(N_A^0)} = \frac{\mathrm{var}(\mathbb{E}(N_A^0|\chi_j)) + \mathbb{E}(\ \mathrm{var}(N_A^0|\chi_j))}{\mathbb{E}(N_A^0)}$.

Pour calculer $\mathbb{E}(\ \mathrm{var}(N_A^0|\chi_j))$, on utilise la relation $N_A^0 = \sum_{l=1}^{N_i^0} Z_{i,1}^l$, donc

$$\mathrm{var}(N_A^0|\chi_j) = \ \mathrm{var}(\sum_{l=1}^{N_i^0} Z_{i,1}^l|\chi_j) = \sum_{l=1}^{N_i^0} \ \mathrm{var}(Z_{i,1}^l|X_j^l)$$

Et en utilisant la formule (106),

$$\mathbb{E}(\ \mathrm{var}(N_A^0|\chi_j)) = \mathbb{E}(N_j^0)\mathbb{E}(\ \mathrm{var}(Z_{i,1}|X_j))$$

Enfin, puisque $\mathbb{E}(N_A^0) = P_{A|j}^*\mathbb{E}(N_j^0)$ et $P_{A|j}^* = a_j P_{A|j}$, on obtient le résultat recherché. $\square$

ANNEXE 2 : PREUVES-SYSTÈMES PARTICULAIRES (CHAP.2)

Preuve du théorème 2.4.1 Nous souhaitons montrer que l'évolution de μ_n est régie par le système dynamique non-linéaire suivant :

$$\mu_{n+1} = \Phi_{n+1}(\mu_n) \tag{107}$$

avec $\Phi : \mathcal{P}_n(E) \rightarrow \mathcal{P}(E)$

$$\Phi_{n+1}(\mu) = \Psi_n(\mu)\mathcal{M}_{n+1}$$

où

- $\Psi_n : \mathcal{P}_n(E) \rightarrow \mathcal{P}_n(E)$ et

$$\Psi_n(\mu)(dx_n) = \frac{g_n(x_n)}{\mu(g_n)}\mu(dx_n)$$

  Ψ_n est la **transformation de Gibbs-Boltzman**
- $\mathcal{M}_{n+1}$ représente le noyau de transition de la chaîne de Markov χ_{n+1}, ie

$$\mathcal{M}_{n+1}(u, dv) = \mathbb{P}(\chi_{n+1} \in dv | \chi_n = u)$$

Démonstration. — 1. On montre d'abord que $\mu_n = \hat{\mu}_{n-1}\mathcal{M}_n$. En effet,

$$\begin{aligned}
\gamma_n(f) &= \mathbb{E}(f(\chi_n)\prod_{p=0}^{n} g_p(\chi_p)) = \mathbb{E}(\mathbb{E}(f(\chi_n)|\chi_1, \cdots, \chi_{n-1})\prod_{p=0}^{n} g_p(\chi_p)) \\
&= \mathbb{E}(\mathbb{E}(f(\chi_n)|\chi_{n-1})\prod_{p=0}^{n} g_p(\chi_p)) = \mathbb{E}(\mathcal{M}_n(f)(\chi_{n-1})\prod_{p=0}^{n} g_p(\chi_p)) \\
&= \hat{\gamma}_{n-1}(\mathcal{M}_n(f))
\end{aligned}$$

Or $\mu_n(f) = \frac{\gamma_n(f)}{\gamma_n(1)} = \frac{\hat{\gamma}_{n-1}(\mathcal{M}_n(f))}{\hat{\gamma}_{n-1}(1)} = \hat{\mu}_{n-1}(\mathcal{M}_n(f))$.

2. Ensuite, on montre que $\hat{\mu}_n = \Psi_n(\mu_n)$. En effet,

$$\begin{aligned}\Psi_n(\mu_n)(f) &= \int f(x_n)\Psi_n(\mu_n)(dx_n) = \frac{1}{\mu_n(g_n)}\int f(x_n)g_n(x_n)\mu_n(dx_n) \\ &= \frac{\mu_n(fg_n)}{\mu_n(g_n)} = \hat{\mu}_n(f)\end{aligned}$$

3. Ainsi, $\mu_n = \hat{\mu}_{n-1}\mathcal{M}_n = \Psi_n(\mu_n)\mathcal{M}_n = \Phi_n(\mu_n)$.

$\square$

Outils pour le calcul du biais et le TCL Définissons $Q_{p,n}$ par $Q_{p,n} = Q_{p+1}\cdots Q_n$ avec

$$Q_n(x_{n-1}, dx_n) = g_{n-1}(x_{n-1})\mathcal{M}_n(x_{n-1}, dx_n)$$

On pose $Q_{n,n} = Id$. Alors $Q_{p,n}(f)(x_p) = \mathbb{E}_{p,x_p}(f(\chi_n)\prod_{q=p}^{n-1} g_q(\chi_q))$.

Proposition 7.0.1. — *On a alors les relations suivantes*

$$\mu_n(f) = \frac{\mu_p(Q_{p,n}(f))}{\mu_p(Q_{p,n}(1)} \quad et \quad \gamma_n(1) = \gamma_p(Q_{p,n}(1))$$

Proposition 7.0.2. — *On définit le processus $\Gamma^N_{\bullet,n}(f)$ de $p \in \{0,\cdots n\}$ dans $\mathbb{R}$ par*

$$\Gamma^N_{p,n}(f) = \gamma^N_p(Q_{p,n}f)\mathbf{1}_{\tau^N \geq p} - \gamma_p(Q_{p,n}f)$$

Alors $\forall p \leq n$, $\Gamma^N_{\bullet,n}(f)$ a une décomposition martingale :

$$\Gamma^N_{p,n}(f) = \sum_{q=0}^{p} \gamma^N_q(1)\mathbf{1}_{\tau^N \geq q}[\mu^N_q(Q_{q,n}f) - \mu^N_{q-1}K_{q,\mu^N_{q-1}}(Q_{q,n}f)]$$

Démonstration. — On montre d'abord que (somme télescopique)

$$\gamma^N_p(\varphi)\mathbf{1}_{\tau^N \geq p} - \gamma_p(\varphi) = \sum_{q=0}^{p}[\gamma^N_q(Q_{q,p}\varphi)\mathbf{1}_{\tau^N \geq q} - \gamma^N_{q-1}(Q_{q-1,p}\varphi)\mathbf{1}_{\tau^N \geq q-1}]$$

Pour le premier terme, on a

$$\gamma^N_q(Q_{q,p}\varphi)\mathbf{1}_{\tau^N \geq q} = \gamma^N_q(1)\mathbf{1}_{\tau^N \geq q}\mu^N_q(Q_{q,p}\varphi)$$

En effet, $\gamma_q^N(\varphi) = \mu_q^N(\varphi)\prod_{p=0}^{q-1}\mu_q^N(g_p) = \mu_q^N(\varphi)\gamma_q^N(1)$, d'où le résultat.

Pour le second terme, on a

$$\gamma_{q-1}^N(Q_{q-1,p}\varphi)\mathbf{1}_{\tau^N\geq q-1} = \gamma_q^N(1)\mathbf{1}_{\tau^N\geq q}\Phi_q(\mu_{q-1}^N)(Q_{q,p}\varphi)$$

En effet, on a d'abord

$$Q_{q-1,p}\varphi = Q_q(Q_{q+1}\cdots Q_p)\varphi = QqQ_{q,p}\varphi = g_{q-1}\mathcal{M}_q(Q_{q,p}\varphi)$$

Donc

$$\begin{aligned}\gamma_{q-1}^N(Q_{q-1,p}\varphi)\mathbf{1}_{\tau^N\geq q-1} &= \gamma_{q-1}^N(g_{q-1}\mathcal{M}_q(Q_{q,p}\varphi))\mathbf{1}_{\tau^N\geq q-1}\\ &= \gamma_{q-1}^N(1)\mu_{q-1}^N(g_{q-1}\mathcal{M}_q(Q_{q,p}\varphi))\mathbf{1}_{\tau^N\geq q-1}\end{aligned}$$

Mais - $1 = \mathbf{1}_{\mu_{q-1}^N(g_{q-1})=0} + \mathbf{1}_{\mu_{q-1}^N(g_{q-1})>0}$

- $\mathbf{1}_{\mu_{q-1}^N(g_{q-1})>0}\mathbf{1}_{\tau^N\geq q-1} = \mathbf{1}_{\tau^N\geq q}$

- $\mu_{q-1}^N(g_{q-1}\mathcal{M}_q(Q_{q,p}\varphi))\mathbf{1}_{\tau^N\geq q-1} = 0$

Ainsi

$$\begin{aligned}\gamma_{q-1}^N(Q_{q-1,p}\varphi)\mathbf{1}_{\tau^N\geq q-1} &= \gamma_{q-1}^N(1)\mathbf{1}_{\tau^N\geq q}\mu_{q-1}^N(g_{q-1}\mathcal{M}_q(Q_{q,p}\varphi)) = \gamma_{q-1}^N(1)\mathbf{1}_{\tau^N\geq q}\mu_{q-1}^N(Q_q(Q_{q,p}\varphi))\\ &= \gamma_{q-1}^N(1)\mathbf{1}_{\tau^N\geq q}\mu_{q-1}^N(Q_{q-1,p}\varphi) = \gamma_{q-1}^N(1)\mathbf{1}_{\tau^N\geq q}\Phi_q(\mu_{q-1}^N)(Q_{q,p}\varphi)\end{aligned}$$

D'où

$$\begin{aligned}\gamma_p^N(\varphi)\mathbf{1}_{\tau^N\geq p} - \gamma_p(\varphi) &= \sum_{q=0}^{p}\gamma_q^N(1)\mathbf{1}_{\tau^N\geq q}[\mu_q^N(Q_{q,p}\varphi) - \Phi_q(\mu_{q-1}^N)(Q_{q,p}\varphi)]\\ &= \sum_{q=0}^{p}\gamma_q^N(1)\mathbf{1}_{\tau^N\geq q}[\mu_q^N(Q_{q,p}\varphi) - \mu_{q-1}^N K_{q,\mu_{q-1}^N}(Q_{q,p}\varphi)]\end{aligned}$$

On applique ce résultat à $\varphi = Q_{q,n}(f)$:

$$\gamma_p^N(Q_{p,n}f)\mathbf{1}_{\tau^N\geq p} - \gamma_p(Q_{p,n}f) = \sum_{q=0}^{p}\gamma_q^N(1)\mathbf{1}_{\tau^N\geq q}[\mu_q^N(Q_{q,n}\varphi) - \mu_{q-1}^N K_{q,\mu_{q-1}^N}(Q_{q,n}\varphi)]$$

puis à $p = n$ et on obtient le résultat. □

Proposition 7.0.3. — *Définissons* $V_q^N(f) = \sqrt{N}[\mu_q^N(f) - \mu_{q-1}^N K_{q,\mu_{q-1}^N}(f)]$.

Alors $(V_q^N)_{0\leq q\leq n}$ *converge en loi vers* $(V_q)_{0\leq q\leq n}$ *(n+1) va indépendantes gaussiennes telles que*

$$\mathbb{E}(V_q(f)) = 0$$

$$\mathbb{E}(V_q(f_1)V_q(f_2)) = \mu_{q-1}\{K_{q,\mu_{q-1}}[f_1 - K_{q,\mu_{q-1}}(f_1)][f_2 - K_{q,\mu_{q-1}}(f_2)]\}$$

Lemme 7.0.1. — *Lemme de Slutsky*
Soit (E, d) *un espace métrique. Supposons que*

$$\begin{cases} X_n \rightarrow_{n\rightarrow\infty,\ \mathcal{L}} X \\ d(X_n, Y_n) \rightarrow_{n\rightarrow\infty,\ \mathbb{P}} 0 \end{cases}$$

alors $Y_n \rightarrow_{n\rightarrow\infty,\ \mathcal{L}} X$

Proposition 7.0.4. — δ*-méthode*
Soient $(U_0^N, \cdots, U_n^N)_{N\geq 1}$ *une suite de vecteurs aléatoires de* $\mathbb{R}^{n+1}$ *définie sur un espace de probabilité et* $(u_p)_{0\leq p\leq n}$ *un point de* $\mathbb{R}^{n+1}$ *tels que*

$$\sqrt{N}(U_0^N - u_0, \cdots, U_n^N - u_n)$$

converge en loi, quand $N \rightarrow \infty$*, vers un vecteur aléatoire noté* $(U_0, \cdots U_n)$*. Alors, pour toute fonction* $F_n : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ *différentiable au point* $(u_p)_{0\leq p\leq n}$*, la suite*

$$\sqrt{N}[F_n(U_0^N, \cdots, U_n^N) - F_n(u_0, \cdots, u_n)]$$

converge en loi quand $N \rightarrow \infty$ *vers la variable aléatoire* $\sum_{p=0}^{n} \frac{\partial F_n}{\partial u_i}(u_0, \cdots, u_n)U_p$*.*

Preuve du théorème 2.4.3 Rappelons que l'estimateur de $\mathbb{P}(V(X) \geq a)$. Nous voulons montrer qu'il est non biaisé.

Démonstration. — On utilise le fait que $\Gamma_{\bullet,M+2}^N(1)$ est une martingale et on l'applique à $p = M + 2$: on a alors

$$\mathbb{E}(\Gamma_{M+2,M+2}^N(1)) = \mathbb{E}(\Gamma_{0,M+2}^N(1)) = \mathbb{E}(\gamma_0^N(1)\mathbf{1}_{\tau^N\geq 0} - \gamma_0(1)) = 0$$

Mais $\mathbb{E}(\Gamma_{M+2,M+2}^N) = \gamma_{M+2}^N(1)\mathbf{1}_{\tau^N\geq M+2} - \gamma_n(1)$, d'où le résultat. □

Preuve du théorème 2.4.4 Nous souhaitons montrer le théorème central limite suivant :

$$\sqrt{N}(\mathbf{1}_{\tau^N\geq M+2}\gamma_{M+2}^N(1) - \mathbb{P}(V(X) \geq a)) \underset{\mathcal{L},N\rightarrow+\infty}{\rightarrow} L_{M+2}$$

où L_{M+2} est un processus gaussien centré de variance

$$v_{M+1}^2 = \sum_{q=0}^{M+2} \gamma_q(1)^2 \mu_{q-1}\{K_{q,\mu_{q-1}}[Q_{q,M+2}(1) - K_{q,\mu_{q-1}}(Q_{q,M+2}(1))]^2\}$$

Démonstration. — Posons $n = M + 2$. Par la décomposition martingale de $\Gamma^N_{p,n}(1)$, on remarque que

$$\Gamma^N_{n,n}(1) := \sum_{q=0}^{n} \gamma_q^N(1)\mathbf{1}_{\tau^N \geq q} \frac{V_q^N(Q_{q,n}(1))}{\sqrt{N}} := F_n(U^N_{0,n}, \cdots, U^N_{n,n})$$

avec

$$\begin{cases} U^N_{q,n} := \gamma_q^N(1)\mathbf{1}_{\tau^N \geq q} \frac{V_q^N(Q_{q,n}(1))}{\sqrt{N}} \\ F_n(u_0, \cdots, u_n) := \sum_{q=0}^n u_q \end{cases}$$

Or

$$\begin{cases} \gamma_q^N(1)\mathbf{1}_{\tau^N \geq q} \underset{\mathbb{P}}{\rightarrow} \gamma_q(1) & \text{par le théorème 2.4.2} \\ V_q^N(Q_{q,n}1) \underset{\mathcal{L}}{\rightarrow} V_q(Q_{q,n}1) & \text{par la proposition 7.0.3} \end{cases}$$

D'où par Slutsky,

$$\sqrt{N}U^N_{q,n} \rightarrow_{\mathcal{L}} \gamma_q(1)V_q(Q_{q,n}1) := \text{"}\sqrt{N}u_q\text{"}$$

Donc $\sqrt{N}(U^N_{q,n} - u_q) \underset{\mathcal{L}}{\rightarrow} 0 := U_{q,n}$. On applique alors la δ-méthode pour conclure que

$$\sqrt{N}[F_n(U^N_{0,n}, \cdots, U^N_{n,n}) - F_n(u_0, \cdots, u_n)] \underset{\mathcal{L}}{\rightarrow} \sum_{p=0}^{n} \frac{\partial F_n}{\partial u_p}(u_0, \cdots u_n)U_{p,n} = 0$$

Ce qui se traduit par le résultat que l'on cherchait.

Pour le calcul de la variance on utilise la proposition 7.0.3,

$$\begin{aligned} var(L_n) &= \sum_{q=0}^{n} \gamma_q(1)^2 var V_q(Q_{q,n}(1)) = \sum_{q=0}^{n} \gamma_q(1)^2 \mathbb{E}(V_q(Q_{q,n}(1))^2 \\ &= \sum_{q=0}^{n} \gamma_q(1)^2 \mu_{q-1}\{K_{q,\mu_{q-1}}[Q_{q,n}(1) - K_{q,\mu_{q-1}}(Q_{q,n}(1))]^2\} \end{aligned}$$

□

Preuve de la proposition 2.4.2 Nous souhaitons montrer que

$$v^2_{M+1} = \mathbb{P}\left(V(X) \geq a\right)^2 (M+1) \left(\frac{1}{P_0} - 1\right)$$

Démonstration. — Notons $n = M+1$ et rappelons que

$$v_n^2 = \sum_{q=1}^{n+1} \gamma_q(1)^2 \mu_{q-1}\{K_{q,\mu_{q-1}}\left[Q_{q,n+1}(1) - K_{q,\mu_{q-1}}\left(Q_{q,n+1}(1)\right)\right]^2\}$$

que l'on réécrit aisément en

$$v_n^2 = \gamma_{n+1}(1)^2 s_n$$

avec

$$s_n := \sum_{p=1}^{n} \mathbb{E}\left(\left[\frac{\mathbb{P}(B_n|B_p)}{\mathbb{P}(B_n|B_{p-1})}\mathbf{1}_{B_p} - 1\right]^2 |B_{p-1}\right)$$

En effet, on a tout d'abord, en notant $\tilde{\varphi}_{q,\eta} = \varphi - \Phi_q(\eta)(\varphi)$,

$$\begin{array}{rcl} K_{q,\eta}(\varphi - K_{q,\eta}(\varphi))^2 & = & K_{q,\eta}(\tilde{\varphi}_{q,\eta} + \Phi_q(\eta)(\varphi) - K_{q,\eta}(\tilde{\varphi}_{q,\eta} + \Phi_q(\eta)(\varphi)))^2 \\ & = & K_{q,\eta}(\tilde{\varphi}_{q,\eta} - K_{q,\eta}(\tilde{\varphi}_{q,\eta}))^2 \\ & = & K_{q,\eta}(\varphi - \Phi_q(\eta)(\varphi))^2 - G_{q-1}^2\left[M_q(\varphi) - \Phi_q(\eta)(\varphi)\right]^2 \end{array}$$

D'où

$$\begin{array}{rl} & \mu_{q-1}\{K_{q,\mu_{q-1}}\left[Q_{q,n}(1) - K_{q,\mu_{q-1}}(Q_{q,n}(1))\right]^2\} \\ = & \mu_{q-1}\{K_{q,\mu_{q-1}}(Q_{q,n}(1) - \Phi_q(\mu_{q-1})(Q_{q,n}(1)))^2 - G_{q-1}^2\left[M_q(Q_{q,n}(1)) - \Phi_q(\mu_{q-1})(Q_{q,n}(1))\right]^2\} \\ = & \mu_q(Q_{q,n}(1) - \mu_q Q_{q,n}(1))^2 - \mu_{q-1}\{G_{q-1}^2\left[M_q(Q_{q,n}(1)) - \Phi_q(\mu_{q-1})(Q_{q,n}(1))\right]^2\} \end{array}$$

Ainsi $s_n = s_n^1 - s_n^2$ avec

$$\begin{cases} s_n^1 := \frac{1}{\gamma_n(1)^2}\sum_{p=1}^{n+1}\gamma_p(1)^2\mu_q(Q_{q,n}(1) - \mu_q Q_{q,n}(1))^2 \\ s_n^2 := \frac{1}{\gamma_{n+1}(1)^2}\sum_{p=1}^{n+1}\gamma_p(1)^2\mu_{q-1}\{G_{q-1}^2\left[M_q(Q_{q,n}(1)) - \Phi_q(\mu_{q-1})(Q_{q,n}(1))\right]^2\} \end{cases}$$

On remarque que dans notre cas particulier $s_n^2 = 0$ et que

$$s_n^1 = \sum_{p=1}^{n} \mathbb{E}\left(\left[\frac{\mathbb{P}(B_n|B_p)}{\mathbb{P}(B_n|B_{p-1})}\mathbf{1}_{B_p} - 1\right]^2 |B_{p-1}\right)$$

D'où le résultat recherché. Mais

$$\begin{aligned}
s_n &= \sum_{p=1}^{n} \left[\mathbb{P}\left(B_p^c | B_{p-1}\right) + \mathbb{P}\left(B_p | B_{p-1}\right) \left(\frac{\mathbb{P}\left(B_n | B_p\right)}{\mathbb{P}\left(B_n | B_{p-1}\right)} - 1 \right)^2 \right] \\
&= \sum_{p=1}^{n} \left[1 - P_p + P_p \left(\frac{P_{p+1} \cdots P_n}{P_p \cdots P_n} - 1 \right)^2 \right] = \sum_{p=1}^{n} \left[1 - P_p + P_p \left(\frac{1}{P_p} - 1 \right)^2 \right] \\
&= \sum_{p=1}^{n} \left(\frac{1}{P_p} - 1 \right)
\end{aligned}$$

D'où le résultat.

En prenant les probabilités de transition égales à P_0, s_n se simplifie en

$$s_n = n \left(\frac{1}{P_0} - 1 \right)$$

qui conduit au résultat du Corollaire 2.4.1. □

ANNEXE 3 : ALGORITHME MCMC

Simulation d'une chaîne de Markov $(X_n)_{n\geq 0}$ à valeurs dans un ensemble discret I, de loi initiale λ et de matrice de transition P

Puisque $\sum_{i\in I} \lambda_i = 1$, on peut créer une partition de $[0,1]$ en des sous-intervalles disjoints $(A_i)_{i\in I}$ de longueur $(\lambda_i)_{i\in I}$:

$$|A_i| = \lambda_i$$

De même, on peut, pour tout $i \in I$, créer une partition de $[0,1]$ en des sous-intervalles disjoints $(A_{ij})_{j\in I}$ de longueur $(p_{ij})_{j\in I}$:

$$|A_{ij}| = p_{ij}$$

On définit ensuite les fonctions

$$G_0 : [0,1] \to I \ ; \ u \to i \ \text{ si } \ u \in A_i$$

$$G : I \times [0,1] \to I \ ; \ (u,i) \to j \ \text{ si } \ u \in A_{ij}$$

Soit $(U_0, U_1, \cdots)$ une suite de variables aléatoires i.i.d de loi uniforme sur $[0,1]$ et posons

$$\begin{cases} X_0 = G_0(U_0) \\ X_{n+1} = G(X_n, U_{n+1}) \ \text{ pour } \quad n \geq 0 \end{cases}$$

Alors $\mathbb{P}(X_0 = i) = \mathbb{P}(U_0 \in A_i) = \lambda_i$ et

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n) = \mathbb{P}(U_{n+1} \in A_{i_n i_{n+1}}) = p_{i_n i_{n+1}}$$

Ainsi $(X_n)_{n\geq 0}$ est une chaîne de Markov (λ, P).

Cette technique permet d'étudier empiriquement les aspects du comportement de la chaîne qui s'avèreraient impossibles à étudier de façon théorique.

Algorithme de Markov Chain Monte Carlo

On se donne une loi $\pi = (\pi_i, i \in I)$ avec I discret et on désire calculer

$$\sum_{i \in I} \pi_i f_i$$

en supposant que π n'est pas facile à simuler directement. L'idée basique de l'algorithme MCMC est la suivante : simuler une chaîne de Markov $(X_n)_{n \geq 0}$ à valeurs dans I et de mesure invariante π. Alors en supposant l'apériodicité et l'irréductibilité, on sait que pour $n \to +\infty$ la loi de X_n converge vers π et

$$\frac{1}{n} \sum_{k=1}^{n-1} f(X_k) \to \sum_{i \in I} \pi_i f_i$$

Comme on ne sait pas simuler directement selon π, simuler une chaîne de Markov entière plutôt qu'une simple variable aléatoire de loi π nous permet de résoudre le problème. Ce qui est assez troublant.

Il faut maintenant exhiber une matrice de transition P pour la chaîne telle que π est invariante pour P c-à-d

$$\pi_i P_{ij} = \pi_j P_{ji} \tag{108}$$

Il y a plusieurs choix possibles pour cela.

Soit R une matrice de transition quelconque. On peut déterminer un P qui vérifiera (108) en prenant

$$\begin{cases} \pi_i P_{ij} = (\pi_i R_{ij}) \wedge (\pi_j R_{ji}) & i \neq j \\ P_{ii} = 1 - \sum_{j \neq i} P_{ij} \end{cases}$$

On vérifie aisément que (108) est satisfaite.

Interprétation : Si $X_n = i$, on simule une nouvelle variable Y_n telle que $Y_n = j$ avec la probabilité R_{ij}. Si $Y_n = j$, on pose

$$X_{n+1} = \begin{cases} Y_n & \text{avec la probabilité} \quad \frac{\pi_j R_{ji}}{\pi_i R_{ij}} \wedge 1 \\ X_n & \text{sinon} \end{cases}$$

C'est l'algorithme de Hastings ou encore d'acceptation-rejet. Reste le choix de R. Il existe deux cas particuliers très utilisés : l'échantillonneur de Gibbs et l'algorithme de Métropolis.

Pour l'échantillonneur de Gibbs, on prend

$$R_{ij} = (\sum \pi_k)^{-1}\pi_j$$

alors on a aussi

$$P_{ij} = (\sum \pi_k)^{-1}\pi_j$$

Pour l'algorithme de Métropolis, on prend simplement

$$R_{ij} = R_{ji} \text{ pour tout i et j}$$

et on trouve pour $i \neq j$,

$$P_{ij} = R_{ij}(\frac{\pi_j}{\pi_i} \wedge 1) \text{ pour } i \neq j$$

Un exemple simple est le suivant : on prend $R_{ij} = \frac{1}{|I|-1}$ si $i \neq j$. Cela revient à proposer une nouvelle valeur j de façon uniforme et aléatoire. Si $\pi_j > \pi_i$, on adopte tout le temps la nouvelle valeur, sinon on adopte la nouvelle valeur avec la probabilité π_j/π_i. Ce qui permet de favoriser les plus «gros poids».